UNIVERSITY OF CALIFORNIA, SAN DIEGO

Chemical, Biochemical and Molecular Genetic Interrogation of Biosynthetically Robust Marine Cyanobacteria

A Dissertation submitted in partial satisfaction of the requirement for the degree
Doctor of Philosophy

in

Marine Biology

by

Rashel Vina Grindberg

Committee in Charge:

Professor William H. Gerwick, Chair
Professor Pieter Dorrestein
Professor Vivian Hook
Professor Paul Jensen
Professor Bradley S. Moore

2010

The Dissertation of Rashel Vina Grindberg is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____
Chair

University of California, San Diego

2010

DEDICATION

"If I have seen further it is only by standing on the shoulders of giants."
– Sir Isaac Newton, circa 1676

These are my giants.

To David I. Peterson, Velva High School distinguished teacher and coach, for believing in me before I knew Balance, Eyes, Elbow, and Follow–through. You showed me that a goal driven by passion is worth risking the unattainable because in the end "there is no report card for life". I learned from you first that dedication and hard work far outweigh what the scoreboard reads at the end of the game. And I now recognize that "a girl, a basketball, and a dream" is a metaphor limited only by my imagination.

To Dr. Bruce R. Jensen, Jamestown College Professor of Biology, for introducing me to the world of ideas. You were my candle in the dark, lighting the path of reason and logic when I most needed it. I learned from you that science is not merely a discipline but a system of knowing, and for me it became a way of life. Thank you for believing in my potential and for holding me up through all odds.

To Dr. Aswani K. Volety, Florida Gulf Coast University Professor of Marine Science, for providing the greatest undergraduate research opportunity I could have ever wished for. Without your guidance, my stellar laboratory skills would have never been realized. Thank you for your continued mentorship and friendship as I progress through my career.

To Stacy Lipp-Carlin, Jami Calavera-Beffert, Tami Haack-Kocon, Ryan Perhus and Wayne Hoffner for your undivided support through the years, for the constant stream of good-times and for not taking me to the hospital after hours of begging for an M.D. Among many things, I'll never forget 3-man, state fair road trips, "quad pact" and jenga. Well, maybe I'll try to forget about jenga.

To Micheal C. Wilson for being my pillar of support, voice of reason and source of sanity when I needed it most. I will forever cherish the laughs, hugs, trails and slopes we've shared.

To my Dad, Anna Marie, Archie, Sarah, and Cody, the best family a silly little girl from North Dakota could ask for. Thank you for believing in me, unconditionally.

Lastly and mostly, to the person that has made extraordinary sacrifices for me and to whom I attribute nothing less than all of my success. She is the bedrock upon which I stand and the loving person I strive to become; my mother, Rachely Sayo Beroncal-Grindberg. Thank you for all you've done mom, forever and always.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A	adenylation
aa	amino acid
ACP	acyl carrier protein
AMP	adenosine monophosphate
AT	acyl gransferase
BLAST	basic local alignment search tool
bp	base pair
C	condensation
C-C	carbon-carbon
CoA	coenzyme A
CP	carrier protein
Cy	cyclization
DH	dehydratase
DIG	digoxigenin
DNA	deoxyribonucleic acid
DMSO	demethylsulfoxide
E	epimerase
ECH	enoyl-CoA hydratase
ER	enoyl reductase
FTICR	fourier transform ion cyclotron resonance
gDNA	genomic DNA
GNAT	GCN5-related *N*-acetyltransferase
GNMT	glycine *N*-methyltransferase
HCS	3-hydroxy-3-methyl-glutaryl-CoA synthase
HMGCS	3-hydroxy-3-methyl-glutaryl-CoA synthase
HMW	high molecular weight
HPLC	high pressure liquid chromatography

Kb	kilobase
kDA	kilodalton
KR	ketoreductase
KS	ketosynthase
MALDI	matrix assisted laser desorption ionization
MDA	multiple displacement amplification
MS	mass spectrometry
MT	methyl transferase
NCBI	National Center for Biotechnology Information
NMR	nuclear magnetic resonance
N-Mt	*N*-methyltransferase
NRPS	nonribosomal peptide synthetase
ORF	open reading frame
PBS	phosphate buffered saline
PCP	peptidyl carrier protein
PCR	polymerase chain reaction
PKS	polyketide synthase
SAM	S-adenosyl methionine
T	thiolation
TE	thioesterase
TOF	time of flight
VLC	vacuum liquid chromatography

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

CURRICULUM VITAE

Education

2005-2010 **Ph.D**., Marine Biology,
Scripps Institution of Oceanography

2003-2005 Pre-doctoral student, Molecular and Cellular Biology,
Oregon State University

2000-2002 **B.A.**, Biology, Florida Gulf Coast University

1995-2000 **B.A.**, Physical Education, Jamestown College

Professional Experience

2005-present *Graduate Research Assistant* - Center for Marine Biotechnology & Biomedicine, Scripps Institution of Oceanography under the guidance of Dr. William H. Gerwick.

Isolated, annotated and characterized the biosynthetic pathway for the highly cytotoxic drug candidate, apratoxin A.

- created and screened genomic DNA libraries
- used bioinformatic analyses to annotate gene clusters
- cloned and expressed proteins for biochemical interrogation
- chromatography, nuclear magnetic resonance and mass spectrometry techniques for substrate and product verification and new natural product isolation

2007 *Graduate Teaching Assistant* – University of California at San Diego, Recombinant DNA Techniques,
Dr. Ramlah Nehring

2003-2005 *Research Fellow*, Oregon State University, OSU Pipeline Fellowship, Dr. William H. Gerwick

2002-2003 *Undergraduate Research Assistant*, Florida Gulf Coast University, Dr. Aswani Volety

Studied the impacts of watershed alteration on oyster health and the role of oyster reefs as essential fish habitat.

- Determine prevalence of oyster parasite, *Perkinsus marinus*
- Identify, and quantify oyster reef decapods and fish
- Investigate physiological and ecological effects on oysters in response to environmental and contaminant stress

2001 *Undergraduate Laboratory Technician*, Florida Gulf Coast University, Dr. S.G.Tolley
Managed the Marine Science Laboratory
- Supervised the field and laboratory work of six interns

Specialized Skills and Techniques

Laboratory: molecular biology: genomic DNA isolation and library production, RNA isolation, library screening by Southern hybridization, polymerase chain reaction, gene cloning, vector construct design for protein over-expression, protein detection and purification, biochemistry chromatography and detection methods: vacuum-, fast protein-, and reverse phase high pressure- liquid chromatography (LC), LC- mass spectrometry (MS), MALDI-MS.

Computer and bio-informatics: Vector NTI, MEGA, ClustalW, ORF finder, nucleotide and protein blast programs, blastx, tblastn, tblastx, molecular modeling, chemdraw, marine lit.

Selected Honors and Awards

2006	Teaching Assistantship Award, University of California at San Diego
2004	Teaching Assistantship Award, Oregon State University
2003	OSU Pipeline Fellowship, Oregon State University
2002	Florida Gulf Coast University Outstanding Student Research Award

---

Publications

Grindberg, R.V., Shuman, C.F., Sorrels, C.M., Wingerd, J., Gerwick. W.H. (2008) *Neurotoxic Alkaloids from Cyanobacteria*. **Modern Alkaloids**: Structure, Isolation, Synthesis and Biology (pp.139-162). WILEY-VCH Verlag GmbH & Co.

Gerwick, W. H.; Coates, R. C.; Engene, N.; Gerwick, L. G.; Grindberg, R. V.; Jones, A.C.; Sorrels, C. M., Giant Marine Cyanobacteria Produce Exciting Potential Pharmaceuticals**.** ***Microbe,*** 2008**,** 3, (6), 8.

Gu, L., Wang, B., Kulkarni, A., Geders, T.W., Grindberg, R.V., Gerwick, Lena G., Håkansson, K., Wipf, P., Smith, J., Gerwick, W.H., & Sherman, D.H. Metamorphic enzyme assembly in polyketide diversification. ***Nature,*** 2009.
459(7247): p. 731-35.

Grindberg, R.V., Pevzner P., Roger Lasken, R., Ishoey T., Gerwick, L., Dorrestein, P.,Gerwick, W.H. Single Cell Genome Amplification Cultivates Discovery of the Biosynthetic Gene Cluster for the Highly Cytotoxic Cyclodepsipeptide, Apratoxin A. ***Nature Chemical Biology,*** 2009. Manuscript in progress.

Conference Presentations

2009 Rashel V. Grindberg, Pavel Pevzner, Dumitru Brinza, Roger Lasken, Thomas Ishoey, Lena Gerwick, Pieter Dorrestein, William H. Gerwick. Single-cell Genome Amplification of *Lyngbya bouillonii* Cultivates Discovery of the Elusive Apratoxin Gene Cluster American Society of Pharmacognosy, National Meeting, Oahu, Hawaii.

2008 Rashel V. Grindberg, L. Gu, D. H. Sherman, W. H. Gerwick. Biochemical Investigations into the β-alkylating events in the Jamaicamide Biosynthetic Pathway. Society for Industrial Microbiology (SIM), San Diego, California.

2006 Rashel V. Grindberg and W. H. Gerwick. Partial isolation of the biosynthetic pathway for the cytotoxic, cyclodepsipeptide, Apratoxin, from the marine cyanobacteria, *Lyngbya bouillonii*. Gordon Research Conference, Ventura, California.

2004 Rashel V. Grindberg, P. Flatt, W. H. Gerwick. Isolation of the cyclodepsipeptide, Apratoxin A, from the marine cyanobacterium, *Lyngbya bouillonii*. Bioorganic Conference. Mount Rainier, Washington.

2002 Rashel V. Grindberg, E. Rasnake, A. Volety. Monitoring the impact of water management practices on estuarine health. Estuarine Research Federation (ERF), Mystic, Connecticut.

ABSTRACT OF THE DISSERTATION

Chemical, Biochemical and Molecular Genetic Interrogation of Biosynthetically Robust Marine Cyanobacteria

by

Rashel Vina Grindberg

Doctor of Philosophy in Marine Biology

University of California, San Diego, 2010

Professor William H. Gerwick, Chair

Marine cyanobacteria exhibit a high production of bioactive and structurally diverse natural products. A number of these secondary metabolites or their derivatives are lead compounds in drug development programs aimed at providing new therapies to treat cancer, bacterial infections, inflammatory responses, and in crop protection to kill harmful microbial pathogens and insects. Isolation and structural analysis of marine and terrestrial cyanobacterial natural products has provided access to a large number of mixed non-ribosomal peptide synthetase/polyketide synthase (NRPS/PKS) systems. In chapter 1 of this

dissertation I provide a brief history of marine natural products, arguing that the discipline was slow to begin but has grown to become a highly valuable and resourceful discipline. Also, I provide background on the genetic architecture of secondary metabolite pathways citing examples of both canonical and less-common catalytic domains in these megasynthases. Chapter 2 profiles the biochemical robustness of the marine cyanobacterium, *Lyngbya bouillonii.* Due in part to little physiological variation among marine cyanobacteria, I argue that many of the compounds attributed to a *Lyngbya* sp. or *Lyngbya majuscula* were likely isolated from *L. bouillonii.* Chapters 3 and 4 are dedicated to the attempt of isolating the apratoxin A gene cluster from two different *L. bouillonii* strains. The first of these chapters describes my effort to use traditional molecular genetic techniques on a strain that was not growing in culture. This approach resulted in the isolation of a pathway inconsistent with the predicted biosynthesis of apratoxin A. In my second attempt, described in chapter 4, I used the new and fundamentally necessary approach of single cell isolation and whole genome amplification, together with traditional techniques to successfully target and isolate the apratoxin gene cluster. The project described in chapter 5 provided me the opportunity to investigate another dimension of marine cyanobacteria. I assayed the biochemical activity of the specialized ECH_1, ECH_2 and ER catalytic domains. I was able to show that the jamaicamide enzymes were catalytically active towards the non-chlorinated jam substrate.

Chapter 1. General Introduction

1.1 Marine Natural Products

Natural products are a functionally diverse class of biochemically-synthesized compounds. Many are classified as secondary metabolites; that is, they rarely have a role in primary metabolism, growth, or reproduction but have evolved to somehow benefit the producing organisms in other, adaptive ways. In the nineteenth and earlier centuries, natural product extracts, particularly those derived from botanical species, provided the main source of folk medicines (Schmidt, Ribnicky et al. 2007). However, in the latter part of the nineteenth century, biologically active organic molecules began to be isolated in relatively pure form for medicinal use. For example, salicylic acid (**1**), the precursor of aspirin, was isolated in 1874 from willow bark (Raskin 1992). Other more potent painkillers, such as morphine (**2**) and codeine (**3**), were isolated from the opium poppy (Ziegler, Beasley et al. 1975). The anti-malarial agent, quinine (**4**), was separated from cinchona (china bark) (Middeldorf and Montigny 1951). The leaves of the purple foxglove plant, *Digitalis purpurea*, provided an excellent source of digitoxin (**5**) that was purified for use against heart disease (Van Dyke and Li 1935; Ahmed 2007) (Figure 1.1). There are numerous other examples. Synthesis of the first synthetic pharmaceutical drug, aspirin, occurred in the latter half of the nineteenth century. However, it was not until the early 1900s that the recognition of aspirin as a universal pain reliever was realized and this discovery spawned the era of therapeutic agents.

Decades later, in the late-1950s, the concept of 'drugs from the sea' finally attracted interest within the scientific community. Beginning in 1951, Werner Bergmann published reports (Bergmann and Feeney 1950; Bergmann and Feeney 1951; Bergmann and Burke 1955) of unusual arabino- nucleosides obtained from marine sponges collected in Florida (spongothymidine (**6**) and spongouridine (**7**)). This was a significant finding in that the discovery of naturally produced ribo-pentosyl nucleosides overturned the existing dogma of the time:

> "For a nucleoside to have biological activity, it had to have ribose or deoxyribose as the sugar, but the base could comprise a multiplicity of heterocycles and even carbocycles" (Newman and Cragg 2004) .

This paradigm shift ushered in new incentive to look for biologically active nucleosides, which directly led to the development of the chemical derivatives ara-A (adenine arabinoside) (**8**) and ara-C (cytosine arabinoside) (**9**), two nucleosides with significant antiviral and anticancer properties, respectively (Newman and Cragg 2004). From the period of 1950 to the early 1970s, the discovery of marine natural products accelerated by mere increase in the level of sophistication of the techniques used. High-field nuclear magnetic resonance or NMR (originally 200 MHz and then up through 600-800 MHz), mass spectrometry that involved MS-MS techniques, and chromatographic methods of all types were typical tools of the trade. Finally, by the late 1970s, the use of high-pressure liquid chromatography (HPLC) became generally available (Newman and Cragg 2004). By 1975, there were already three parallel tracks of research in marine natural product chemistry: marine toxins, marine

bioproducts and marine chemical ecology (Burja, Banaigs et al. 2001). Today, marine natural products chemists have determined the chemical structures of over 16,000 unique chemical entities.

In the areas of cancer and infectious disease, 60% and 75% of new drugs, respectively, originated from natural sources between 1981 and 2002 (Newman, Cragg et al. 2003). Between 2001 and 2005, 23 new drugs derived from natural products were introduced for the treatment of disorders such as bacterial and fungal infections, cancer, diabetes, dyslipidemia, atopic dermatitis, Alzheimer's disease and genetic diseases such as tyrosinaemia and Gaucher disease (Table 1.1). However, only very recently in 2004, did the first drug from the sea, Ziconotide (**10**) get approved in the United States. Ziconotide or by other designations, Prialt®, SNX-111 and ω-conotoxin MVIIA, is a 25 amino acid linear peptide possessing three disulfide bonds. It was originally isolated from the venom of the predatory Indo-Pacific marine mollusk, *Conus magus* (Olivera 2000). *C. magus* and other *Conus* species are fish-hunting mollusks that use their venom to paralyze their prey (Kohn 1956). The remarkable analgesic activity of ziconotide (the compound proved to be 1,000 times more active than morphine in animal models of antinociception) is due to the blockage of calcium channels (Olivera 2000), (McCleskey, Fox et al. 1987) and is currently in use for the treatment of chronic pain.

A second drug from the sea is the tetrahydroisoquinoline alkaloid trabectedin (Yondelis/ecteinascidin-743/ET-743) (**11**) which was originally purified from the marine tunicate *Ecteinascidia turbinata* in 1990 (Rinehart, Holt et al. 1990). This

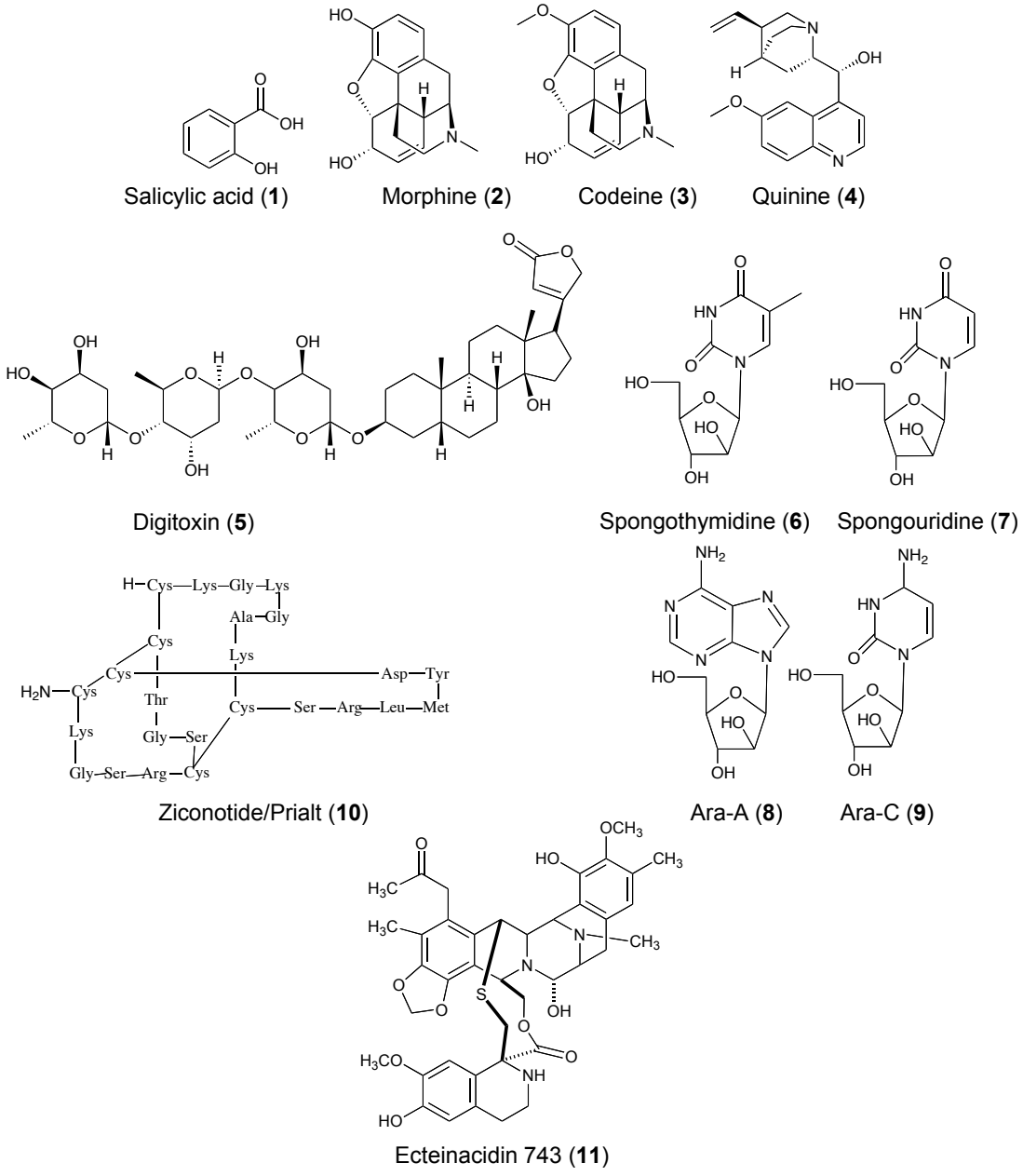


Figure 1.1. Biologically active compounds first isolated from natural sources.

drug exhibited broad-spectrum anti-tumour activity and was especially effective against solid tumors such as sarcomas and breast cancer (Valoti, Nicoletti et al. 1998). Trabectedin was approved by the European Union in October 2007 for the treatment of soft-tissue sarcoma (NA 2006) and also has the potential to be a beneficial treatment for advanced or recurrent ovarian cancer (Carter and Keam 2007).

Several other candidate compounds from marine sources are presently in the drug discovery pipeline and being evaluated in Phase I–III clinical trials in the United States and in Europe for the treatment of various cancers (Chin, Balunas et al. 2006; Butler 2008; Blunt, Copp et al. 2009), (Fig. 1.2, table 1.2). Some examples include aplidin (**12**) isolated from the Mediterranean tunicate, *Aplidium albicans,* (Rinehart, Gloer et al. 1981), halichondrin B (**13**) isolated from *Halichondria okadai* which is a common, widely distributed sponge off the Pacific coast of Japan (Uemura 1986), bryostatin 1 (**14**) from the bryozoan *Bulgula neritina* (Pettit, Herald et al. 1982), dolastatin 15 (**15**) isolated from the Indian Ocean sea hare *Dolabela auricularia*, (Pettit, Kamano et al. 1989), and kahalalide F (**17**) isolated from the mollusk *Elysia rufescens* (Hamann and Scheuer 1993).

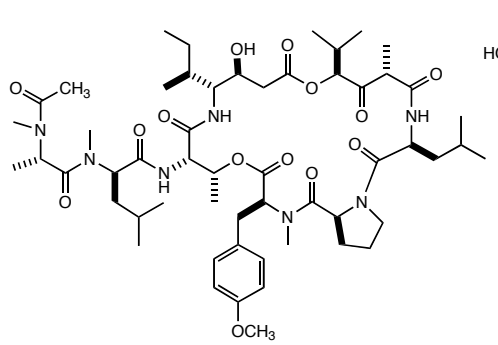
Marine cyanobacteria have a rich history of producing a multitude of secondary metabolites including alkaloids like somocystinamide (**18**) (Nogle and Gerwick 2002), polyketides such as kalkipyrone (**19**) (Graber and Gerwick 1998),

Table 1.1 Drugs derived from natural products launched in Europe, Japan and the United States 2001–2005. Adapted from (Lam 2007)

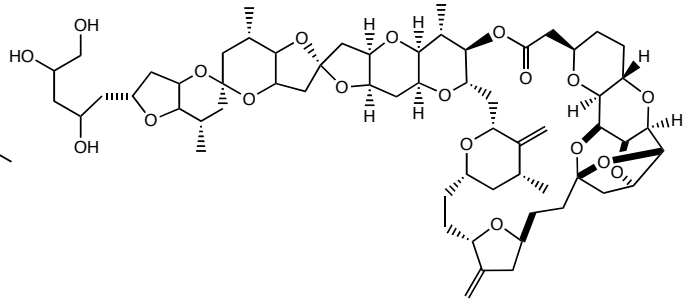
Year	Generic name (trade name)	Natural product	Indications
2001	Caspofungin (Cancidas®)	Pneumocandin B	antifungal
2001	Pimecrolimus (Elidel®)	Ascomycin	atopic dermatitis
2001	Telithromycin (Ketek®)	Erythromycin	antibacterial
2002	Amrubicin hydrochloride (Calsed®)	Doxorubicin	anticancer
2002	Biapenem (Omegacin®)	Thienamycin	antibacterial
2002	Ertapenem (Invanz™)	Thienamycin	antibacterial
2002	Fulvestrant (Faslodex®)	Estradiol	anticancer
2002	Galantamine (Reminyl®)	Galantamine	alzheimer's disease
2002	Micafungin (Funguard®)	FR901379	antifungal
2002	Nitisinone (Orfadin®)	Leptospermone	antityrosinaemia
2003	Daptomycin (Cubicin™)	Daptomycin	antibacterial
2003	Miglustat (Zavesca®)	1-deoxynojirimycin	type 1 Gaucher disease
2003	Mycophenolate sodium (Myfortic®)	Mycophenolic acid	immunosuppression
2003	Pitavastatin (Livalo®)	Mevastatin	dyslipidemia
2003	Rosuvastatin (Crestor®)	Mevastatin	dyslipidemia
2004	Everolimus (Certican™)	Sirolimus	immunosuppression
2004	Talaporfin sodium (Laserphyrin®)	Chlorophyll and l-aspartic acid	anticancer
2005	Doripenem (Finibax®)	Carbapenem	antibacterial
2005	Extenatide (Byetta®)	Incretin	anti-diabetic
2005	Paclitaxel nanoparticles (Abraxane®)	Taxol	anticancer
2005	Pramlintide acetate (Symlin®)	Amylin	anti-diabetic
2005	Tigecycline (Tigacil®)	Tetracycline	antibacterial
2005	Zicontide (Prialt™)	MVIIA	pain management

Table 1.2 List of lead compounds in the drug discovery pipeline. Table adapted from (Butler 2008) and updated**.**

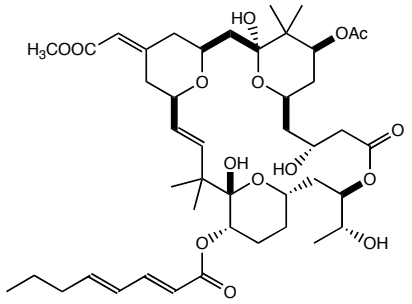
Lead Compound and Source	Name (synonym)	Mechanism of Action	Development Status	Developer
Ascidian				
plitidepsin (aplidin) (**12**)	plitidepsin (Aplidin)	VEGF and VEGFR1 inhibitor, G1/G2 phase cell cycle inhibitor	Phase II	PharmaMar
Sponge				
halichondrin B (**13**)	eribulin (E7389, NSC707389)	tubulin assembly inhibition	Phase II/III	Eisai
hemiasterlin	E7974	tubulin assembly inhibition	Phase I	Eisai
psammaplin A (**16**)	panobinostat (LBH-589)	HDAC inhibition	Phase I/II/III	Novartis
Others				
bryostatin 1 (**14**)	bryostatin 1	Protein kinase C inhibition	Phase I/II	NCI
jorumycin	Zalypsis (PM00104/50)	DNA binding and transcriptional activity	Phase I	PharmaMar
dolastatin 15 (**15**)	tasidotin (synthadotin, ILX-651)	tubulin assembly inhibition	Phase II	Genzyme
kahalalide F (**17**)	alters lysosomal membrane function	alters lysosomal membrane function	Phase I	PharmaMar



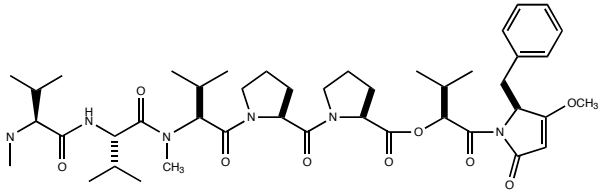
Aplidine (**12**)



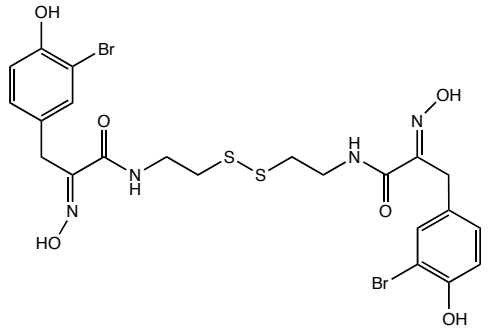
Halichondrin B (**13**)



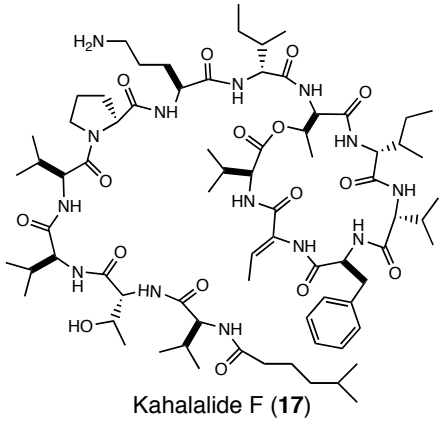
Bryostatin 1 (**14**)



Dolastatin 15 (**15**)



Psammaplin A (**16**)



Kahalalide F (**17**)

Figure 1.2. Lead compounds currently in the drug discovery pipeline.

and nonribosomal peptides, such as the dragonamides (**20**) (Jimenez and Scheuer 2001; Gunasekera, Ross et al. 2008) some of which are potent toxins (Fig 1.4) (Apeldoorn, Egmond et al. 2007). Representatives of these compounds span the chemical classes of hydrocarbons, long chain alcohols, fatty acids, glycolypids, aromatic polyketides, lactones, linear peptides, simple- and macrocyclic- peptides, cyclic depsipeptides, linear and ring containing alkaloids, isoprenoids, proteins and aromatic compounds to name a few (Van Wagoner, Drummond et al. 2007) (for an exhaustive list of compounds isolated from marine cyanobacteria and their corresponding biological activities, the reader is directed to the references (Burja, Banaigs et al. 2001; Gerwick, Tan et al. 2001; Tidgewell, Clark et al. 2010). Cyanobacteria in general are a large and morphologically diverse group of phototrophic prokaryotes which occur in almost every habitat on earth (Thajuddin and Subramanian 2005). Species and strains from all of the common planktonic cyanobacterial genera including *Anabaena, Aphanizomenon, Cylindrospermopsis, Microcystis, Nodularia, Nostoc, Oscillatoria, Lyngbya, Scytonema* and *Tolypothrix* produce unique chemical substances (Burja, Banaigs et al. 2001). More specifically however, nearly 800 compounds have been reported from marine cyanobacteria (marinLit), and three orders within the Cyanophyta produce the vast majority of compounds. These are the *Chroococcales*, *Nostocales* and *Oscillatoriales* which, in turn produce nearly half of the reported compounds (Fig. I.3 from (Gerwick, Coates et al. 2008)). From the 9 genera currently listed in the *Oscillatoriales*, *Lyngbya* leads in production with nearly 300 distinct substances coming from this single genus.

There are 12 marine species in this genus, and 236 compounds are ascribed to *L. majuscula* and a further 36 to *L. bouillonii* (Gerwick, Coates et al. 2008). Marine organisms belonging to the genus *Lyngbya,* are filamentous cyanobacteria that grow abundantly within tropical and sub-tropical waters. This genus is the source for such bioactive compounds as antillatoxin (**21**) (Berman, Gerwick et al. 1999), barbamide (**22**) (Orjala and Gerwick 1996), carmabin (**23**), curacin A (**24**) (Berman, Gerwick et al. 1999), jamaicamide A (**25**) (Edwards, Marquez et al. 2004), and lyngbyatoxin A (**26**) (Edwards and Gerwick 2004).

1.2 Modular Type I Polyketide Synthases

Modular polyketide synthases (PKSs) are large (100 to 10,000 kDa), multifunctional enzymes that are responsible for the production of many clinically important drugs such as tetracycline, danuorubicin, erythromycin, rapamycin and lovastatin (Khosla, Gokhale et al. 1999). Analogous to fatty acid synthases, the individual active sites of type I PKSs are part of a single multifunctional polypeptide, whereas type II PKSs comprise several monodomain enzymes each expressed from a distinct gene (Hopwood and Sherman 1990). A third class of PKSs, designated as type III PKSs, are composed of enzymes that iteratively catalyze chain elongation using malonyl-CoA-derived building blocks (Shen 2003). All three types of PKSs use acyl coenzyme A monomers as extender units to form elaborate chemical structures with a ketide backbone. A brief overview of modular type I PKSs is presented in the following section.

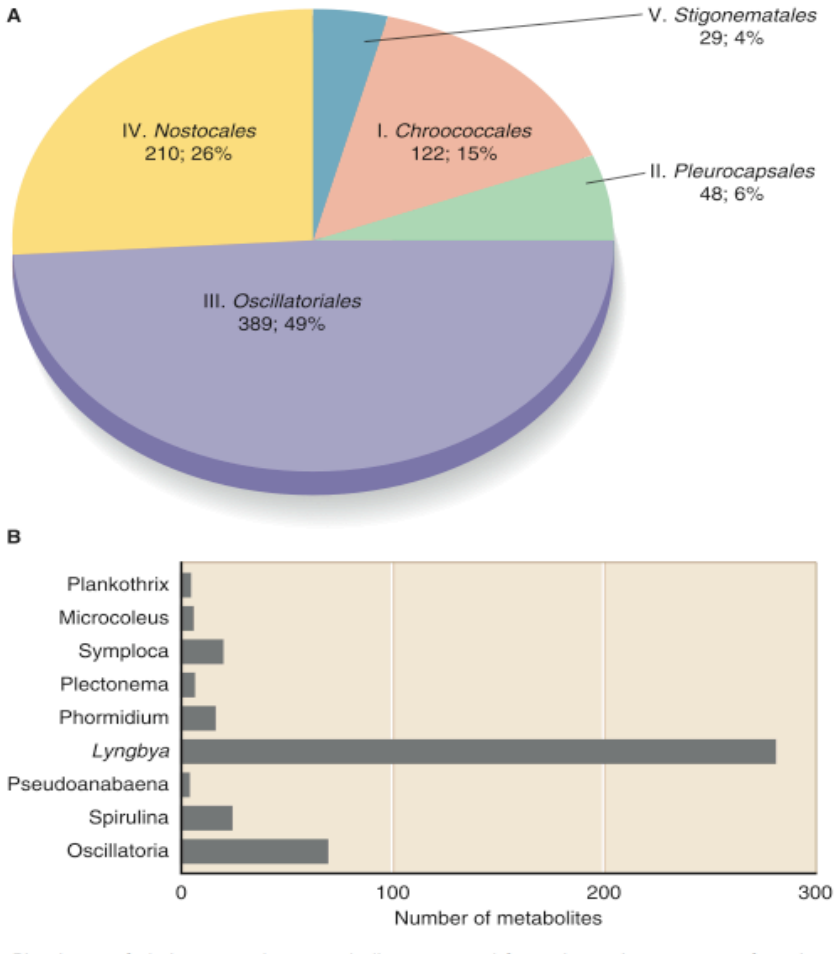
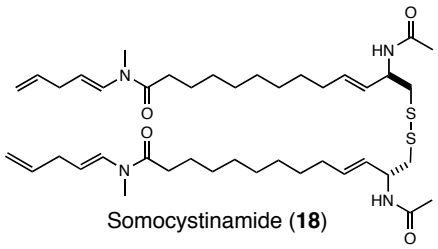
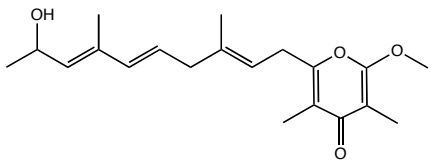


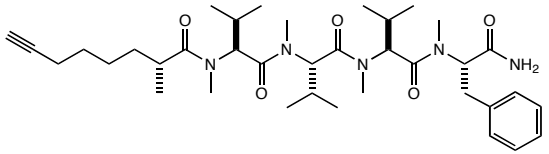
Figure 1.3 Charts of a) secondary metabolites reported from the various groups of marine cyanobacteria (types I to V with botanical order equivalents provided), and b) the metabolites of type III cyanobacteria Oscillatoriales) separated by genus (Gerwick, Coates et al. 2008).



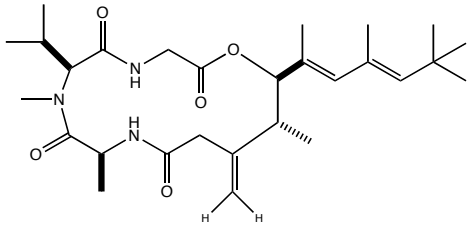
Somocystinamide (**18**)



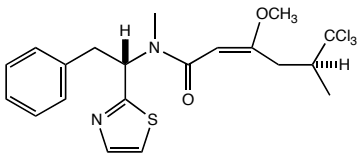
Kalkipyrone (**19**)



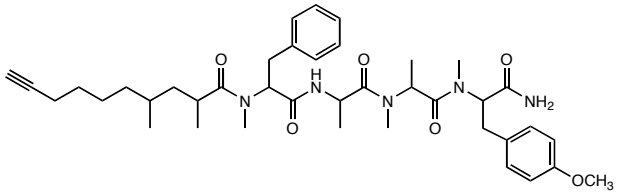
Dragonamide (**20**)



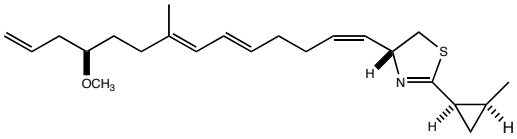
Antillatoxin (**21**)



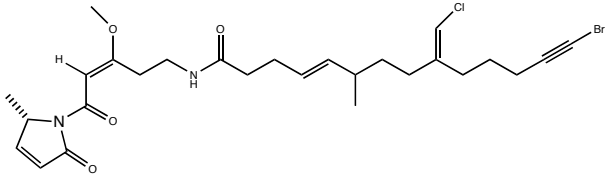
Barbamide (**22**)



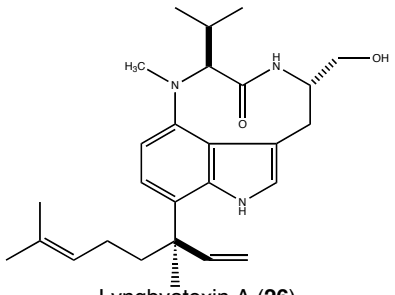
Carmabin (**23**)



Curacin A (**24**)



Jamaicamide A (**25**)



Lyngbyatoxin A (**26**)

Figure 1.4. Natural products isolated from marine cyanobacteria.

Although PKSs are similar to fatty acid synthases (FASs) they can vary considerably in their architecture, whether they are iterative or modular and whether intermediates are tethered to proteins or cofactors. Modular type I PKSs are organized into groups of active sites in which each module is responsible for one cycle of polyketide chain extension and functional group modification. Natural products created by these megasynthases possess polyketides with central carbon skeletons formed by iterative decarboxylative condensations of malonic acid (or substituted malonic acid) thioesters yielding polyacetate, polypropionate, or polybutyrate chains (Figure 1.4). The type I PKS involved in the biosynthesis of 6-deoxyerythronolide B, the non-glycosylated precursor of erythromycin A is one of the best studied (Cortes, Haydock et al. 1990; Donadio, Staver et al. 1991). These large enzymes have both minimum domains and tailoring domains that are covalently linked in long polypeptides and organized into modules. A "minimal module" is termed as such because it contains all of the catalytic domains necessary for the selection and incorporation of a single monomeric unit extending the central carbon skeleton by one acetate unit. These domains include the β-ketosynthase (KS), acyl transferase (AT), and acyl carrier protein (ACP). Ancillary domains that define the oxidation state of the β-carbon are often present, but are not compulsory. These domains include the β-ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER).

The first step in polyketide biosynthesis is the modification of the apo-ACP to the holo form. This involves the posttranslational attachment of the

phosphopantetheinyl arm of coenzyme A onto a conserved serine residue (Copp and Neilan 2006) (Figure 1.7). The terminal phosphopantetheine thiol serves as a tether for the fixation of starting materials and intermediates throughout biosynthesis (Mercer and Burkart 2007) (Figure 1.8). The next step is mediated by the AT domain which transfers the activated precursors of the 2-, 3- and 4- carbon building blocks of acetyl-CoA, propionyl-CoA, butyryl-CoA, (malonyl-, methylmalonyl-, and ethylmalonyl-CoA, respectively) to the corresponding modified carrier domain or ACP (Chan, Podevels et al. 2009). This domain has been referred to in the literature as the "gatekeeper of polyketide biosynthesis" (Khosla, Gokhale et al. 1999). For example in the DEBS pathway, the AT domains within each of the downstream modules (from the AT in the loading module or AT_L) exhibit strict structural and stereochemical specificity for (2*S*)-methylmalonyl-CoA, their native substrate. Polyketide synthases requiring acetate-derived building blocks use alternative AT domains that are specific for malonyl-CoA. The substrate specificity of the AT-domain for malonyl-CoA or methylmalonyl-CoA has been shown to be determined by the presence of certain conserved amino acid residues (Haydock, Aparicio et al. 1995). Pioneering work done by Haydock et. al determined different AT binding pocket motifs and active site amino acids confer the predictable incorporation of an acetate unit *or* a propionate unit into the developing polyketide.

The KS domain mediates the fundamental chain elongation reaction and catalyzes the C-C bond formation through a Claisen-type condensation. The KS domain first decarboxylates the malonyl or methylmalonyl extender unit bound to

the ACP domain. The resulting nucleophile then reacts with the thioester group of the polyketide chain attached to the KS domain. Ultimately this decarboxylative condensation results in the extension of the growing ketide chain by two carbons (Figure 1.5) (Cane and Walsh 1999).

The β-carbonyl processing domains KR, DH and ER catalyze three consecutive reductive reactions (Figure 1.9). Before transfer of the elongated polyketide chain to the KS domain of the next module, the β-keto group can be reduced to a hydroxyl by the KR domain and requires NADPH as a co-factor. This reaction results in a β-carbon stereocenter which was previously non-asymmetric. One could hypothesize at this point whether or not the KR domain confers stereospecificity on that stereocenter.

The study of the KR domains in polyketide biosynthesis has a rich academic and scientific history and as such I will take a few paragraphs here to unfold the story. Understanding the relationship between the stereochemical configuration of a chiral center in a molecule and the primary structure of a protein can provide insight into how a stereochemical outcome is controlled. A good model to illustrate this logic is the molecular basis for Celmer's Rule. A common type of PKS module elongates the acyl chain with a methylmalonyl moiety and reduces the condensation product as far as the alcohol (Khosla, Gokhale, Jacobsen, & Cane, 1999). In this situation, the 2-methyl-3-hydroxyacyl moiety produced has two new chiral centers

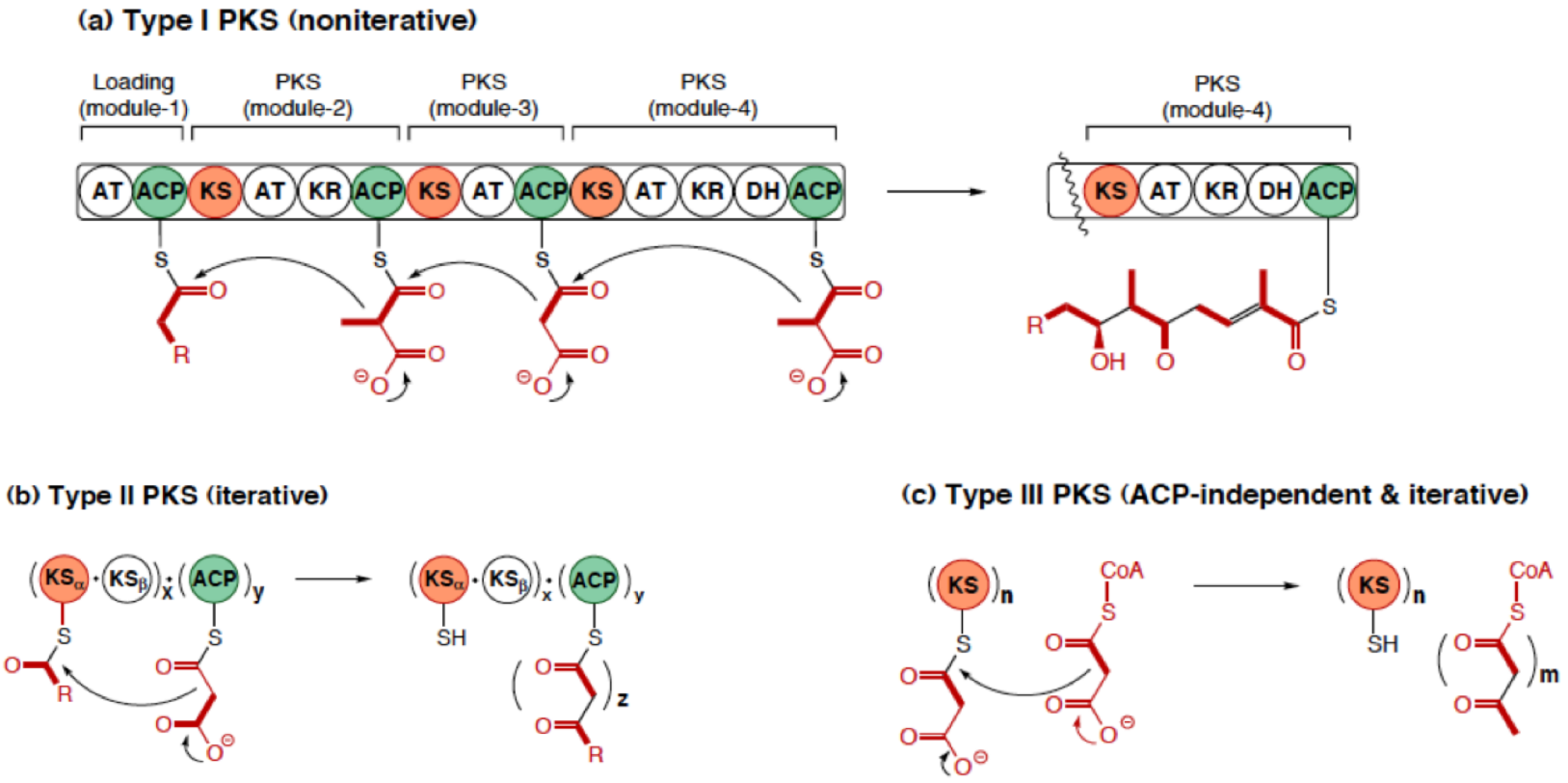


Figure 1.5. Three commonly encountered architectures of bacterial PKSs. (Shen 2003)

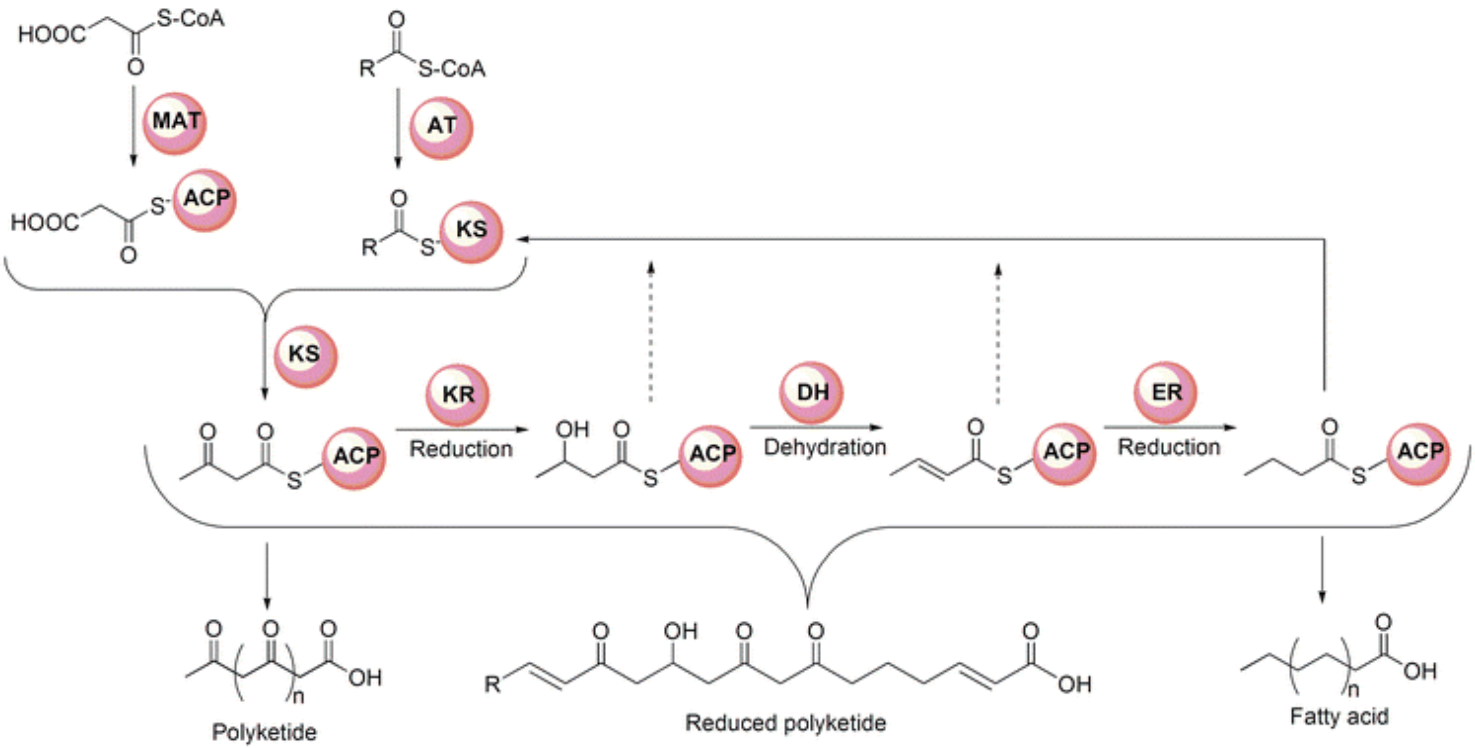


Figure 1.6. Generic reaction scheme for PKSs (Smith and Tsai 2007) MAT = malonyl-acyl transferase, ACP = acyl carrier protein, AT = acyl transferase, KS = ketosynthase, KR = ketoreductase, DH = dehydratase, ER = enoyl reductase

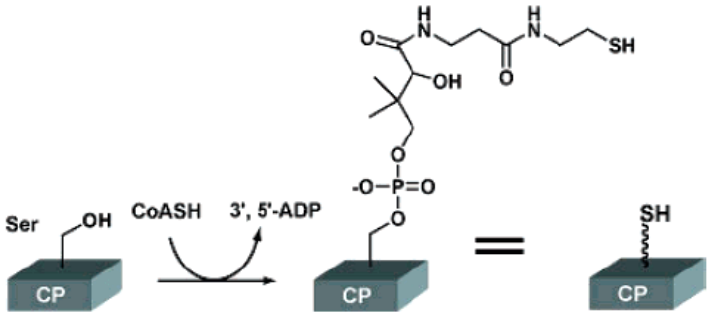


Figure 1.7. Post-translational modification of carrier proteins (CPs).(Lai, Koglin et al. 2006)

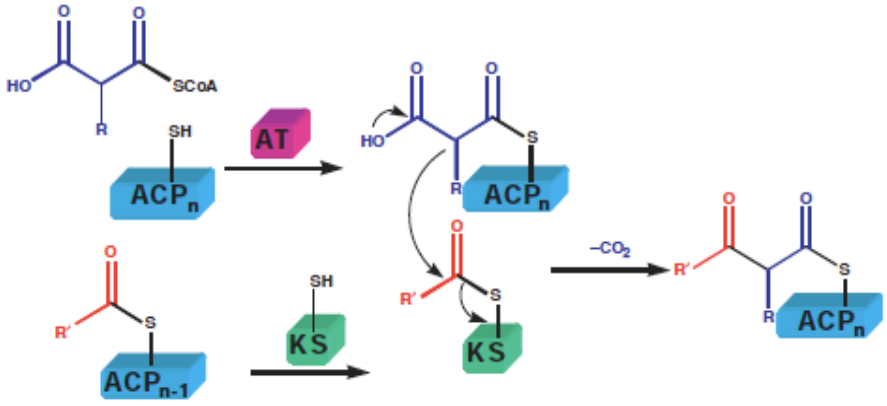


Figure 1.8. Function and mechanism of PKS core catalytic and carrier domains. (Cane and Walsh 1999)

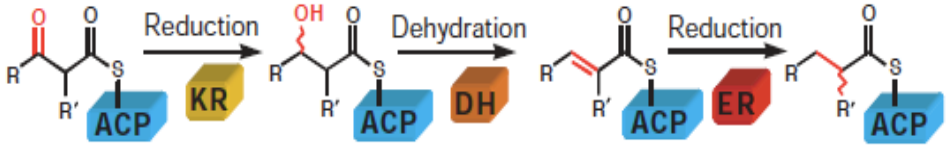


Figure 1.9. Auxiliary PKS domains mediating ketoreduction (KR), dehydration (DH), and enoylacyl reduction (ER) (Cane and Walsh 1999)

and all four combinations of methyl and alcohol stereochemistry are theoretically possible: (2*R*, 3*S*), (2*S*, 3*R*), (2*R*, 3*R*), (2*S*, 3*S*). The (2*S* or 2*R*, 3*R*) configurations are most widely seen in nature (Smith and Tsai 2007). As far back as 1965, W. D. Celmer (Celmer 1965) noted that despite the variation within a given polyketide structure, there are strong position-specific structural and stereochemical homologies among the hundreds of known polyketides. His inference that these patterns were indicative of a common genetic origin for stereochemical control at these centers is referred to as Celmer's rule.

More recently, domain exchange experiments have shown that KR domains determine the stereochemical outcome of the ketoreduction reaction (McDaniel, Kao et al. 1997; Kao, McPherson et al. 1998). It has been established that there are two types of ketoreductases that lead to two different alcohol stereochemistries, 3*R* or 3*S* (Caffrey 2003). Based on sequence alignment, the stereoselective signature motifs for the modular PKS KRs have been proposed to be "LDD' and PxxxN, and the presence of these motifs predicts formation of the 3*R* steroisomer (or type B-KR) whereas the absence of these motifs predicts the 3*S* stereoisomer (or type A-KR) (Reid, Piagentini et al. 2003).

Continuing with the rest of the ancillary domains, the DH catalyzes the dehydration of the β-hydroxyl group resulting in the formation of a C=C double bond. Finally, the ER domain uses an NADPH co-factor to form a saturated bond in the ketide chain (Figure 1.9) (Cane and Walsh 1999). When a newly added priopionate extender unit undergoes reduction by an ER domain in a modular PKS system, the resulting methyl branch may have either S or R configuration. A correlation between

the presence or absence of a unique tyrosine residue in the ER active site and the chirality of the methyl branch that is introduced has been recently discovered (Kwan, Sun et al. 2008). When a tyrosine residue occupies this position in the active site, the methyl branch has S configuration, otherwise it has R configuration. This is a powerful tool with highly predictive properties and could be used to address difficult stereochemical questions that current spectroscopic strategies cannot resolve. For example, despite concerted efforts to determine the nature of the stereocenter C9 in jamaicamide A, it remains unknown. If the logic of stereochemical control by the ER domains is applicable when the methyl group is introduced by SAM (as opposed to a propionate extender unit), then bioinformatic analysis allows us to assign C9 in jamaicamide A with *R* configuration. After the final extension, the nascent ketide chain is transferred to the TE domain which is responsible for product release and in some cases, cyclization.

1.3 Non-canonical Biosynthetic Mechanisms Prominent in Marine Cyanobacterial Secondary Metabolite Biosynthesis

1.3.1 SAM-dependent methyltransferases and stereochemical considerations

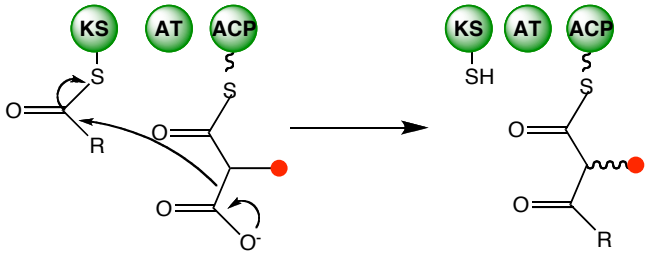
There are three known ways to produce a pendant methyl group or its equivalent on the carbon skeleton of polyketides. The first is by incorporation of an activated form of methylmalonate as an extender unit introducing a propionyl moiety into the nascent polyketide chain after condensation (Khosla, Gokhale et al. 1999) which produces the pendant methyl carbon at the C-2 postion. The second and third ways are via SAM-dependent methylation and *via* an HMG-CoA synthase (HCS) like

reaction. As cyanobacteria are known to utilize only the latter two ways of producing a C-methyl functionality, I will provide an overview of SAM-dependent methylation and HCS-like reactions.

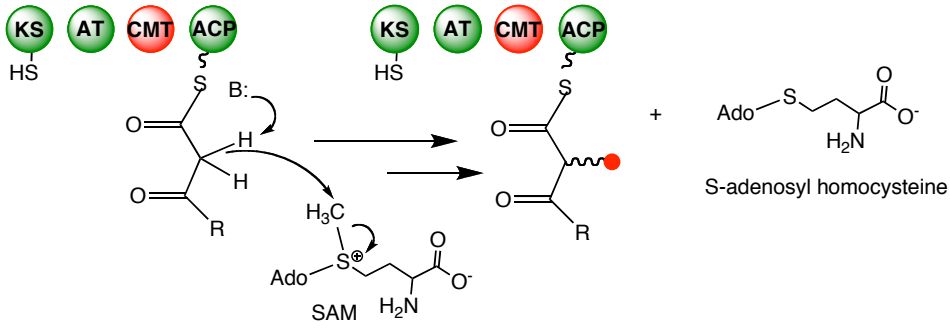
S-adenosylmethionine (AdoMet or SAM) is a very commonly used cofactor as a methyl donor in alkylation reactions central to cellular biochemistry and primary metabolism. SAM-dependent methyltransferases (SAM-MTs) are a family of enzymes that utilize SAM to methylate a myriad of substrates and several atom types. Nearly 120 members of this family have been discovered and classified into one of five different subclasses based on the atom targeted for methylation and substrate specificity; small molecule, protein, DNA, RNA and other MTases. Of those 117 types of SAM-MTs, 7 types methylate sulfur, 14 methylate carbon, 41 methylate nitrogen and 55 methylate oxygen.

From the mid 1970's to the mid 1980's research on SAM-dependent methylation chemistry showed convincingly that methyltransferase reactions proceed *via* a nucleophilic attack on the methyl group of SAM (Hegazi, Borchard et al. 1976). It was also observed that the inversion of configuration at the methyl group indicates that an odd number of nucleophilic displacements takes place and thus suggests an S_N2-like transition state (Woodard, Tsai et al. 1980).

1) Use methylmalonyl as an extender unit



2) SAM-dependent methyltransferase



3) HMG-Co-A synthase-like reaction

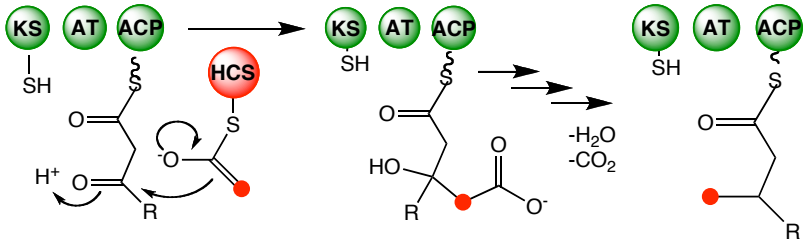


Figure 1.10 Three different routes to pendant methyl group formation on polyketide backbones.

Although the S_N2-type mechanism appears to be conserved across the family of enzymes, the presence of an inherent nucleophile on the acceptor molecule needs further consideration. It is likely (but not yet definitively shown) that N, O, and S-methylation occur *via* a straightforward attack by the oxygen, nitrogen or sulfur lone pair of electrons. In contrast, an *sp*3 hybridized carbon atom does not possess a lone pair of electrons that could attack the SAM methyl group. It is therefore necessary for a CMTase to carry out steps during catalysis to generate a carbanion equivalent at the target carbon.

A number of compounds produced by marine cyanobacteria display C-methylation domains that are predicted to utilize SAM to impart a branched methyl functionality. For example, JamJ in jamaicamide A (Edwards, Marquez et al. 2004), Cur J in the curacin pathway (Chang, Sitachitta et al. 2004), and HctD in the hectochlorin pathway (Ramaswamy, Sorrels et al. 2007) all code for CMT domains.

1.3.2 β-Alkylation *via* HMG-CoA Synthase Like Reactions

Of considerable interest are the recently discovered β-alkylation events involving an HCS-like gene and associated tailoring domains embedded within PKS/NRPS pathways. To date, several biosynthetic gene clusters for the production of secondary metabolites possessing this biosynthetic motif have been reported, such as the pathways for curacin A (Chang, Sitachitta et al. 2004), jamaicamide A (Edwards, Marquez et al. 2004), mupirocin (El-Sayed, Hothersall et al. 2003), bacillaene (Calderone, Kowtoniuk et al. 2006; Butcher, Schroeder et al. 2007; Moldenhauer, Chen Rainer et al. 2007), pederin (Piel, 2002), and myxovirescin (Simunovic, Zapp et al.

2006). Each have been annotated as having stand alone homologs of the HCS gene, a set of genes encoding one or more ACPs, a mutant KS with a cys-to-ser active site substitution (KSs) and two homologs of the enoyl-CoA hydratase (ECH_1 and ECH_2) family. The current working hypothesis is that these gene products condense acetyl-CoA with the ß-ketoacyl-S-ACP intermediate of the growing polyketide to eventually result in the addition of the C-2 acetate carbon in the structure (Figure 1.12, A). Furthermore, elaboration upon that carbon has yielded diverse functionalities including an α-β unsaturated pendant methyl group in bacillaene and mupirocin, a cyclopropyl ring in curacin A, a vinylic chloride in jamaicamide A and an α-β saturated pendant methyl group in apratoxin A (Figure 1.11).

Recently, the *pks*X cluster, from *B. subtilus*, that codes for the production of bacillaene was characterized as having this type of architecture. Calderone *et al.* (Calderone, Kowtoniuk et al. 2006) observed the following: 1) an upstream AT domain (PksC) selectively loaded malonyl-CoA (over acetyl-CoA and methylmalonyl-CoA) and transferred malonyl to the carrier protein, AcpK, and was decarboxylated by PksF (KSq). 2) The resulting acetate functionality, in a PksG-mediated reaction (HCS), underwent an aldol condensation with acetoacetyl-S-PksL to yield HMG-S-PksL (Figure 1.12, A). The metabolic reactions for the downstream enzymes in the curacin A pathway were also characterized. Gu *et al.* (2006) demonstrated that the CurE/CurF ECH_1-ECH_2 enzyme pair catalyzes the successive dehydration and decarboxylation of (S)-HMG-ACP to generate a 3-methylcrotonyl-ACP intermediate for subsequent formation of the cyclopropane ring (Figure 1.12, B).

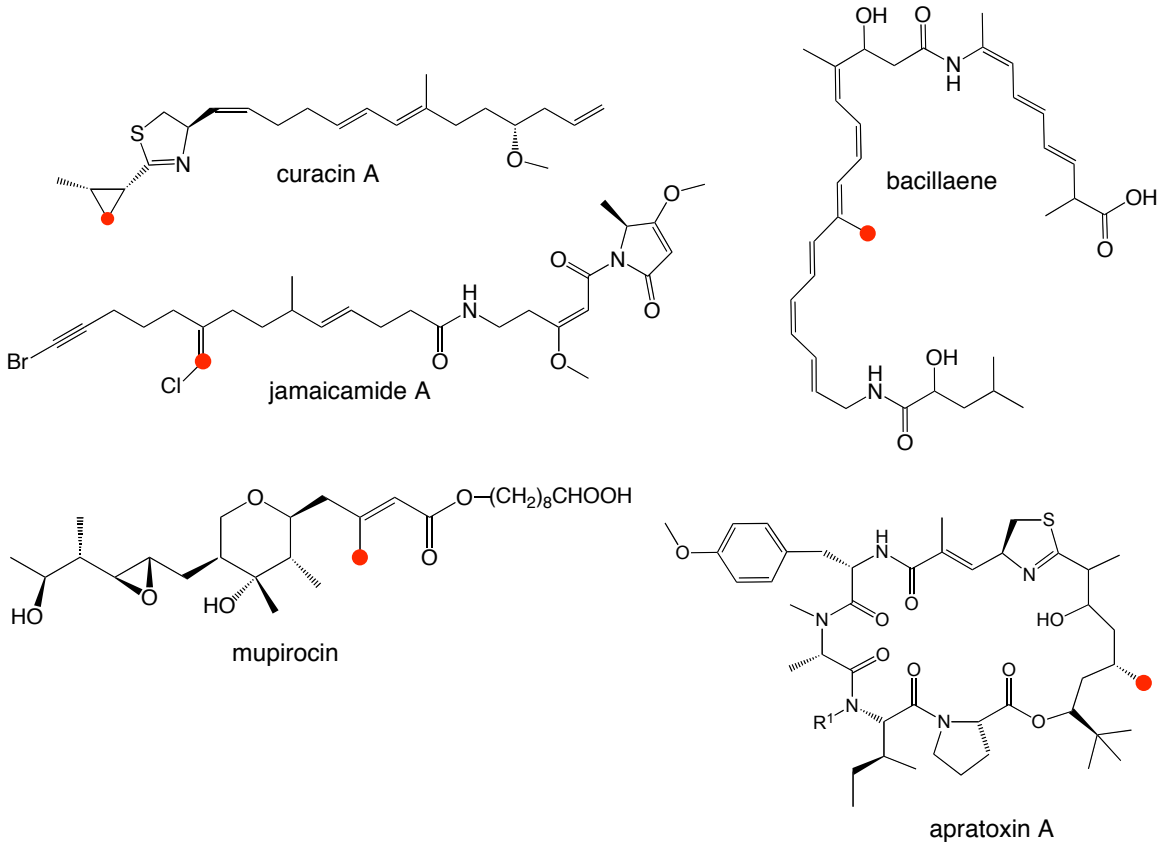


Figure 1.11. Secondary metabolites with an (HCS)-like gene and associated tailoring domains responsible for yielding diverse functionalities (red dot indicates C-2 of acetate) including an α-β unsaturated pendant methyl group in bacillaene from *Bacillus subtilis* and mupirocin from *Pseudomonas fluorescens*, a fully saturated methyl moiety in apratoxin A from *Lyngbya bouillonii*, a cyclopropyl ring in curacin A, and a vinylic chloride in jamaicamide A both from *Lyngbya majuscula*.

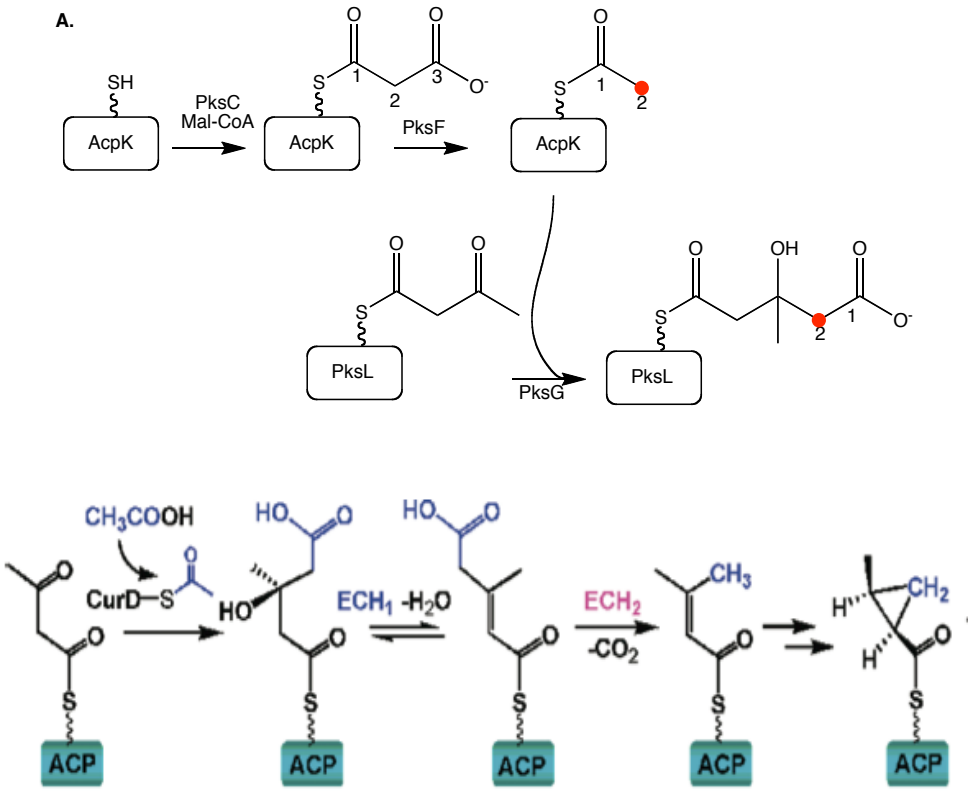


Figure 1.12. A. PksX scheme for loading of malonyl and decarboxylation of malonyl-ACP to acetate, which is then transferred to the HCS domain (Calderone, Kowtoniuk et al. 2006). B. Curacin A scheme for the activity of ECH_1 and ECH_2 for subsequent formation of the cyclopropane ring (Gu, Jia et al. 2006).

1.4 Nonribosomal Peptide Synthetases

Nonribosomally produced peptides are some of the most functionally diverse and structurally complex entities of all natural products. This biochemical pathway has yielded several compounds with important bioactivities such as antibiotics (tyrocidin A (Mootz and Marahiel 1997), gramicidin S (K. Turgay 1992)), immunosuppressives (cyclosporine A (Weber, Schörgendorfer et al. 2004)), and cytostatic activity (yersiniabactin (Gehring, DeMoll et al. 1998) and enterobactin (Gehring, Mori et al. 1998)).

Nonribosomal peptide synthetases (NRPSs) are comprised of multimodular enzymatic assembly lines that typically possess one module for each amino acid monomer incorporated (Schwarzer, Finking et al. 2003). A minimal module consists of a carrier protein/peptidyl carrier protein/thiolation (CP/PCP/T) domain and two catalytic regions, the 50 kDa adenylation (A) and 50 kDa condensation (C) domains. The usual order of domains in a module is C-A-T. In analogy to the PKSs, the NRPSs similarly have domains that are non-compulsory. These include, but are not limited to, catalytic domains that epimerize (E), heterocyclize (Cy), and/or N- , C- and O-methylate (N-Mt, C-Mt, and O-Mt) the developing compound (Finking and Marahiel 2004).

The adenylation (A) domains are responsible for the selection of the amino- (or carboxy-) acid substrates which are then activated as acyl adenylates at the expense of ATP (Figure 1.10) and subsequently incorporated into the nascent peptide chain. The C domain is the chain-elongating catalyst, joining an upstream electrophilic peptidyl-carrier

protein to the down stream nucleophilic aminoacyl-carrier protein. Chain elongation by one aminoacyl residue occurs concomitantly with chain translocation to the downstream carrier protein. Product release is achieved by a two-step process that involves an acyl-*O*-TE-enzyme intermediate that is subsequently attacked by either a peptide-internal nucleophile or water which results either in a macrocyclic product or the linear peptide (Cane and Walsh 1999).

Although the Cy-domain has been categorized as a tailoring domain, it is the only enzyme that can replace a core-domain of the NRPS. Because Cy-domains replace C-domains in a given module, they catalyze, in addition to heterocyclization of cysteine, serine, and threonine residues, the formation of the peptide bond. The E domain catalyzes the epimerization of the PCP-bound l-amino acid or C-terminal amino acid of the growing polypeptide. NRPSs contain methyltransferases that are responsible for the *N*, *C*, or *O*-methylation of amino acid residues, thus making the peptide less susceptible to proteolytic breakdown. These methyltransferases (*N*-MT, *C*-MT) use *S*-adenosyl methionine as the methyl donor. (Figure 1.14) (Cane and Walsh 1999)

Nonribosomally produced peptides exhibit of a number of structural features that distinguish them from ribosomally synthesized compounds. For example, they contain non-proteinogenic amino acids like ornithine, the structures are typically macrocyclic, branched macrocyclic or represent dimers of two or trimers of three identical structural elements. Also common among them are small heterocyclic rings like thiazoles, oxazolines or thiazolines. Finally, theses peptides contain a high degree of *N*, *C*, and *O*-methylations and glycosylations as well as insertion of acetate or propionate units (Fischbach and Walsh 2006).

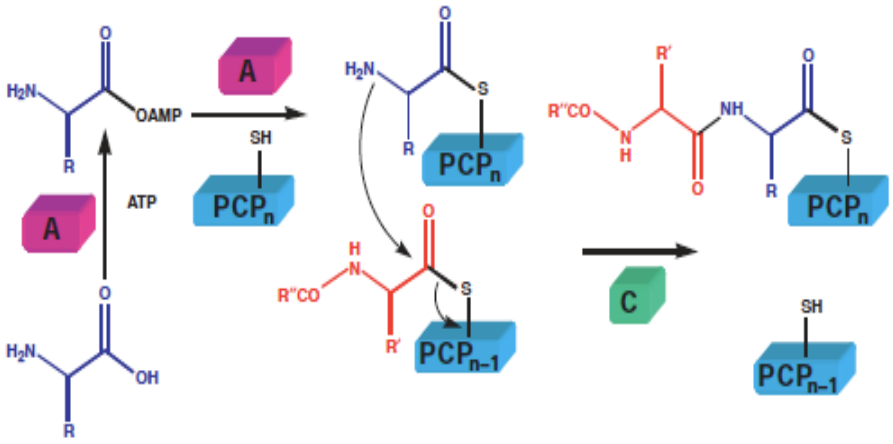


Figure 1.13. Function and mechanism of NRPS core catalytic and PCP domains. (Cane and Walsh 1999)

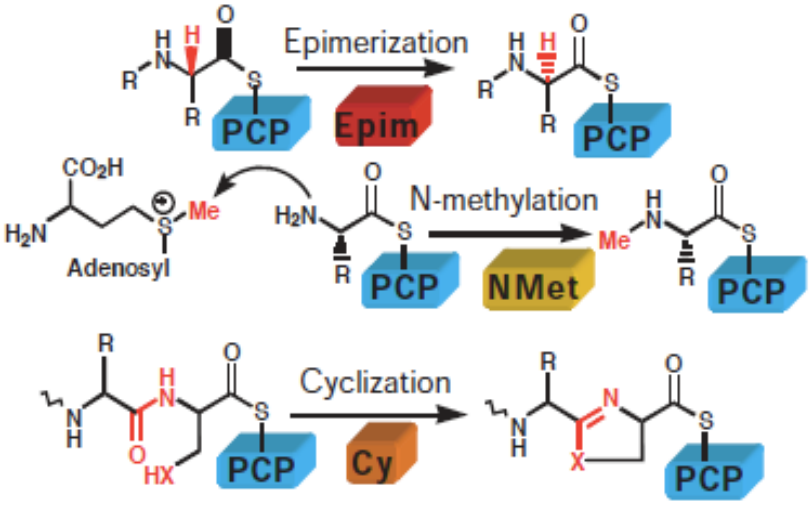


Figure 1.14. Auxiliary NRPS domains mediating epimerization (Epim), peptide N-methylation (NMet), and heterocyclization (Cy) of cysteine or serine residues. (Cane and Walsh 1999)

1.5 References

Ahmed, A. (2007). "Digoxin and reduction in mortality and hospitalization in geriatric heart failure: importance of low doses and low serum concentrations." J Gerontol A Biol Sci Med Sci **62**(3): 323-9.

Apeldoorn, M. E. V., H. P. V. Egmond, et al. (2007). "Toxins of cyanobacteria." Molecular Nutrition & Food Research **51**(1): 7-60.

Bergmann, W. and D. C. Burke (1955). "Contributions to the study of marine products. XXXIX. The nucleosides of sponges. III. Spongothymidine and spongouridine. ." The Journal of Organic Chemistry **20**(11): 1501-1507.

Bergmann, W. and R. J. Feeney (1950). "The isolation of a new thymine pentoside from sponges. ." Journal of the American Chemical Society **72**(6): 2809-2810.

Bergmann, W. and R. J. Feeney (1951). "Contributions to the study of marine products. XXXII. The nucleosides of sponges. I. ." The Journal of Organic Chemistry **16**(6): 981-987.

Berman, F. W., W. H. Gerwick, et al. (1999). "Antillatoxin and kalkitoxin, ichthyotoxins from the tropical cyanobacterium *Lyngbya majuscula*, induce distinct temporal patterns of NMDA receptor-mediated neurotoxicity." Toxicon **37**(11): 1645-8.

Blunt, J. W., B. R. Copp, et al. (2009). "Marine natural products." Nat Prod Rep **26**(2): 170-244.

Burja, A. M., B. Banaigs, et al. (2001). "Marine cyanobacteria-a prolific source of natural products." Tetrahedron **57**(46): 9347-9377.

Butcher, R. A., F. C. Schroeder, et al. (2007). "The identification of bacillaene, the product of the pksX megacomplex in *Bacillus subtilis*." Proc Natl Acad Sci U S A **104**(5): 1506-9.

Butler, M. S. (2008). "Natural products to drugs: natural product-derived compounds in clinical trials." Natural Product Reports **25**(3): 475-516.

Caffrey, P. (2003). "Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases." Chembiochem **4**(7): 654-7.

Calderone, C. T., W. E. Kowtoniuk, et al. (2006). "Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*." Proc Natl Acad Sci U S A **103**(24): 8977-82.

Cane, D. E. and C. T. Walsh (1999). "The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases." Chem Biol **6**(12): R319-25.

Carter, N. J. and S. J. Keam (2007). "Trabectedin: a review of its use in the management of soft tissue sarcoma and ovarian cancer." Drugs **67**(15): 2257-76.

Celmer, W. D. (1965). "Macrolide Stereochemistry. 3. A Configurational Model for Macrolide Antibiotics." J Am Chem Soc **87**: 1801-2.

Chan, Y. A., A. M. Podevels, et al. (2009). "Biosynthesis of polyketide synthase extender units." Nat Prod Rep **26**(1): 90-114.

Chang, Z., N. Sitachitta, et al. (2004). "Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **67**(8): 1356-67.

Chin, Y. W., M. J. Balunas, et al. (2006). "Drug discovery from natural sources." The AAPS Journal **8**(2): 239-253.

Copp, J. N. and B. A. Neilan (2006). "The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria." Appl Environ Microbiol **72**(4): 2298-305.

Cortes, J., S. F. Haydock, et al. (1990). "An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea." Nature **348**(6297): 176-178.

Donadio, S., M. J. Staver, et al. (1991). "Modular organization of genes required for complex polyketide biosynthesis." Science **252**(5006): 675-679.

Edwards, D. J. and W. H. Gerwick (2004). "Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase." J Am Chem Soc **126**(37): 11432-3.

Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*." Chem Biol **11**(6): 817-33.

El-Sayed, A. K., J. Hothersall, et al. (2003). "Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586." Chem Biol **10**(5): 419-30.

Finking, R. and M. A. Marahiel (2004). "Biosynthesis of nonribosomal peptides." Annu Rev Microbiol **58**: 453-88.

Fischbach, M. A. and C. T. Walsh (2006). "Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms." Chem Rev **106**(8): 3468-96.

Gehring, A. M., E. DeMoll, et al. (1998). "Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*." **5**(10): 573-586.

Gehring, A. M., I. Mori, et al. (1998). "Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF." Biochemistry **37**(8): 2648-2659.

Gerwick, W. H., R. C. Coates, et al. (2008). "Giant marine cyanobacteria produce exciting potential pharmaceuticals." Microbe **3**(6): 8.

Gerwick, W. H., L. T. Tan, et al. (2001). "Nitrogen-containing metabolites from marine cyanobacteria." Alkaloids Chem Biol **57**: 75-184.

Graber, M. A. and W. H. Gerwick (1998). "Kalkipyrone, a toxic gamma-pyrone from an assemblage of the marine cyanobacteria *Lyngbya majuscula* and *Tolypothrix* sp." J Nat Prod **61**(5): 677-80.

Gu, L., J. Jia, et al. (2006). "Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair." J Am Chem Soc **128**(28): 9014-5.

Gunasekera, S. P., C. Ross, et al. (2008). "Dragonamides C and D, linear lipopeptides from the marine cyanobacterium brown *Lyngbya polychroa*." J Nat Prod **71**(5): 887-90.

Hamann, M. T. and P. J. Scheuer (1993). "Kahalalide F: a bioactive depsipeptide from the sacoglossan mollusk *Elysia rufescens* and the green alga *Bryopsis* sp." Journal of the American Chemical Society **115**(13): 5825-5826.

Haydock, S. F., J. F. Aparicio, et al. (1995). "Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases." FEBS Lett **374**(2): 246-8.

Hegazi, M. F., R. T. Borchard, et al. (1976). "Letter: SN2-like transition state for methyl transfer catalyzed by catechol-O-methyl-transferase." J Am Chem Soc **98**(10): 3048-9.

Hopwood, D. A. and D. H. Sherman (1990). "Molecular genetics of polyketides and its comparison to fatty acid biosynthesis." Annu Rev Genet **24**: 37-66.

Jimenez, J. I. and P. J. Scheuer (2001). "New lipopeptides from the caribbean cyanobacterium *Lyngbya majuscula*." J Nat Prod **64**(2): 200-3.

K. Turgay, M. K. M. A. M. (1992). "Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes." Molecular Microbiology **6**(4): 529-546.

Kao, C. M., M. McPherson, et al. (1998). "Alcohol stereochemistry in polyketide backbones is controlled by the beta-ketoreductase domains of modular polyketide synthases." Journal of the American Chemical Society **120**(10): 2478-2479.

Khosla, C., R. S. Gokhale, et al. (1999). "Tolerance and specificity of polyketide synthases." Annual Review of Biochemistry **68**(1): 219-253.

Kohn, A. J. (1956). "Piscivorous gastropods of the genus *Conus*." Proc Natl Acad Sci U S A **42**(3): 168-71.

Kwan, D. H., Y. Sun, et al. (2008). "Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases." **15**(11): 1231-1240.

Lai, J. R., A. Koglin, et al. (2006). "Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis." Biochemistry **45**(50): 14869-79.

Lam, K. S. (2007). "New aspects of natural products in drug discovery." Trends in Microbiology **15**(6): 279-289.

McCleskey, E. W., A. P. Fox, et al. (1987). "Omega-conotoxin: direct and persistent blockade of specific types of calcium channels in neurons but not muscle." Proc Natl Acad Sci U S A **84**(12): 4327-31.

McDaniel, R., C. M. Kao, et al. (1997). "Engineered intermodular and intramodular polyketide synthase fusions." **4**(9): 667-674.

Mercer, A. C. and M. D. Burkart (2007). "The ubiquitous carrier protein- a window to metabolite biosynthesis." Nat Prod Rep **24**(4): 750-73.

Middeldorf, R. and W. Montigny (1951). "Examination of quinine for accessory cinchona alkaloids." Pharmazie **6**(3): 94-8.

Moldenhauer, J., X. H. Chen Rainer, et al. (2007). "Biosynthesis of the antibiotic bacillaene, the product of a giant polyketide synthase complex of the *trans*-AT family." Angewandte Chemie International Edition **46**(43): 8195-8197.

Mootz, H. D. and M. A. Marahiel (1997). "The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains." J. Bacteriol. **179**(21): 6843-6850.

NA (2006). "Trabectedin: ecteinascidin 743, ecteinascidin-743, ET 743, ET-743, NSC 684766." Drugs R D **7**(5): 317-28.

Newman, D. J. and G. M. Cragg (2004). "Advanced preclinical and clinical trials of natural products and related compounds from marine sources." Curr Med Chem **11**(13): 1693-713.

Newman, D. J. and G. M. Cragg (2004). "Marine natural products and related compounds in clinical and advanced preclinical trials." J Nat Prod **67**(8): 1216-38.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural products as sources of new drugs over the period 1981 - 2002." Journal of Natural Products **66**(7): 1022-1037.

Nogle, L. M. and W. H. Gerwick (2002). "Somocystinamide A, a novel cytotoxic disulfide dimer from a Fijian marine cyanobacterial mixed assemblage." Org Lett **4**(7): 1095-8.

Olivera, B. M. (2000). omega-Conotoxin MVIIA: from marine snail venom to analgesic drug. Drugs from the Sea. N. Fusetani, Karger**:** 74-85.

Orjala, J. and W. H. Gerwick (1996). "Barbamide, a chlorinated metabolite with molluscicidal activity from the caribbean cyanobacterium *Lyngbya majuscula*." J Nat Prod **59**(4): 427-30.

Pettit, G. R., C. L. Herald, et al. (1982). "Isolation and structure of bryostatin 1." Journal of the American Chemical Society **104**(24): 6846-6848.

Pettit, G. R., Y. Kamano, et al. (1989). "Isolation and structure of the cytostatic linear depsipeptide dolastatin 15." The Journal of Organic Chemistry **54**(26): 6005-6006.

Ramaswamy, A. V., C. M. Sorrels, et al. (2007). "Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **70**(12): 1977-86.

Raskin, I. (1992). "Role of salicylic-acid in plants." Annual Review of Plant Physiology and Plant Molecular Biology **43**: 439-463.

Reid, R., M. Piagentini, et al. (2003). "A model of structure and catalysis for ketoreductase domains in modular polyketide synthases." Biochemistry **42**(1): 72-79.

Rinehart, K. L., J. B. Gloer, et al. (1981). "Structures of the didemnins, antiviral and cytotoxic depsipeptides from a Caribbean tunicate." Journal of the American Chemical Society **103**(7): 1857-1859.

Rinehart, K. L., T. G. Holt, et al. (1990). "Bioactive compounds from aquatic and terrestrial sources." J Nat Prod **53**(4): 771-92.

Schmidt, B. M., D. M. Ribnicky, et al. (2007). "Revisiting the ancient concept of botanical therapeutics." Nat Chem Biol **3**(7): 360-366.

Schwarzer, D., R. Finking, et al. (2003). "Nonribosomal peptides: from genes to products." Nat Prod Rep **20**(3): 275-87.

Shen, B. (2003). "Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms." Curr Opin Chem Biol **7**(2): 285-95.

Simunovic, V., J. Zapp, et al. (2006). "Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases." Chembiochem **7**(8): 1206-20.

Smith, S. and S. C. Tsai (2007). "The type I fatty acid and polyketide synthases: a tale of two megasynthases." Nat Prod Rep **24**(5): 1041-72.

Thajuddin, N. and G. Subramanian (2005). "Cyanobacterial biodiversity and potential applications in biotechnology." Current Science **89**(1): 47-57.

Tidgewell, K., B. R. Clark, et al. (2010). The natural products chemistry of cyanobacteria.

Uemura, Y. H. a. D. (1986). "Halichondrins—antitumor polyether macrolides from a marine sponge " Pure Appl. Chem. **58**: 701-710.

Valoti, G., M. I. Nicoletti, et al. (1998). "Ecteinascidin-743, a new marine natural product with potent antitumor activity on human ovarian carcinoma xenografts." Clin Cancer Res **4**(8): 1977-83.

Van Dyke, H. B. and R. C. Li (1935). "A study of the standardization of digitalis. Ii. The relationship between laboratory methods of assay and potency as determined by experimental cumulative poisoning and clinical standardization." J Clin Invest **14**(6): 733-7.

Van Wagoner, R. M., A. K. Drummond, et al. (2007). "Biogenetic diversity of cyanobacterial metabolites." Adv Appl Microbiol **61**: 89-217.

Weber, G., K. Schörgendorfer, et al. (2004). "The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame " Current Genetics **26**(2).

Woodard, R. W., M. D. Tsai, et al. (1980). "Stereochemical course of the transmethylation catalyzed by catechol O-methyltransferase." J Biol Chem **255**(19): 9124-7.

Ziegler, H. W., T. H. Beasley, Sr., et al. (1975). "Simultaneous assay for six alkaloids in opium, using high-performance liquid chromatography." J Assoc Off Anal Chem **58**(5): 888-97.

Chapter 2. The Chemistry of *Lyngbya bouillonii*

2.1 Introduction

Cyanobacterial classification systems have historically relied heavily on morphological and phenotypic traits due to relatively little physiological variation among them (Brun and Shimkets 2000). This however, has led to extensive misidentification of species and as a result cyanobacterial taxonomy is in a state of transition (Holt, Krieg et al. 1994). Driving this change is the emergence of molecular genetics and genome sequencing which utilizes and integrates genetic markers such as the16S rRNA gene (Nubel, Garcia-Pichel et al. 1997) and the internally transcribed (ITS) region (Boyer, Flechtner et al. 2001) into the current morphology data sets for a more universal and robust classification system. The most commonly used cyanobacterial taxonomic system developed by Rippka and coworkers is based mostly on morphological characteristics (Rippka, Deruelles et al. 1981; Rippka 1988). In this system cyanobacteria are divided into five subgroups: 1 and 2 contain unicellular cyanobacteria (orders *Chroococcales* and *Pleurocapsales*), and 3 through 5 (orders *Oscillatoriales*, *Nostocales*, and *Stigonematales* respectively) consist of filamentous cyanobacteria in which vegetative trichomes are surrounded by a sheath (Rippka, Deruelles et al. 1981) (Figure 2.1). Generally, cyanobacteria are prokaryotic photoautotrophs and therefore undergo oxygenic photosynthesis. The rich evolutionary history of the interplay between cellular respiration (O_2 production) and nitrogen fixation (via the O_2 labile enzyme, nitrogenase) within cyanobacteria is fascinating, and is reviewed in several articles (Farquhar, Bao et al. 2000; Kasting and Siefert 2002; Beerling 2009).

The genus *Lyngbya* belongs to the order *Oscillatoriales* which by definition are filamentous, undergo binary fission in a single plane and are non-heterocyst and non-akinete producers (Holt, Krieg et al. 1994). Members of this subgroup occupy diverse habitats. The *Lyngbya* species, *L. bouillonii*, was described in 1991 from coral reefs in Papua New Guinea primarily on morphological data. They were observed to form extensive mats strongly attached to madrepores in the infralittoral zone (Hoffmann and Demoulin 1991). Interestingly, this species takes on a distinct morphological pattern, often described as thin, veil-like (or cobweb-like) coverings over holes in the reef, usually occurring between 7 and 35 m deep (Tidgewell, Clark et al. 2010) (Figure 2.1). Often a small shrimp, *Alpheus frontalis*, is found associated with this organism. In 2002, Lohse and colleagues reported on a related shrimp, *Alpheus heterochaelis*, and its predatory mechanism. The shrimp has two claws, one resembling an oversize boxing glove, which it uses to stun prey, such as small crabs, by snapping the oversize claw shut. Interestingly, they reported that the stunning snap comes not from the clap of the claws coming together but from a bubble generated by the claw's rapid closing motion (Versluis, Schmitz et al. 2000). It is thought that the shrimp, *Alpheus frontalis*, uses this dual-purpose predatory technique as a territorial defense mechanism.

Our research group has made several collections of *L. bouillonii* from regions in Papua New Guinea since 2001. The collective chemical profiling efforts of this cyanobacterium over the past 20 years from several research groups have yielded eight distinct classes of metabolites, consisting of some 36 different compounds (Figures 2.3 and 2.4). In the next few pages I will detail the isolation and bioactivity, as well as propose biosynthetic routes to the compounds produced by *L. bouillonii*. A detailed treatment on the isolation of the apratoxin series of compounds is reserved for a later section.

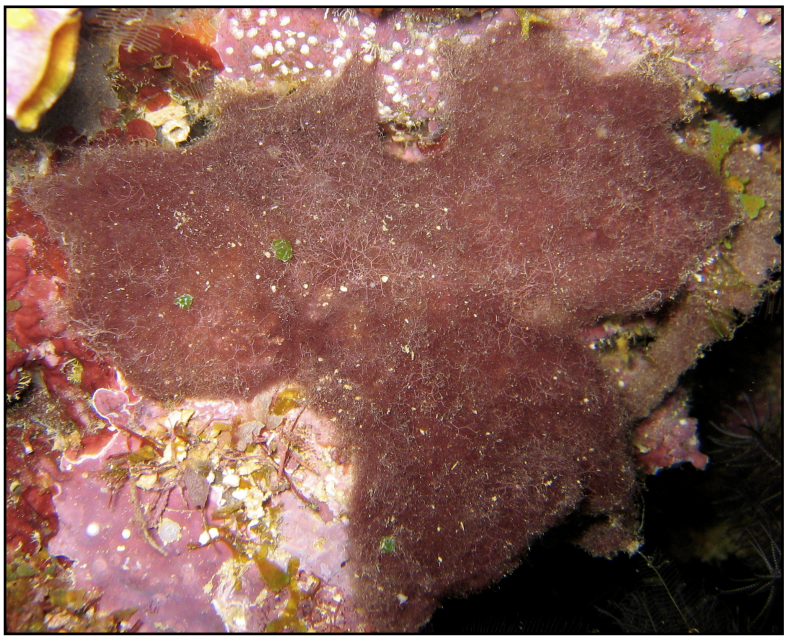


Figure 2.1. Underwater picture of *Lyngbya bouillonii* Hoffman and Demoulin from Papua New Guinea in its natural habitat (photo by Rashel V. Grindberg, PNG 2005).

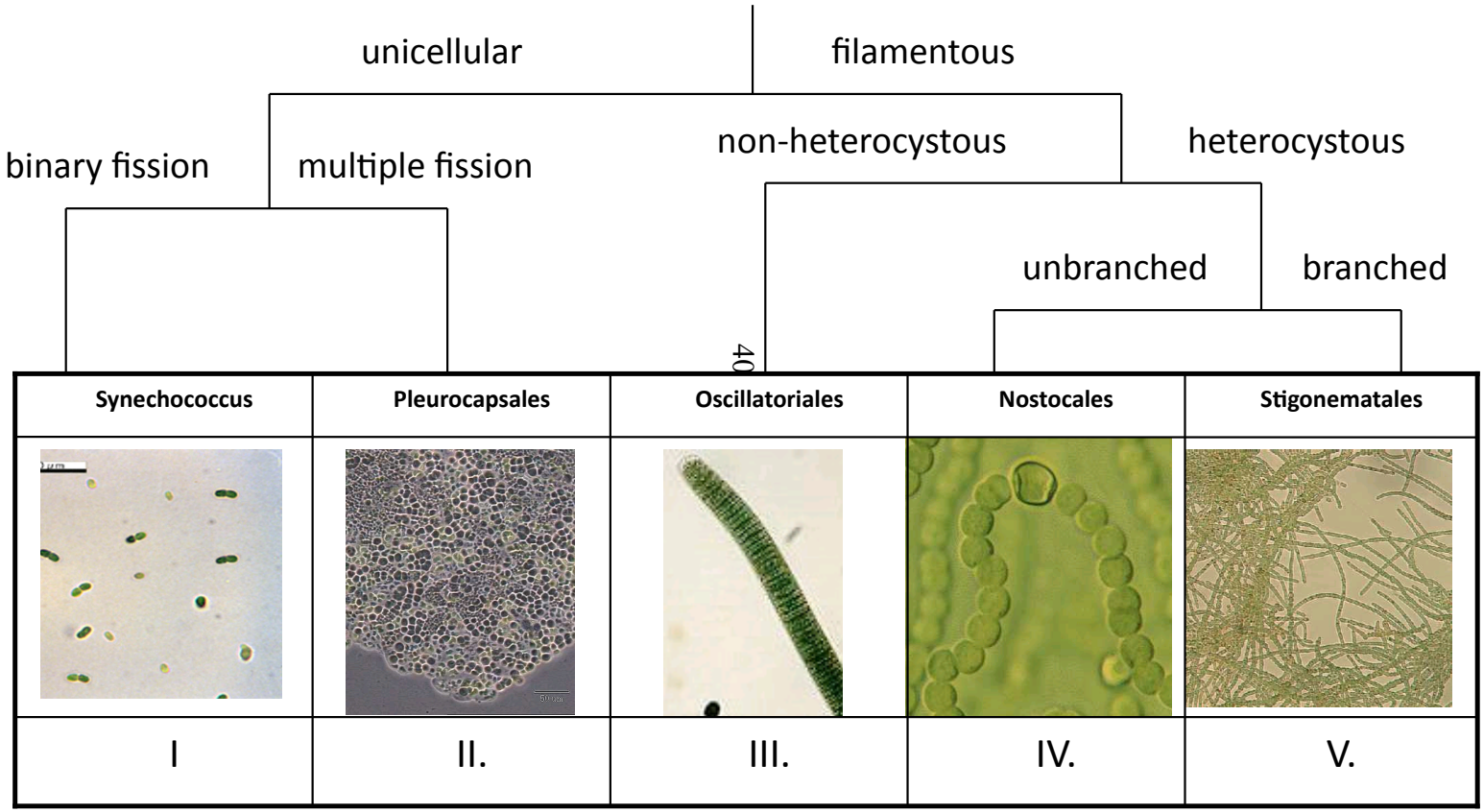


Figure 2.2. Morphological classification scheme of Cyanophyta (Rippka 1988).

2.2. Compounds isolated from *Lyngbya bouillonii*

The species, *Lyngbya bouillonii*, was relatively recently described (Hoffmann and Demoulin 1991). Species identification within the genus *Lyngbya* has been expressed as rudimentary, at best. Together, the relatively young species characterization and taxonomic difficulties in this group have confounded the ability to accurately describe this organism. In the recent past, several compounds were discovered by members of the Richard Moore lab at the University of Hawaii at Manoa, and the producing strains were initially reported as *L. majuscula* or a *Lyngbya* sp . Soon after however, as the process of *L. bouillonii* species identification matured, the validity of the initial identification of the producers were re-evaluated. With regard to the specific samples, VP417a, VP417b and NIH309, R. Thacker and V. Paul addressed the question by amplifying partial 16S rRNA genes (rDNA) from each of the samples and used phylogenetic analyses to confirm that the samples belonged to the genus *Lyngbya* (Thacker and Paul 2004). Furthermore, Williams and Luesch et al. described those and other specific collections as being "cyanobacterium that closely match *L. bouillonii* Hoffmann et Demoulin in its general habitat as well as in its morphology" (Luesch, Yoshida et al. 2002) and are "easily identifiable by the presence of a small shrimp within the algal mat" (Williams, Luesch et al. 2003). Based on the 16S rRNA data and morphological observations it is reasonable to include the compounds in table X as part of the *L. bouillonii* chemical repertoire. For completeness and consistency, I will include the compounds from those samples in the following review.

Between 1996 and 1999 Klein and co-workers pioneered the initial investigations into the natural product composition of *L. bouillonii* species. Sample collection efforts towards a screening program to evaluate cyanobacteria as a source of interesting bioactive metabolites were concentrated to regions in Papua New Guinea (PNG). Extracts from a *L. bouillonii* collection near Laing Island, PNG afforded the first reported compound, laingolide (**27**), a 15-membered macrolide, from the newly described species (Klein, Braekman et al. 1996). No bioactivity data was reported. Later, in 1999, Klein et al reported the structure of two more macrolide derivatives, laingolide A (**28**) and madangolide (**29**), from a collection made in the vicinity of Madang, PNG (Klein, Braekman et al. 1999). In planar structure, **28** is structurally very similar to laingolide, but is deficient in a methyl group at the C-4 position. Similarly, **29** is an analog of laingolide A, but possesses four more carbons (two in the form of olefinic methines) as well as an additional methyl functionality at the α-position in the polyketide region of the molecule. The biosynthesis of this class of compounds arises from the activity of a general hybrid PKS-NRPS-PKS pathway where laingolide, laingolide A, and madangolide each contain mechanistic derivations within the assembly architecture. The tert-butyl functionality at the proposed start the biosynthetic pathway is a hallmark of this family and the apratoxin series (discussed later). Comparatively, laingolide A is deficient in one SAM-dependent C-methyltransferase activity, and madangolide possesses two additional rounds of acetate extension. Also, in madangolide, the biochemical addition of the acetate extension after the NRPS activity yields a

dienamide system in the final product and a SAM-dependent C-methyltransferase acts on the last PKS extension module. No bioactivity of laingolide A or madangolide was reported.

As their screening program for new bioactive chemical entities continued, Klein et al. reported yet another new macrolide isolated from a strain of *L. bouillonii* collected from the region near Laing Island, PNG (Klein, Braekman et al. 1997). Lyngbyaloside (30) is a 16-membered glycosylated macrolide, the second one isolated from a cyanobacterium (Barchi, Moore et al. 1984), but the first one isolated from a marine cyanobacterium. Due to limited sample at the time, no bioactivity assays were performed. Later in 2002, Gerwick, et al. and Luesch et al. each reported on compounds structurally similar to 30. Lyngbouilloside (31) (Tan, Marquez et al. 2002) and lyngbyaloside B (32) (Luesch, Yoshida et al. 2002) represent the second and third macrolide glycosides from a marine cyanobacterium. Lyngbouilloside was isolated from a sample collected near the north coast of New Britain, PNG in 2000. It was reported as moderately cytotoxic to neuro-2a neuroblastoma cells (IC_{50} = 17 μM). The weakly cytotoxic analog, 32, was isolated from a strain collected from the Ulong Channel in Palau. Presumably, the biosynthesis of these compounds is similar for each one in that the linear molecule is primarily PKS derived with the priming of an acetate unit, with 9 (or 10 for lyngbouilloside) subsequent acetate extension units. They differ in the tailoring enzymes that are responsible for the bromination and methylation patterns, and post-translational modifications such as the glycosylation reactions.

The first compound from the lyngbyapeptin family of modified peptides was isolated in 1999, again by Klein et al., from a collection off the coast of Laing Island, PNG. Lyngbyapeptin A (**33**), contains a 3-methoxy-2-butenoyl moiety on the *N*-terminus. Later, in 2002, Luesch et al. discovered two analogs isolated from collections around Palau, lyngbyapeptins B **(34)** and C **(35)** which were reported as noncytotoxic at <5 μM against KB human nasopharyngeal carcinoma and LoVo human colon adenocarcinoma cells (Luesch, Yoshida et al. 2002). In 2003 Williams et al. reported the structurally similar metabolite 15-norlyngbyapeptin A **(36)** from re-collections of a Guamanian strain of *L. bouillonii*. No bioactivity was reported (Williams, Luesch et al. 2003). The biosyntheisis of these ketopeptides (Tidgewell, Clark et al. 2010) begins with PKS-type loading and extension modules creating a diketide precursor. C-2 undergoes deprotonation forming the α-β unsaturation with concomittent nucleophilic attack of the adjacent oxygen to the methyl group of SAM. The starter unit for lyngbyapeptin C presumably undergoes C-methylation by SAM as well. The biosyntheis continues with nonribosomal peptide synthetase activity extending the nascent compound by five amino acid moieties. Lyngbyapeptin A differs from norlyngbyapeptin A in that it possesses an *N*, *O*-dimethyl tyrosine in the first amino acid position whereas **36** possesses an *N*-methyl tyrosine. The NRPS region of lyngbyapeptins B and C are similar, but differ from **34** in amino acid positions 2, 3 and 4. Lyngbyapeptins B and C incorporate *N*-methyl valine, *N*, *O*-dimethyl tyrosine and alanine whereas lyngbyapeptin A incorporates an *N*-methyl leucine, *N*-methyl isoleucine, and proline respectively.

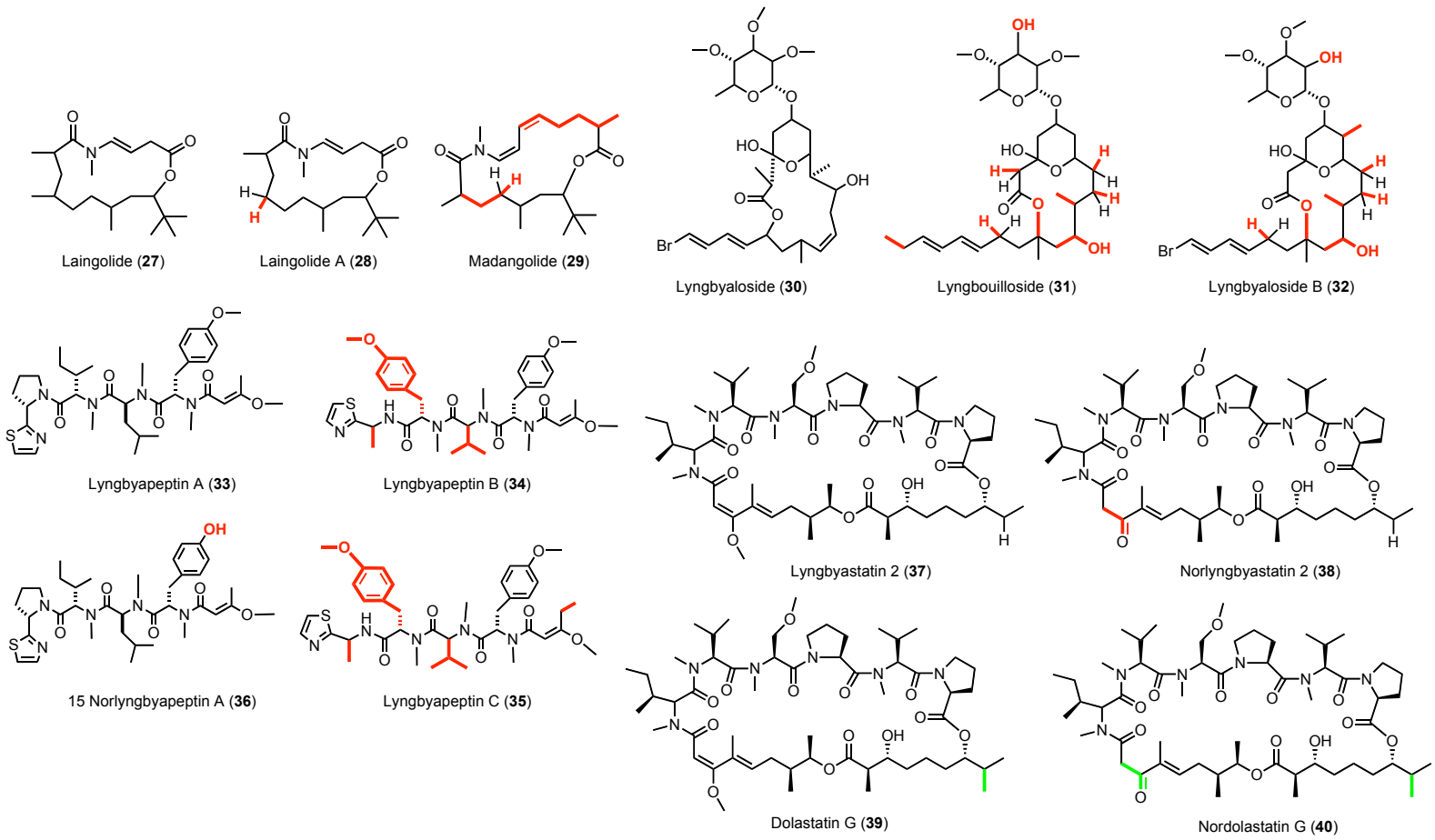


Figure 2.3. Reported compounds likely produced from *Lyngbya bouillonii* organized by compound family. Red moieties mark structural differences from the parent (first isolated) natural product within each family. The green moieties mark structural differences from the parent natural product, but were isolated from a different organism.

The cyclic depsipeptides lyngbyastatin 2 **(37)** and norlyngbyastatin 2 **(38)** were isolated from a Guamanian *Lyngbya* sp. (Luesch, Yoshida et al. 1999). As noted by Luesch and co-workers, these compounds bear a striking resemblance to the previously isolated dolastatin G **(39)** and nordolastatin G **(40)** from the Japanese sea hare *Dolabella auricularia* (Mutou, Kondo et al. 1996) differing only in an isopropyl instead of an ethyl functionality in the terminal polyketide section. Lyngbyastatin 2 was found to be toxic to mice at 3 mg/kg (LD_{100}). At sublethal doses, it was inactive in vivo against the C38 murine adenocarcinoma cell line.

From a biosynthetic perspective, the polyketide portions of lyngbyastatin 2 and norlyngbyastatin 2 have some interesting characteristics. To start, the nonanoic acid moiety is modified with varying levels of ketide reduction. Then, it can be predicted that the first of two SAM-mediated C-2 methylations occurs on the starter unit, and the next is on the third acetate extension. The central region of the polyketide section is extended by incorporation of an α-hydroxy acid by an NRPS type system to form the ester bond, likely recruited as the α-keto acid (Ramaswamy, Sorrels et al. 2007). Subsequent C-1 methylation via an HMG-CoA synthase-like (HCS) mechanism occurs at what was the carboxyl of this α-hydroxy acid. To continue the growing compound, three more PKS modules extend the chain by six carbon atoms, again with varying degrees of reduction and methylation. Six NRPS modules then perform the addition of *N*-methyl isoleucine, *N*-methyl valine, *N*, *O*-dimethyl serine, proline, *N*-methyl valine and proline respectively. Finally, the compound is concomittently off-loaded and cyclized, presumably by a type I thioesterase (TE) domain.

Lyngbyastatin 2 and norlyngbyastatin 2 belong to a family of cyclodepsipeptide compounds that include lyngbyastatins 1, and 3 through 7. However, *L. bouillonii* has not yet been identified as a candidate producer of these additional lyngbyastatin compounds (Harrigan, Yoshida et al. 1998; Bai, Bates et al. 2002; Williams, Moore et al. 2003; Matthew, Ross et al. 2007; Taori, Matthew et al. 2007).

The fifth class of compounds is the peptolides. Lyngbyabellin A **(41)** was first reported by Luesch et al. in 2000 as exhibiting moderate activity against KB and LoVo cells, with IC_{50} values of 0.03 μg/mL and 0.50 μg/mL respectively. Additionally, *in vitro* trials revealed that **41** is toxic to mice (with sublethal doses of 1.2-1.5 mg/kg). No antitumor activity was reported against the C38 murine colon adenocarcinoma, or the M16 mammary adenocarcinoma cell lines. Further mechanism of action studies revealed that the cytoskeleton disrupting effects of **41** are specific to microfilaments, with no disruption to microtubules (Luesch, Yoshida et al. 2000).

Soon after, lyngbyabellin B **(42)** was reported in 2000 by two different research teams from samples collected from two different hemispheres of the earth. With collections from a reef-inhabiting strain from Apra Harbor, Guam, Luesch et al described the metabolite as slightly less cytotoxic *in vitro* as **41**. It exhibited IC_{50} values of 0.10 and 0.83 μg/mL against KB and LoVo cells respectively. Mode of action studies were not carried out. In parallel, Gerwick et al. collected a strain of *Lyngbya* near the Dry Tortugas National Park, Florida, USA. After extraction and vacuum liquid chromatography (VLC) of the organic phase, **42** was isolated. It displayed toxicity

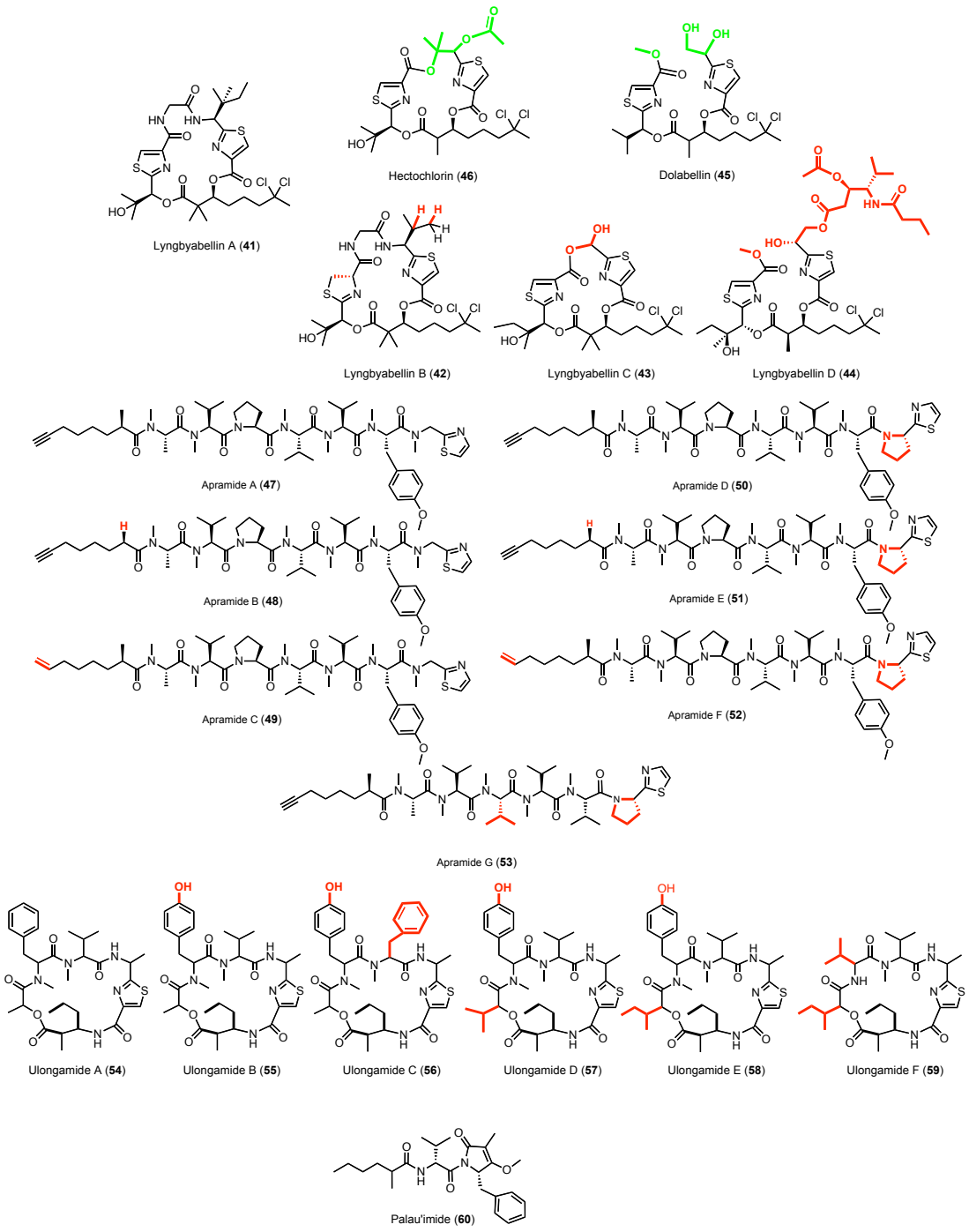


Figure 2.4. Reported compounds likely produced from *Lyngbya bouillonii* organized by compound family. Red moieties mark structural differences from the parent (first isolated) natural product within each family. The green moieties mark structural differences from the parent natural product, but were isolated from a different organism.

against brine shrimp with a LC_{50} value of 3.3 ppm and inhibited growth of *Candida albicans* (ATCC 14053) (Gerwick, Mrozek et al. 1989).

Lyngbyabellin C **(43)** was isolated from a *L. bouillonii* species collected near Short Dropoff, Palau and reported in 2002 by Luesch et al. It exhibited weak cytotoxicity against KB and LoVo cells with IC_{50} values of 2.1 μM and 5.3 μM, respectively.

Lyngbyabellin D **(44)**, reported by Williams et. al in 2003, is an acyclic analog of this family. To consider the biosynthesis of this class of compounds, one must note the striking structural similarity to two other related compounds, dolabellin **(45)** (Sone, Kondo et al. 1995) and hectochlorin **(46)** (Marquez, Watts et al. 2002). Dolabellin was isolated from *Dolabella auricularia* and hectochlorin was isolated in 2002 from a cultured strain of *Lyngbya majuscula* collected in Hector Bay, Jamaica. Other cyclic and acyclic analogs of this class of compounds (lyngbyabellin E through G) were also isolated from a *L. majuscula* strain collected from Alotau Bay, PNG in 2002 (Han, PhD dissertation, 2005). Recently, the biosynthetic gene cluster for **46** was described (Ramaswamy, Sorrels et al. 2007). Lyngbyabellin biosynthesis is envisioned as being similar to hectochlorin and dolabellin production. The beginning portion of the compound contains an adapted octanoic acid unit. The molecule continues with two consecutive additions of the modified isovaleric acid and modified cysteine extender units. Final hydrolytic cyclization by a thioesterase leaves the mature peptolide. In the case of **44** and **45**, ring opening and modification of the second isovaleric acid unit occurs. Futher, **44**

undergoes extension of the free hydroxyl with a modified heptanoic acid unit. The timing of this latter addition is unknown, but presumably occurs after the modular construction steps and ring opening.

The next seven compounds belong to a family of lipopeptides that were reported together in 2000 by Luesch and coworkers. Apramides A through F (**47 – 53)** are ketolides, starting with a PKS region and continuing with eight NRPS incorporated amino acid moieties. The biosynthetically interesting functionality in this series of molecules is the terminal alkyne in A,B,D, E and G. Apramides C and F bear a terminal olefin instead. This terminating functionality is also seen in jamaicamides B and C. The mechanistic biochemistry of the formation of these functionalities is not yet understood, but presumably uses a motif related to a fatty acid desaturase (Edwards, Marquez et al. 2004). The bioactivity of apramides A through F was not reported.

In 2002, a six-member family of β-amino acid containing cyclic depsipeptides were isolated from a Palauan strain of *L. bouillonii* and reported by Luesch, et al (Luesch, Williams et al. 2002). Biosynthesis of ulongamides A through F (**54 – 59**) presumably starts by a PKS-type initiation and extension to yield a hexanoic acid precursor that is acted upon by a SAM methyltransferase and an amino transferase. Further extension occurs by the addition of 2-hydroxyisovaleric acid and four modified proteinogenic amino acids with final thioesterase mediated hydrolytic cleavage predicted to form the macrocyclic product. The particular macrolactam functionality that is presumably created by an existing TE is rare in

marine cyanobacterial secondary metabolites and to bacterial natural products in general. To my knowledge, the one other example that contains a fatty acid modifying amino transferase to create a β-amino fatty acid is mycosubtilin, a potent antifungal lipopeptide isolated from the *Bacillus subtillis* strain ATCC6633 (Duitman, Hamoen et al. 1999).

The final molecular class to discuss (excluding the apratoxin series) is palau'imide (**60**). It exhibited mild bioactivity with IC_{50} values for *in vitro* cytotoxicity of 1.4 and 0.36 μM against KB and LoVo cells, respectively (Luesch, Yoshida et al. 2002). Although it is structurally unrelated to any other compound isolated from this species to date, its biosynthesis is very interesting. A 2-methylhexanoic acid moiety is extended by valine and then tyrosine. Next, ketide extension incorporates an acetate unit which is SAM methylated at C-2. This C-2 presumably undergoes deprotonation forming the α-β unsaturated double bond with concomittent nucleophilic attack by the adjacent oxygen to a second methyl group of SAM. This intermediate is subsequently cyclized to form a pyrrolinone ring. From a biosynthetic perspective, the final pyrrolinone ring formation is comparable to the related ring functionality in the jamaicamide series of compounds produced by *Lyngyba majuscula.* However, the penultimate monomer of jamaicamides A-C is not tyrosine, but glycine. Mechanistically, it is envisioned that hydrolytic cyclization occurrs via nucleophilic attack of the amide nitrogen to the thioester carbon (Edwards, Marquez et al. 2004).

This review introduces and describes 30 compounds reported between 1996 and

2003 that were isolated from various collections of *Lyngbya* sp. with morphological characteristics consistent with a *L. bouillonii* strain. These cyanobacteria were obtained from several collection locations; Guam, Palau, and Papua New Guinea. This organism has produced molecules belonging to six different classes of compounds; macrolides, glycosylated macrolides, modified peptides, peptolides, lipopeptides and cyclic depsipeptides, thus revealing it's impressive genetic capacity to make structurally diverse bioactive chemical entities. The biosynthetic routes suggested above span polyketides, simple and complex ketopeptides (polyketides transitioning to peptides), nonribosomal peptides, hybrid variations and modifications within.

2.3 The Apratoxins

In this section, I will give a review of the discovery, isolation, syntheses and bioactivities of a family of cyclodepsipeptide compounds known as the apratoxins. This family of natural products includes apratoxin A (**61**), and its analogs B through G (**62** – **67**). (Gutierrez, Suyama et al. 2008; Matthew, Schupp et al. 2008) and (Tidgewell, unpublished data). This series of compounds has been isolated from marine cyanobacteria inhabiting coral reefs surrounding the islands of Guam (Matthew, Schupp et al. 2008), Palau (Luesch, Yoshida et al. 2002; Luesch, Chanda et al. 2006) and Papua New Guinea (Gutierrez, Suyama et al. 2008). They are classified as complex ketopeptides or cyclodepsipeptides of mixed biosynthetic origin possessing both polyketide and peptide segments.

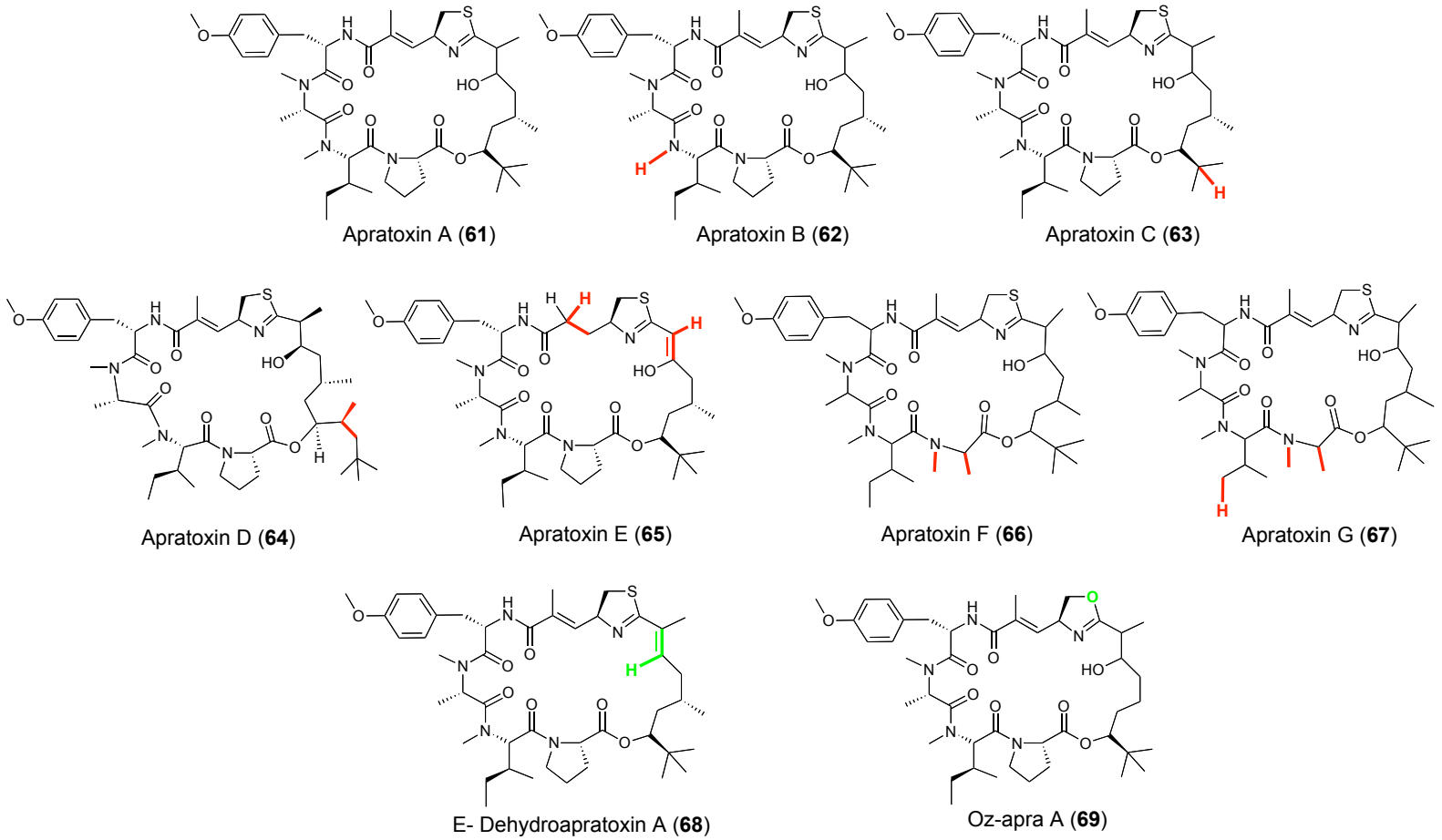


Figure 2.5 Reported compounds likely produced from *Lyngbya bouillonii* organized by compound family. Red moieties mark structural differences from the parent (first isolated) natural product within each family. The green moieties mark structural differences from the parent natural product, but are synthetically derived analogs.

Table 2.1. All reported compounds likely produced by *L. bouillonii* ordered by date of publication.

Year Reported	Compound Name	Compound Class	Organism reported	Collection ID
(Klein, Braekman et al. 1996)	laingolide	macrolide	*Lyngbya bouillonii*	NR
(Klein, Braekman et al. 1997)	lyngbyaloside	glycosylated macrolide	*Lyngbya bouillonii*	NR
(Klein, Braekman et al. 1999)	laingolide A	macrolide	*Lyngbya bouillonii*	NR
(Klein, Braekman et al. 1999)	madangolide	macrolide	*Lyngbya bouillonii*	NR
(Klein, Braekman et al. 1999)/ (Luesch, Yoshida et al. 2000)	lyngbyapeptin A	modified peptide	*Lyngbya bouillonii*/ [a]*Lyngbya majuscula*	NR/ VP417
(Luesch, Yoshida et al. 1999)	lyngbyastatin 2	cyclic depsipeptide	[a]*Lyngbya majuscula*	VP417, VP503
(Luesch, Yoshida et al. 1999)	norlyngbyastatin 2	cyclic depsipeptide	[a]*Lyngbya majuscula*	VP417, VP503
(Luesch, Yoshida et al. (Luesch, Yoshida et al. 2000)	lyngbyabellin A	peptolide	[a]*Lyngbya majuscula*	VP417
(Luesch, Yoshida et al. 2000)	lyngbyabellin B	peptolide	[a]*Lyngbya majuscula*	VP417
(Luesch, Yoshida et al. 2000)	apramide A	lipopeptide	[a]*Lyngbya majuscula*	VP503
(Luesch, Yoshida et al. 2000)	apramide B	lipopeptide	[a]*Lyngbya majuscula*	VP503
(Luesch, Yoshida et al. 2000)	apramide C	lipopeptide	[a]*Lyngbya majuscula*	VP503
(Luesch, Yoshida et al. 2000)	apramide D	lipopeptide	[a]*Lyngbya majuscula*	VP503
(Luesch, Yoshida et al. 2000)	apramide E	lipopeptide	[a]*Lyngbya majuscula*	VP503
(Luesch, Yoshida et al. 2000)	apramide F	lipopeptide	[a]*Lyngbya majuscula*	VP503
(Luesch, Yoshida et al. 2000)	apramide G	lipopeptide	[a]*Lyngbya majuscula*	VP417, VP503
(Luesch, Yoshida et al. 2001)	apratoxin A	cyclic depsipeptide	[a]*Lyngbya majuscula*	VP312, VP496, VP503, VP417

[a] samples determined 99.8% similar to *L. bouillonii* by partial 16S rDNA analyses (Thacker and Paul 2004)

[b] morphology and general habitat similar to *L. bouillonii* (Williams, Luesch et al. 2003)

Table 2.1 continued.

Year reported	Compound Name	Compound Class	Organism	Collection ID
(Luesch, Yoshida et al. 2002)	apratoxin B	cyclic depsipeptide	[a,b]*Lyngbya sp.*	VP417b
(Luesch, Yoshida et al. 2002)	apratoxin C	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Yoshida et al. 2002)	lyngbyabellin C	peptolide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Yoshida et al. 2002)	lyngbyapeptin B	modified peptide	[b]*Lyngbya sp.*	Ulong Channel
(Luesch, Yoshida et al. 2002)	lyngbyapeptin C	modified peptide	[b]*Lyngbya sp.*	mixed collection
(Luesch, Yoshida et al. 2002)	palau'imide	ketopeptide	[b]*Lyngbya sp.*	mixed collection
(Luesch, Yoshida et al. 2002)	lyngbyaloside B	glycosylated macrolide	[a,b]*Lyngbya sp.*	NIH309
(Tan, Marquez et al. 2002)	lyngbouilloside	glycosylated macrolide	*Lyngbya bouillonii*	PNGE9 5/Dec/99-1
(Luesch, Williams et al. 2002)	ulongamide A	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Williams et al. 2002)	ulongamide B	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Williams et al. 2002)	ulongamide C	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Williams et al. 2002)	ulongamide D	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Williams et al. 2002)	ulongamide E	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Luesch, Williams et(Luesch, Williams et al. 2002)	ulongamide F	cyclic depsipeptide	[a,b]*Lyngbya sp.*	NIH309
(Williams, Luesch et al. 2003)	15-norlyngbyapeptin A	modified peptide	[a,b]*Lyngbya sp.*	VP417
(Williams, Luesch et al. 2003)	lyngbyabellin D	peptolide	[a,b]*Lyngbya sp.*	VP417
(Matthew, Schupp et al. 2008)	apratoxin E	cyclic depsipeptide	*Lyngbya bouillonii*	PS372
in prep	apratoxin F	cyclic depsipeptide	*Lyngbya bouillonii*	
in prep	apratoxin G	cyclic depsipeptide	*Lyngbya bouillonii*	
[a](samples determined 99.8% similar to *Lyngbya sp.* by partial 16S rDNA analyses (Thacker and Paul 2004)				
[b](morphology and general habitat similar to *L. bouillonii* (Williams, Luesch et al. 2003)				

Members of this compound family have shown potent *in vitro* cytotoxicity against the KB (0.52-21.3 nM) and LoVo cell lines (0.36-10.8 nM). Efforts to elucidate the mechanism of action of apratoxin A revealed that this compound induces G1-phase cell-cycle arrest and apoptosis. Three total syntheses of **61** have been reported by the groups of Forsyth (Chen and Forsyth 2003; Chen and Forsyth 2003; Chen and Forsyth 2004), Takahashi (Doi, Numajiri et al. 2006), and Liu (Ma, Zou et al. 2006). From the latter effort four oxazoline analogues of apratoxin A were synthesized and it was found that replacement of the thiazoline ring with an oxazoline ring had only a marginal effect on potency. Interestingly, Shen et al., found that one of the oxazoline analogues of apratoxin A, oz-apraA (**69**) utilizes a new anticancer strategy by promoting the degradation of heat shock protein (Hsp) 90 clients through chaperone-mediated autophagy (Shen, Zhang et al. 2009). Apratoxin A likely inhibits Hsp 90 function by stabilizing the interaction with Hsc70/Hsp70. Other recent mode of action studies have revealed that **61** reversibly inhibits the secretory pathway for several cancer-associated receptors by interfering with their cotranslational translocation (Liu, Law et al. 2009).

2.4 Results and Discussion

2.4.1 Collection, Extraction, and Fractionation of *L. bouillonii* for Apratoxin A Isolation

As part of our drug discovery program first at Oregon State University, and now more recently at SIO, our group has made several trips to Papua New Guinea to collect marine cyanobacteria. Collections of *L. bouillonii* in 1999 and 2000 yielded the isolation of apratoxin A (Lik Tan, PhD dissertation 2001). Our bioassay guided screening of these organic extracts indicated cytoxic activity towards neuro2a mouse neuroblastoma and H-460 cells. These interesting results prompted a concerted effort to isolate more of this natural product for biological testing. In April 2006 a collection trip to Papua New Guinea targeting the apratoxin producer yielded a 3.0 liter composite collection of *Lyngbya bouillonii* preserved in EtOH. The location, GPS coordinates, depth and date of each sample collected is shown in Table 2.2. The large-scale chemistry samples were extracted and fractionated for bioassay and chemical profiling and for apratoxin A isolation (Figures 2.6 and 2.7). Our in-house human lung cancer cell (NIH-H460) screening indicated that the organic fractions E, F, G, H and I for both 1597-1 and -2 (Figure 2.8) were highly cytotoxic activity with IC_{50} values of less than 5 ug/mL. LC-MS analysis of each fraction indicated the presence of apratoxin A in varying amounts in each of the fractions E – I. Further purification of 1597-1 fraction H and -2 fraction H yielded approximately 23.0 and 6.0 mg of apratoxin A respectively (Figures 2.6 and 2.7).

Table 2.2. Date, depth and location of each *L. bouillonii* sample that was added to the 3.0 liter composite collection. NR = not recorded.

Location	GPS	Depth	Date Collected
Kwato Wharf	10° 36.990 S 150° 37.976 E	20-40 ft	4/20/06
Grant Rock	10° 32.763 S 151° 02.510 E	NR	4/21/06
Deacon's Reef	10° 15.612 S 150° 44.878 E	~50 ft	4/23/06
Bentley Bay	10° 13.051 S 150° 36.110 E	NR	4/25/06
Kape Point	10° 14.055 S 150° 49.873 E	30-60 ft	4/26/06
Sewa Bay	10° 10.980 S 150° 58.624 E	NR	4/27/06
Gallons Reef	10° 14.00 S 151° 0.00 E	20-40 ft	4/28/06

Compound isolation efforts from this collection yielded sufficient natural product for further bioactivity testing. Together with our collaborator Dr. Fred Valeriote at the Josephine Ford Cancer Center in Detroit, Michigan, we showed that apratoxin A had a unique and medically relevant profile of cancer cell toxicity. In the *in vitro* disk diffusion assay (or zone of inhibition assay) where a difference in the zones between solid tumor cells and leukemia cells of 250 units or greater defines a sold tumor selective compound (Valeriote, Grieshaber et al. 2002), apratoxin A was found to be both potent and human solid tumor selective. Further pharmacokinetic studies in mouse models are currently being pursued.

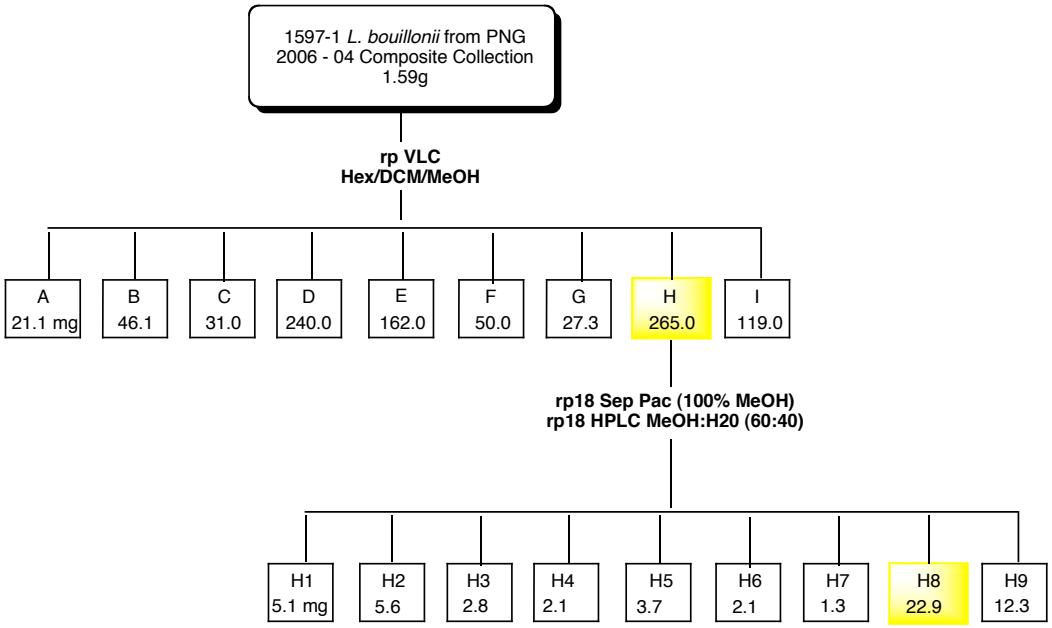


Figure 2.6. Bioassay guided fractionation scheme of sample1597-1. Highlighted box indicates the presence of apratoxin A, as detected by low resolution LC-MS.

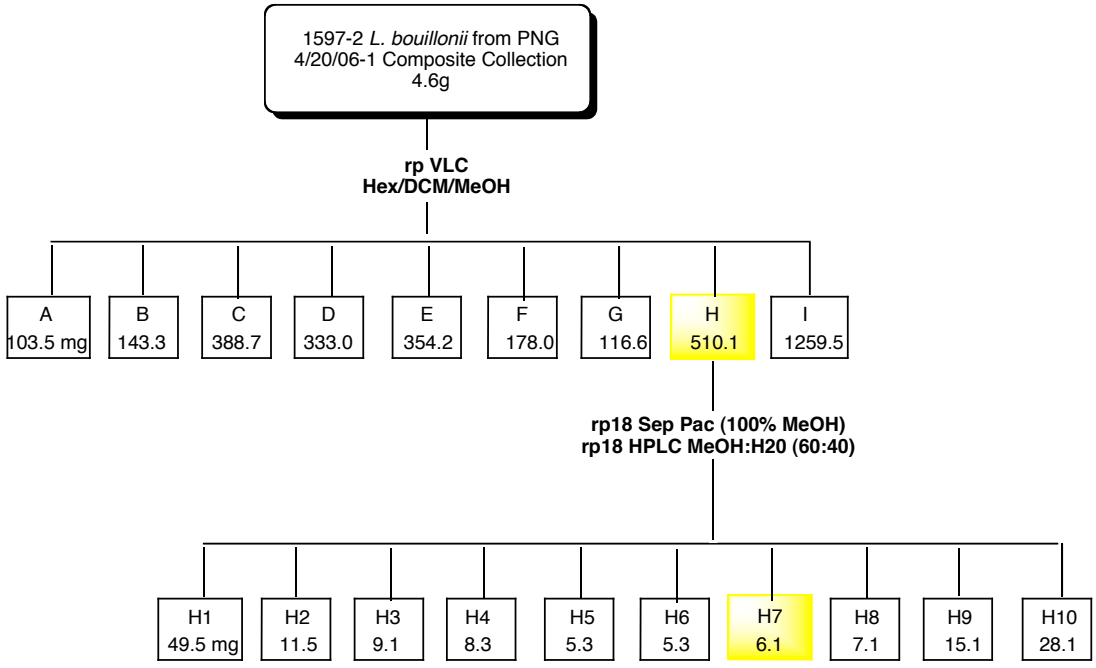


Figure 2.7. Bioassay guided fractionation scheme of sample1597-2. Highlighted box indicates the presence of apratoxin A, as detected by low resolution LC-MS.

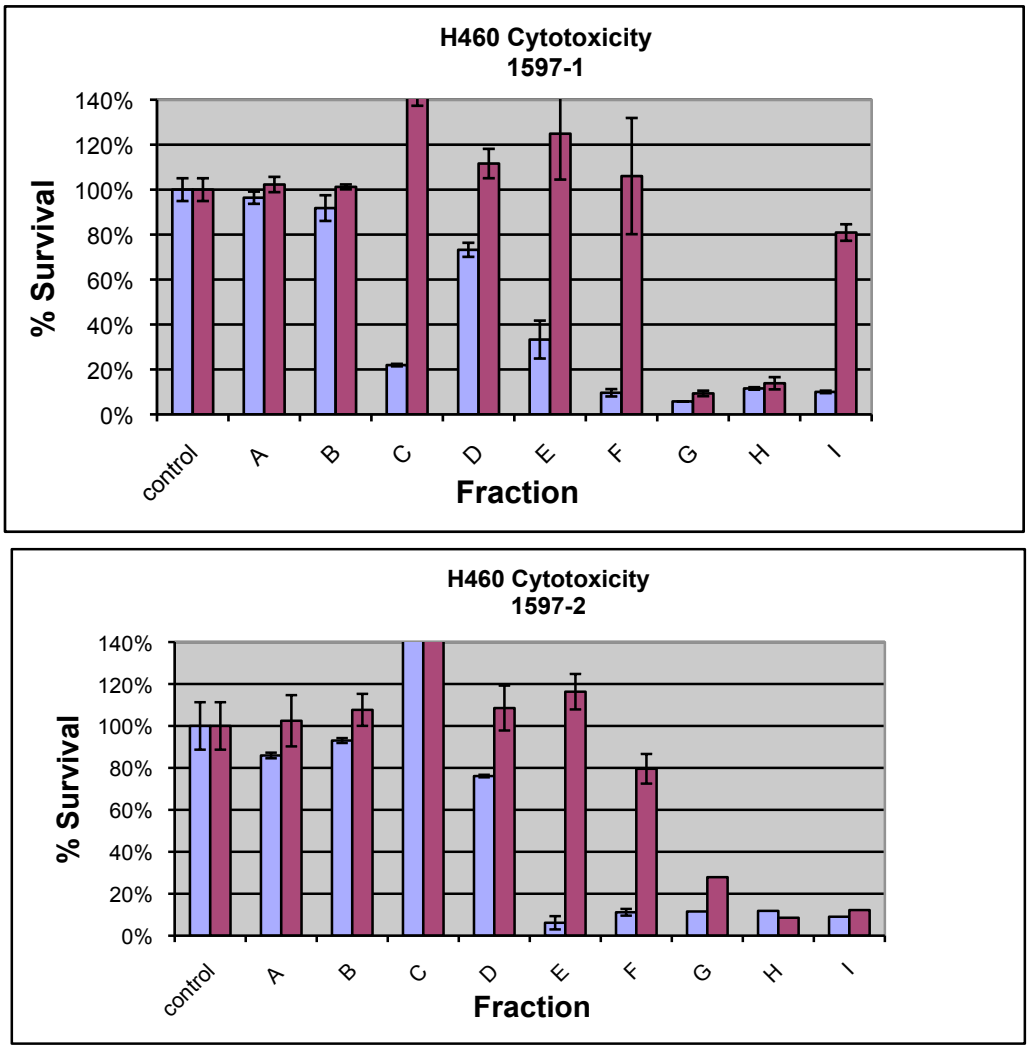


Figure 2.8. Cytotoxicity results of each fraction from the parent extracts 1597-1 (top) and 1597-2 (bottom) against the NIH-H460 cancer cell line. At 50 μg/mL (purple bars), fractions 1597-1 C, E-I and 1597-2 E-I exhibited significant cytotoxicity. At 5 μg/mL (red bars), fractions 1597-1 G, H and 1597-2 G-I exhibited significant cytotoxicity.

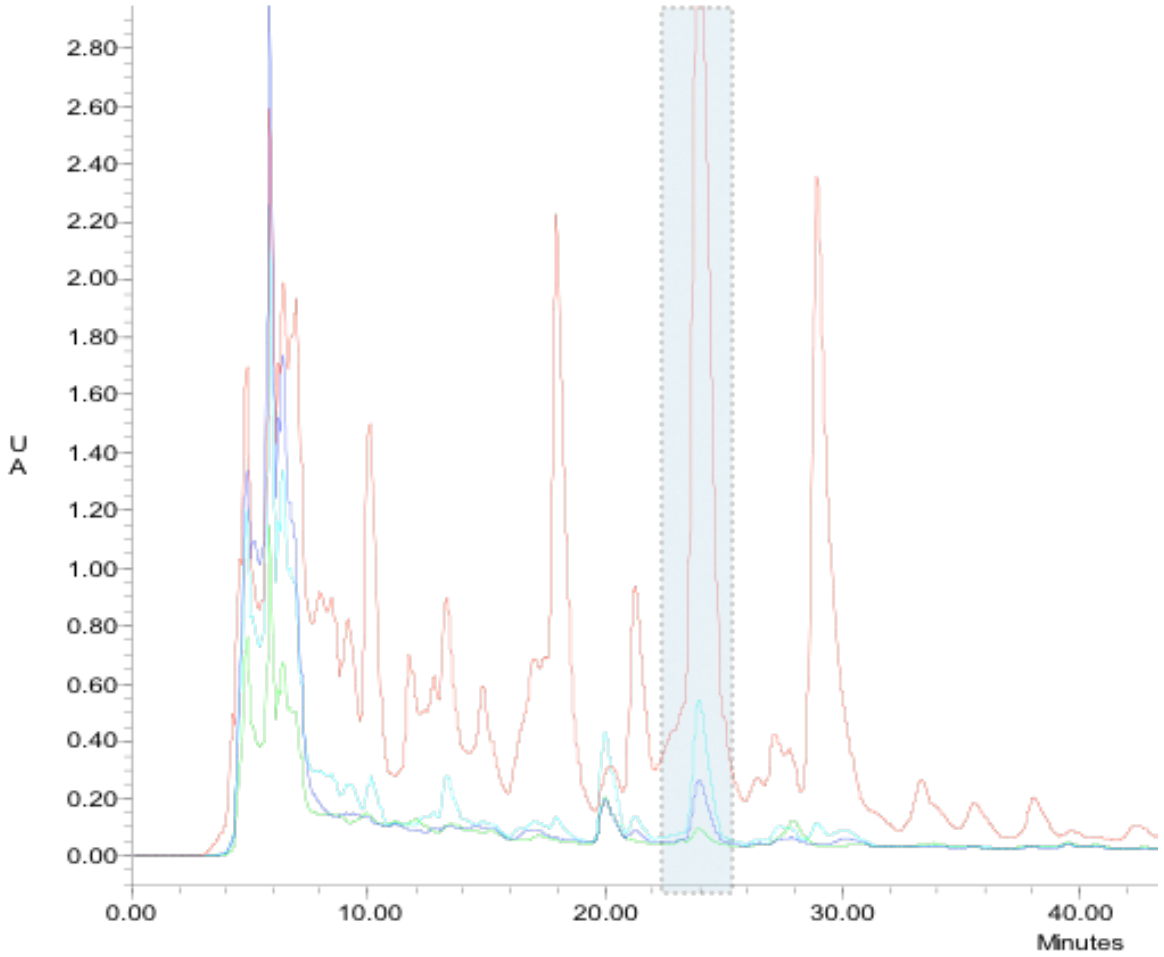


Figure 2.9. HPLC trace of fraction 1597-1H. Apratoxin A (blue box) has a signature retention time at approximately 25.00 minutes.

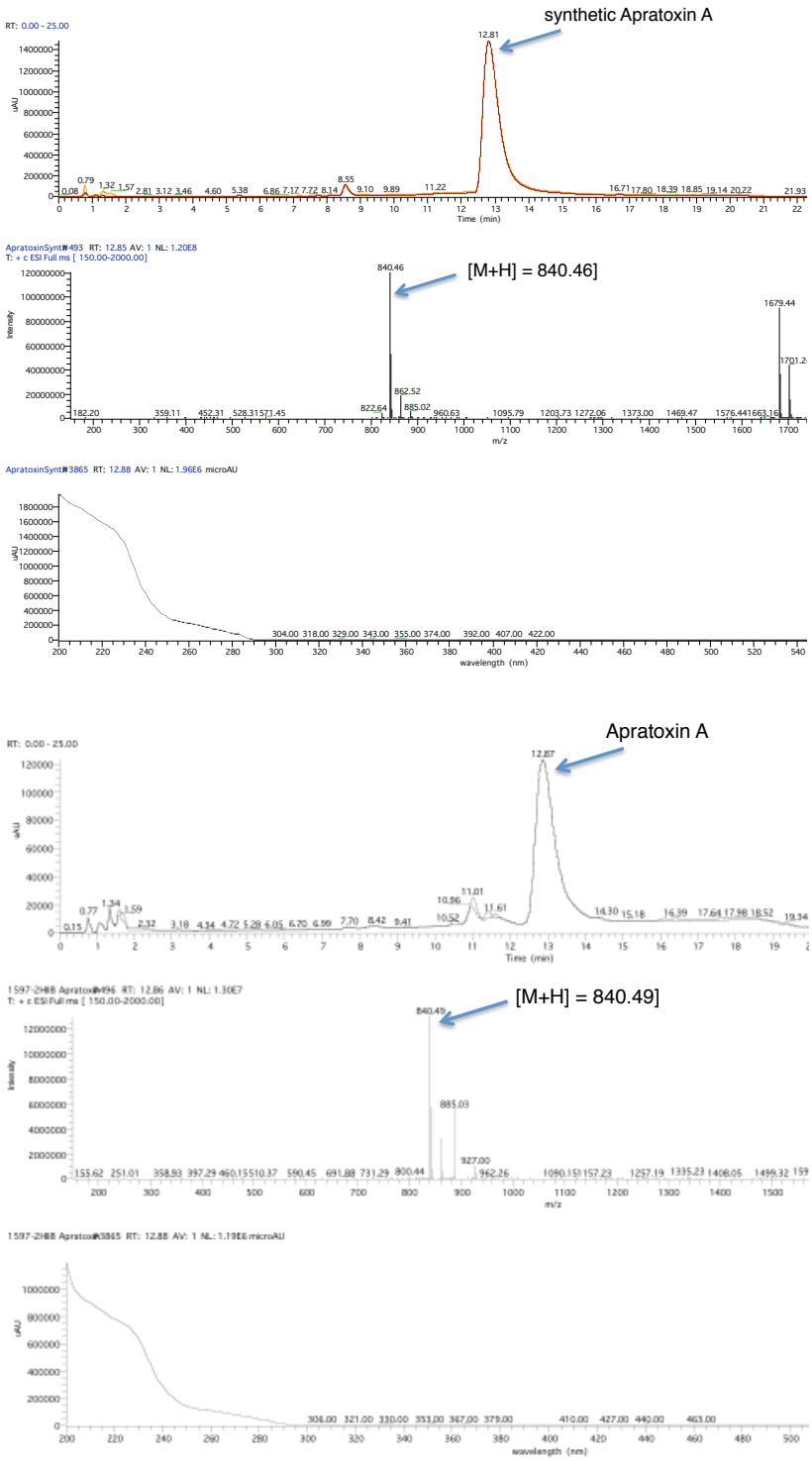


Figure 2.10. Low resolution LC-MS trace of synthetic (top) and natural (bottom) apratoxin A. It exhibits distinct properties under specific LC conditions (see experimental). The retention time is approximately 12.80 minutes (top panel), with an $[M+H]^+$ peak at *m/z* 840.5 (middle panel) and exhibits a UV absorbance at 220 nm (bottom panel).

2.5 EXPERIMENTAL

General Experimental Procedures. Low-resolution ESI-MS data were acquired on a Finnigan LCQAdvantage Max mass spectrometer. Purification of the compounds were carried out on a Waters HPLC system equipped with a Waters 515 binary pump, a Waters 996 PDA detector, and a Phenomenex Jupiter C_{18} column (10 μm, 10 X 250 mm).

Biological Material. The 3.0-liter composite collection of *Lyngbya bouillonii* was made in various locations in Papua New Guinea, April, 2006 and given the collection number PNG/4/20/06-1. Samples of the cyanobacterium were hand collected at depths of 20 – 60 ft using SCUBA. Upon collection, the pooled marine cyanobacterium was preserved in 50 % ethanol and seawater and stored at -20° C until work-up.

Extraction and Fractionation of *L. bouillonii*. A 3.0-liter collection of *L. bouillonii* was thawed and split into 3 X 1 liter samples and assigned the extract numbers 1597-1, 1597-2, and 1597-3. Each 1 liter sample was extracted with 2:1 CH_2Cl_2/MeOH three times and concentrated to dryness *in vacuo* to give 1.59 g, 4.4 g, and 6 g of crude extract, respectively. VLC pre-fractionations of the crude extracts using a gradient with 0-100% EtOAc in hexanes followed by 0-100% of MeOH in EtOAc yielded nine fractions (A-I).

Biological Activity. Cytotoxicity was measured in NCI-H460 human lung tumor cells with cell viability being determined by MTT reduction. Cells were seeded in 96-well plates at 6000 cells/well in 180 μL of medium. After 24 h, the test chemicals were dissolved in DMSO and diluted into medium without fetal bovine serum and then added at 20 μg/well. DMSO was less than 0.5% of the final concentration. After 48 h, the medium was removed and cell viability determined.

2.6 References

Bai, R., R. B. Bates, et al. (2002). "Lyngbyastatin 1 and Ibu-epilyngbyastatin 1: synthesis, stereochemistry, and NMR line broadening." J Nat Prod **65**(12): 1824-9.

Barchi, J. J., R. E. Moore, et al. (1984). "Acutiphycin and 20,21-didehydroacutiphycin, new antineoplastic agents from the cyanophyte *Oscillatoria acutissima*." Journal of the American Chemical Society **106**(26): 8193-8197.

Beerling, D. J. (2009). "Coevolution of photosynthetic organisms and the environment." Geobiology **7**(2): 97-9.

Boyer, S. L., V. R. Flechtner, et al. (2001). "Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria." Mol Biol Evol **18**(6): 1057-69.

Brun, Y. V. and L. J. Shimkets (2000). Prokaryotic Development, ASM Press.

Chen, J. and C. J. Forsyth (2003). "Synthesis of the apratoxin 2,4-disubstituted thiazoline via an intramolecular aza-Wittig reaction." Org Lett **5**(8): 1281-3.

Chen, J. and C. J. Forsyth (2003). "Total synthesis of apratoxin A." J Am Chem Soc **125**(29): 8734-5.

Chen, J. and C. J. Forsyth (2004). "Total synthesis of the marine cyanobacterial cyclodepsipeptide apratoxin A." Proc Natl Acad Sci U S A **101**(33): 12067-72.

Doi, T., Y. Numajiri, et al. (2006). "Total synthesis of apratoxin A." Org Lett **8**(3): 531-4.

Duitman, E. H., L. W. Hamoen, et al. (1999). "The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase." Proc Natl Acad Sci U S A **96**(23): 13294-9.

Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*." Chem Biol **11**(6): 817-33.

Farquhar, J., H. Bao, et al. (2000). "Atmospheric influence of earth's earliest sulfur cycle." Science **289**(5480): 756-758.

Gerwick, W. H., C. Mrozek, et al. (1989). "Novel cytotoxic peptides from the tropical marine cyanobacterium *Hormothamnion enteromorphoides*. 1. Discovery, isolation and initial chemical and biological characterization of the hormothamnins from wild and cultured material." Experientia **45**(2): 115-21.

Gutierrez, M., T. L. Suyama, et al. (2008). "Apratoxin D, a potent cytotoxic cyclodepsipeptide from papua new guinea collections of the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida*." J Nat Prod **71**(6): 1099-103.

Harrigan, G. G., W. Y. Yoshida, et al. (1998). "Isolation, structure determination, and biological activity of dolastatin 12 and lyngbyastatin 1 from *Lyngbya majuscula*/*Schizothrix calcicola* cyanobacterial assemblages." J Nat Prod **61**(10): 1221-5.

Hoffmann, L. and V. Demoulin (1991). "Marine Cyanophyceae of Papua New Guinea. II. *Lyngbya bouillonii* sp. nov., a remarkable tropical reef-inhabiting blue-green alga." Belg. J. Bot. **124**: 82-88.

Holt, J. G., N. R. Krieg, et al. (1994). Bergey's Manual of Determinative Bacteriology. Baltimore, Md., Williams and Wilkins.

Kasting, J. F. and J. L. Siefert (2002). "Life and the evolution of earth's atmosphere." Science **296**(5570): 1066-1068.

Klein, D., J.-C. Braekman, et al. (1999). "Lyngbyapeptin A, a modified tetrapeptide from *Lyngbya bouillonii* (Cyanophyceae)." Tetrahedron Letters **40**(4): 695-696.

Klein, D., J.-C. Braekman, et al. (1996). "Laingolide, a novel 15-membered macrolide from *Lyngbya bouillonii* (cyanophyceae)." Tetrahedron Letters **37**(42): 7519-7520.

Klein, D., J. C. Braekman, et al. (1999). "Madangolide and laingolide A, two novel macrolides from *Lyngbya bouillonii* (Cyanobacteria)." Journal of Natural Products **62**(6): 934-936.

Klein, D., J. C. Braekman, et al. (1997). "Lyngbyaloside, a novel 2,3,4-tri-*O*-methyl-6-deoxy-alpha-mannopyranoside macrolide from *Lyngbya bouillonii* (Cyanobacteria)." Journal of Natural Products **60**(10): 1057-1059.

Liu, Y., B. K. Law, et al. (2009). "Apratoxin A reversibly inhibits the secretory pathway by preventing cotranslational translocation." Mol Pharmacol.

Luesch, H., S. K. Chanda, et al. (2006). "A functional genomics approach to the mode of action of apratoxin A." Nat Chem Biol **2**(3): 158-67.

Luesch, H., P. G. Williams, et al. (2002). "Ulongamides A-F, new beta-amino acid-containing cyclodepsipeptides from Palauan collections of the marine cyanobacterium *Lyngbya* sp." J Nat Prod **65**(7): 996-1000.

Luesch, H., W. Y. Yoshida, et al. (2002). "Lyngbyaloside B, a new glycoside macrolide from a Palauan marine cyanobacterium, *Lyngbya* sp." J Nat Prod **65**(12): 1945-8.

Luesch, H., W. Y. Yoshida, et al. (1999). "Lyngbyastatin 2 and norlyngbyastatin 2, analogues of dolastatin G and nordolastatin G from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **62**(12): 1702-6.

Luesch, H., W. Y. Yoshida, et al. (2000). "Apramides A-G, novel lipopeptides from the marine cyanobacterium Lyngbya majuscula." J Nat Prod **63**(8): 1106-12.

Luesch, H., W. Y. Yoshida, et al. (2000). "Isolation and structure of the cytotoxin lyngbyabellin B and absolute configuration of lyngbyapeptin A from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **63**(10): 1437-9.

Luesch, H., W. Y. Yoshida, et al. (2002). "New apratoxins of marine cyanobacterial origin from Guam and Palau." Bioorg Med Chem **10**(6): 1973-8.

Luesch, H., W. Y. Yoshida, et al. (2002). "Structurally diverse new alkaloids from Palauan collections of the apratoxin-producing marine cyanobacterium *Lyngbya* sp." Tetrahedron **58**(39): 7959-7966.

Luesch, H., W. Y. Yoshida, et al. (2001). "Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula*." J Am Chem Soc **123**(23): 5418-23.

Luesch, H., W. Y. Yoshida, et al. (2000). "Isolation, structure determination, and biological activity of Lyngbyabellin A from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **63**(5): 611-5.

Ma, D., B. Zou, et al. (2006). "Total synthesis of the cyclodepsipeptide apratoxin A and its analogues and assessment of their biological activities." Chemistry **12**(29): 7615-26.

Marquez, B. L., K. S. Watts, et al. (2002). "Structure and absolute stereochemistry of hectochlorin, a potent stimulator of actin assembly." J Nat Prod **65**(6): 866-71.

Matthew, S., C. Ross, et al. (2007). "Lyngbyastatin 4, a dolastatin 13 analogue with elastase and chymotrypsin inhibitory activity from the marine cyanobacterium *Lyngbya confervoides*." J Nat Prod **70**(1): 124-7.

Matthew, S., P. J. Schupp, et al. (2008). "Apratoxin E, a cytotoxic peptolide from a guamanian collection of the marine cyanobacterium *Lyngbya bouillonii*." J Nat Prod **71**(6): 1113-6.

Mutou, T., T. Kondo, et al. (1996). "Isolation and stereostructures of dolastatin G and nordolastatin G, cytotoxic 35-membered cyclodepsipeptides from the Japanese sea hare *Dolabella auricularia*." J Org Chem **61**(18): 6340-6345.

Nubel, U., F. Garcia-Pichel, et al. (1997). "PCR primers to amplify 16S rRNA genes from cyanobacteria." Appl Environ Microbiol **63**(8): 3327-32.

Ramaswamy, A. V., C. M. Sorrels, et al. (2007). "Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **70**(12): 1977-86.

Rippka, R. (1988). "Recognition and identification of cyanobacteria." Methods Enzymology **167**: 28-67.

Rippka, R. J., J. Deruelles, et al. (1981). Provisional generic assignments for cyanobacteria in pure culture. The Prokaryotes. H. G. Schlegel. Heidelberg, Springer-Verlag. **1**.

Shen, S., P. Zhang, et al. (2009). "Cyclodepsipeptide toxin promotes the degradation of Hsp90 client proteins through chaperone-mediated autophagy." J Cell Biol.

Sone, H., T. Kondo, et al. (1995). "Dolabellin, a cytotoxic bisthiazole metabolite from the sea hare *Dolabella auricularia*: structural determination and synthesis." The Journal of Organic Chemistry **60**(15): 4774-4781.

Tan, L. T., B. L. Marquez, et al. (2002). "Lyngbouilloside, a novel glycosidic macrolide from the marine cyanobacterium *Lyngbya bouillonii*." J Nat Prod **65**(6): 925-8.

Taori, K., S. Matthew, et al. (2007). "Lyngbyastatins 5-7, potent elastase inhibitors from Floridian marine cyanobacteria, *Lyngbya* spp." J Nat Prod **70**(10): 1593-600.

Thacker, R. W. and V. J. Paul (2004). "Morphological, chemical, and genetic diversity of tropical marine cyanobacteria *Lyngbya* spp. and *Symploca* spp. (Oscillatoriales)." Appl Environ Microbiol **70**(6): 3305-12.

Tidgewell, K., B. R. Clark, et al. (2010). The natural products chemistry of cyanobacteria.

Valeriote, F., C. K. Grieshaber, et al. (2002). "Discovery and development of anticancer agents from plants." J Exp Ther Oncol **2**(4): 228-36.

Versluis, M., B. Schmitz, et al. (2000). "How snapping shrimp snap: through cavitating bubbles." Science **289**(5487): 2114-2117.

Williams, P. G., H. Luesch, et al. (2003). "Continuing studies on the cyanobacterium *Lyngbya* sp.: isolation and structure determination of 15-norlyngbyapeptin A and lyngbyabellin D." J Nat Prod **66**(5): 595-8.

Williams, P. G., R. E. Moore, et al. (2003). "Isolation and structure determination of lyngbyastatin 3, a lyngbyastatin 1 homologue from the marine cyanobacterium *Lyngbya majuscula* Determination of the configuration of the 4-amino-2,2-dimethyl-3-oxopentanoic acid unit in majusculamide C, dolastatin 12, lyngbyastatin 1, and lyngbyastatin 3 from cyanobacteria." J Nat Prod **66**(10): 1356-63.

Chapter 3. Isolation of a Cryptic Gene Cluster from *Lyngbya bouillonii*

3.1 Abstract

Lyngbya bouillonii has also proven to be exceptionally rich in unique secondary metabolites of which the apratoxins, the laingolides, madangolide, and others have been isolated. However, no biosynthetic pathways have been isolated from this organism yet. Studies were undertaken toward the isolation of the biosynthetic pathway for apratoxin A, which was shown to mediate its potent cytotoxicity through the induction of G_1 cell cycle arrest and an apoptotic cascade (Luesch, Chanda et al. 2006). Apratoxin A is predicted to be of mixed PKS/NRPS biosynthesis bearing a thiazoline ring flanked by polyketide portions that possess unusual methylation patterns.

To isolate the gene cluster from *L. bouillonii* that specifies the biosynthesis of apratoxin A, a 6X genomic DNA fosmid library was created and screened using digoxigenin (DIG) labeled, polymerase chain reaction (PCR)-generated probes (NRPS probes for tyrosine and isoleucine and six different PKS probes). The NRPS and PKS probes hybridized to a total of 35 fosmids. Candidate fosmids were further analyzed by restriction mapping, Southern blot, and PCR screening. Sequence analysis and BLAST searches of PCR amplified regions of the 35 fosmids revealed adenylation- domain binding pockets for valine, isoleucine, phenylalanine and tyrosine and for ketosynthase domains. Two overlapping fosmids were shown to harbor PKS regions as well as an HMG-CoA synthase like domain. These were therefore identified as likely candidates for containing the putative apratoxin A biosynthetic pathway. Shotgun sequencing, blastp, Pfam, and other bioinformatic analyses of one fosmid revealed 108 total open

reading frames (ORFs), 17 of which are arranged in a classic modular type I PKS fashion, and are thought to make up the upstream portion of a secondary metabolite pathway. A unique tandem-methyltransferase domain within the loading module was discovered, as well as an HMGCoA synthase-like cassette. This portion of the pathway is most consistent with the predicted pathways for laingolide or madangolide, but not apratoxin.

3.2 Introduction

Natural products play a crucial role in the development of all classes of new pharmaceuticals, especially in the area of oncology. For example, of all cancer drugs available during the period 1940-2002, 40% are natural products or natural product derived, 10% are biologics or vaccines, 20% are synthetics that mimic natural product pharmacophores, and just 30% are of uniquely synthetic origin. (Newman, Cragg et al. 2003)

Cyanobacteria have proven to be a valuable source of natural products, with a broad spectrum of biological activities, in particular anti-tumor activity (Jaspars and Lawton 1998). To date, six natural product biosynthetic pathways have been isolated from the marine cyanobacterium, *Lyngbya majuscula:* barbamide (Chang, Flatt et al. 2002), carmabin (Gerwick *et al*., unpublished data), curacin A (Chang, Sitachitta et al. 2004), hectochlorin (Ramaswamy, Sorrels et al. 2007), jamaicamide A (Edwards, Marquez et al. 2004), and lyngbyatoxin (Edwards and Gerwick 2004) (Figure 3.1). Each pathway primarily follows a co-linear route to synthesis utilizing both PKS and NRPS elements.

However, each pathway also possesses its own set of unique domain architectures and underlying mechanistic chemistries. For example, the intriguing trichloromethyl moiety in barbamide was the subject of very revealing biosynthetic experiments. Feeding studies with isotopically labeled precursors established that it is biosynthetically derived from the *pro*-R methyl group of *L*-leucine (Flatt, O'Connell et al. 2006). This led to the discovery that a tandem pair of novel nonheme Fe^{II} halogenases is responsible for catalyzing the

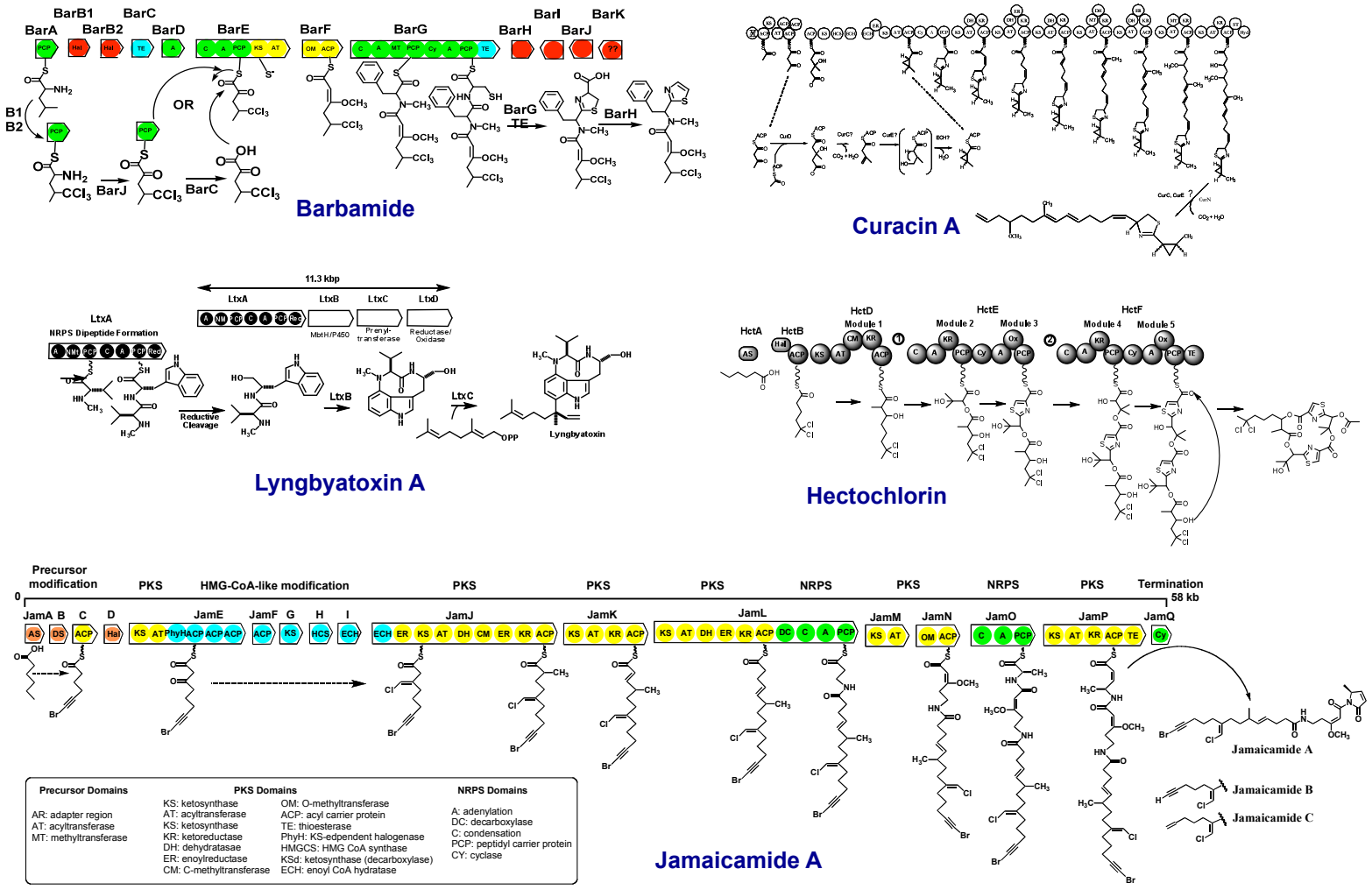


Figure 3.1 Biosynthetic pathways from the marine cyanobacterium *Lyngbya majuscula*.

dichlorination (by BarB2) and final conversion to a trichloromethyl group by BarB1, (Vaillancourt, Yeh et al. 2006). Also, the curacin and jamaicamide pathways provide an unprecedented opportunity to study a unique evolutionary divergence of biochemistry that is perfectly mirrored by their gene identities. Bioinformatic analyses of the HMG-CoA synthase-like manifold in these two parallel assemblies show an extraordinarily high sequence identity in the upstream portion of the manifold, but decreases further downstream. The enzymes include a halogenase, a 3-hydroxy-3-methylglutaryl-CoA synthase (HMG-CoA synthase) cassette for polyketide β-branching, and an enoyl reductase domain. Comparative analysis of these curacin (Cur) and jamaicamide (Jam) enzymes revealed that the sequence identities of the halogenasess, ACPs, KSs, HCSs and ECH_1s are extraordinarily high (~90%), whereas the ECH_2s and ERs show significantly lower sequence identity (~60%) (Gu, Wang et al. 2009). The pathways biochemically diverge to form a β-branched cyclopropane in the Cur pathway, and a vinyl chloride group in the Jam pathway. These discoveries provide new information towards the growing knowledge base of secondary metabolite biosynthesis. Having access to these gene clusters also provides the foundation to potentially address issues of supply by either bioengineering individual domains for precursor-directed biosynthesis or by heterologous expression of complete pathways.

A related, filamentous marine cyanobacterium, *Lyngbya bouillonii,* has proven to be exceptionally rich in unique secondary metabolites. The discoveries of thirty new natural products from this productive organism have been reported (Klein, Braekman et al. 1997; Klein, Braekman et al. 1999; Luesch, Yoshida et al. 1999; Luesch, Yoshida et

al. 2000; Luesch, Williams et al. 2002; Luesch, Yoshida et al. 2002). More specifically, madangolide, laingolide. laingolide A (Klein, Braekman et al. 1996; Klein, Braekman et al. 1999) and the apratoxin series (A, B, C, E, F and G) (Luesch *et al., 2001*; Luesch *et al,* 2002) (Tidgewell, unpublished data), have been isolated from *L. bouillonii* strains inhabiting coral reefs surrounding the islands of Guam, Palau and Papua New Guinea. The compound apratoxin A is of particular interest due to its potent cytotoxic activity against a number of cancer cell lines and has some activity against early-stage colon adenocarcinoma C38 derived tumors in *in vivo* studies. This organism clearly has great propensity to produce unique chemical entities that are predicted to arise from unprecedented mechanistic underpinnings (Figure 3.1). At the outset of this project, no biosynthetic pathways had been isolated from this organism. Together, these characteristics made *L. bouillonii* a great candidate for discovering new biochemistry and to contribute to the growing knowledge base of NRPS and PKS biosynthetic systems.

The molecular genetic methodology towards targeting and isolating a biosynthetic gene cluster from a genome first utilizes knowledge of the compound's chemical structure. Based on current marine cyanobacterial secondary metabolite pathways and PKS/NRPS enzyme logic, predictions can be made about the genetic architecture underlying the production of a particular family of compounds. These predictions can then be used to identify a "genetic handle" that may be unique to the compound of interest. Using this approach, one can design gene primers/molecular probes that will amplify/hybridize to the targeted genes, thereby identifying the desired gene locus in the biosynthetic pathway. Other molecular genetic tools can then be used to isolate the whole gene cluster locus. The versatility of this approach is attractive in that one can

incorporate a general targeting strategy to maximize the potential success of the initial screens. One might also take a more expeditious approach that assumes direct detection of a specific gene locus, or thirdly, one can utilize a combination of these two approaches.

3.2.1 General Gene Cluster Isolation Strategy

Lyngbya bouillonii is known to produce secondary metabolites that have similar structural motifs (Figure 3.2). Isolation of a specific gene cluster from this organism has proven to be very difficult (previous attempts by former post-docs Kerry McPhail and Patricia Flatt were unsuccessful). Therefore, to maximize success it was necessary to carry out a multi-dimensional screening approach. The goal was to isolate a gene cluster or multiple gene clusters that harbor certain genetic elements. These elements were predicted as shared amongst the biosynthetic pathways for apratoxin A, laingolide and madangolide. A combination of gene probes were designed to target the isolation of a hybrid PKS/NRPS gene cluster from a genomic DNA (gDNA) library. From that data set, one could focus in on a more unique feature of the compound of interest. One such combination was to probe for several ketosynthase (KS) and adenylation (A) domains and then narrow the search with an HMG-CoA synthase (HCS) like domain probe. This approach would most likely lead to isolating a gene cluster responsible for the production of an interesting molecule such as apratoxin A, with unique chemical functionalities and therefore new underlying biochemistry.

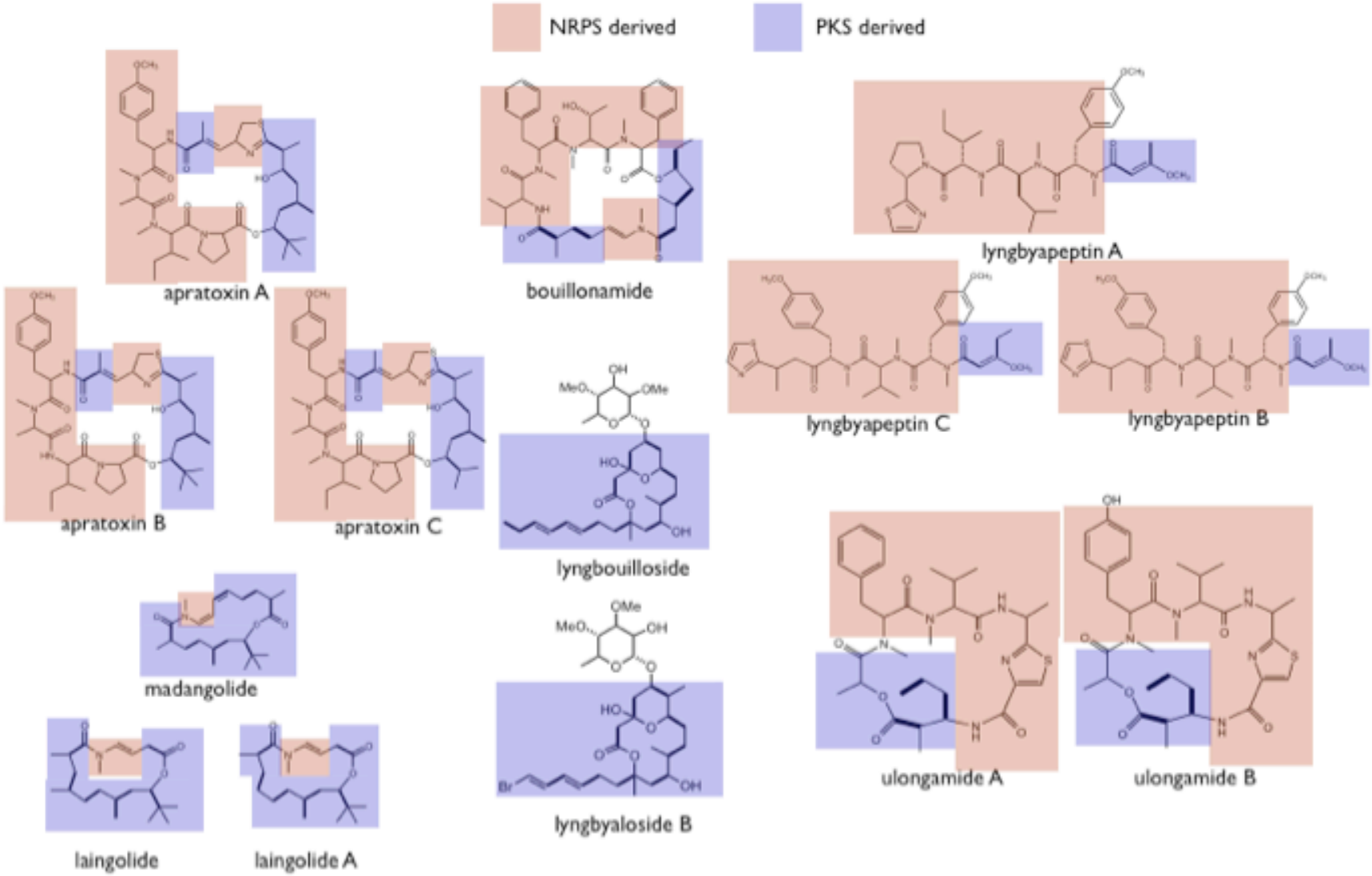
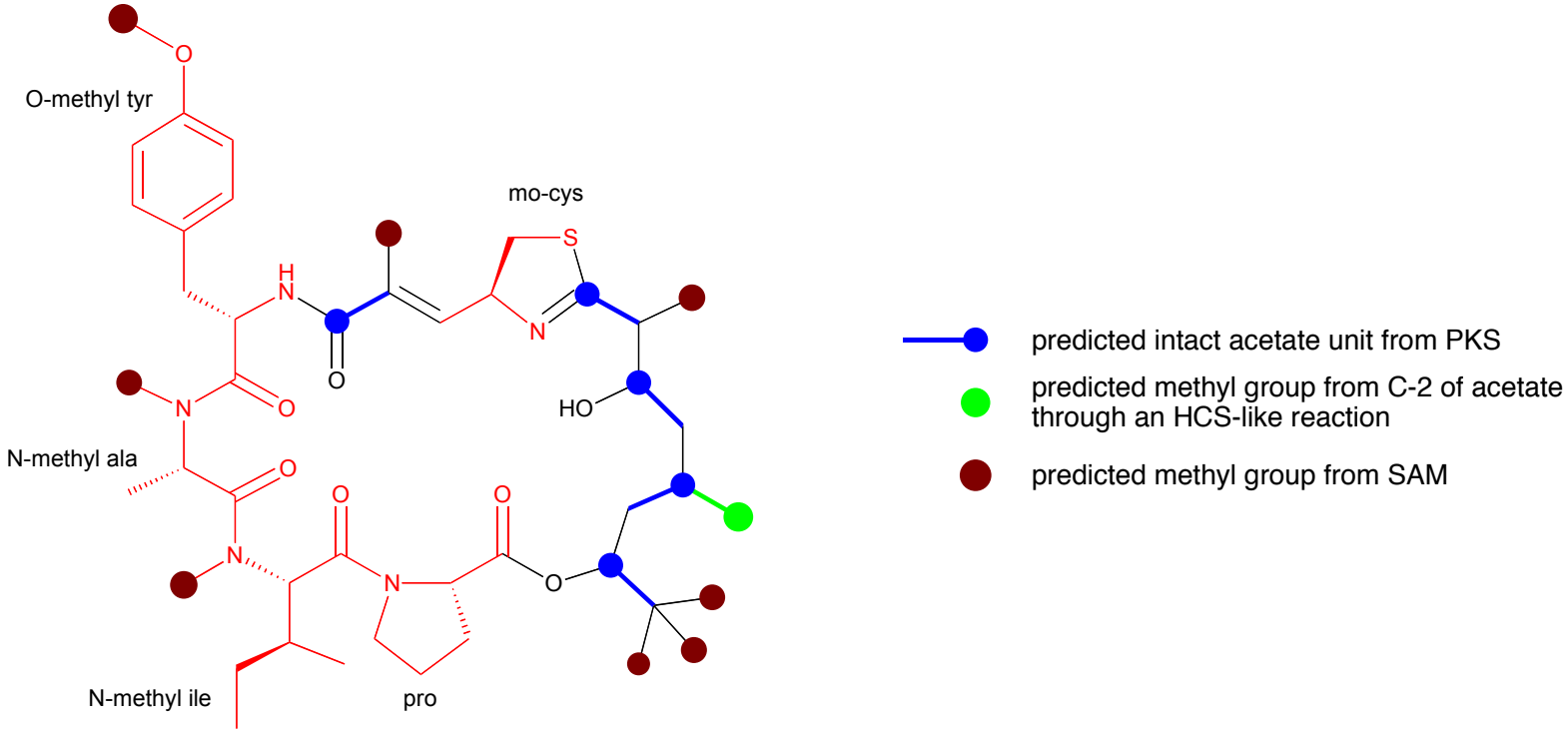


Figure 3.2. Select compounds isolated from *L. bouillonii*. Colors denote moieties arising from NRPS (pink) and PKS (purple) activity. Many compounds share similar biosynthetic elements.

3.3 Results and Discussion

3.3.1 Predicted Biosynthesis of Apratoxin A

In utilizing a molecular targeting approach to identify the biosynthetic pathway for apratoxin A, a dissection of its biosynthetic genesis is necessary. The apratoxin family of compounds are complex cyclodepsipeptides of mixed biosynthetic origin possessing both polyketide and peptide sections (Figure 3.3). The *t*-butyl starter group is a functionality shared by only a few known secondary metabolites from cyanobacteria including antillatoxin (Berman, Gerwick et al. 1999), madangolide, laingolide and the makalika analogs (Gallimore, Galario et al. 2000). This functionality may come from an unusual *t*-pentanoic acid starter, or alternatively from acetate, propionate, or *iso*-butyrate undergoing one to three SAM-derived methylations. Two rounds of ketide extension condenses two acetate units onto the growing substrate. The recognizable HMGCS manifold then presumably adds an acetate unit to the β-ketide position. Further decarboxylation, dehydration and reduction, predicted from the ECH_1-ECH_2-ER domains typically associated with this cassette affords the fully saturated methyl moiety. One more round of ketide extension and C-2 SAM methylation completes the polyketide section. One NRPS module is predicted to condense a modified cysteine onto the growing chain. Ketide extension and SAM-methylation incorporates the secondary methyl group on the C-2 position of the incoming acetate unit. Next, a series of NRPS modules incorporates *O*-methyl tyrosine, *N*-methyl alanine, *N*-methyl isoleucine, and proline. The final hydrolytic macrocyclization is predicted to be from a type I thioesterase (TE).



Apratoxin A

Figure 3.3. Deduced biosynthetic elements of apratoxin A. Mechanistically intriguing features include the *tert*-butyl starter functionality and the HMGCS derived methyl group (green). These features can be used as genetic handles in identifying the gene cluster loci for this compound.

3.3.2 Genomic Library Production

The first step towards generating a genomic library of an environmental sample of *L. bouillonii* was to isolate and purify high molecular weight genomic DNA (HMW gDNA). Several methods were explored to optimize a protocol that consistently yielded intact HMW gDNA in large concentrations and in a form suitable for subsequent manipulation and analysis (see experimental section). Having done so, HMW gDNA was isolated from a Papua New Guinean field collected strain of *Lyngbya bouillonii,* given the collection code 08/30/2001-10, PNG. Approximately 1 μg of high quality gDNA (Figure 3.4) was end repaired, ligated into a fosmid vector, packaged into phage particles and adsorbed onto a strain of engineered *E. coli* cells (EPI 300 cells). These efforts resulted in more than 2000 colonies with approximately 40-45 kb of genomic DNA in each fosmid, representing nearly 6X coverage of the estimated 10 MBp genome. The colonies were arrayed onto nutrient agar plates and stored for further experimentation.

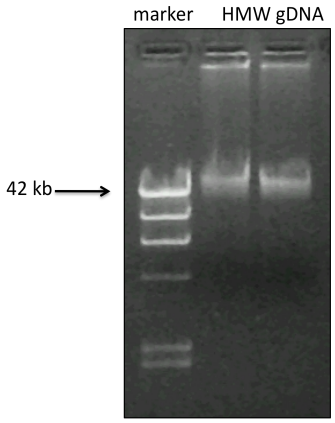


Figure 3.4. High molecular weight (HMW) genomic DNA (gDNA) preparation. Single bands were gel purified, end-repaired and ligated into a vector.

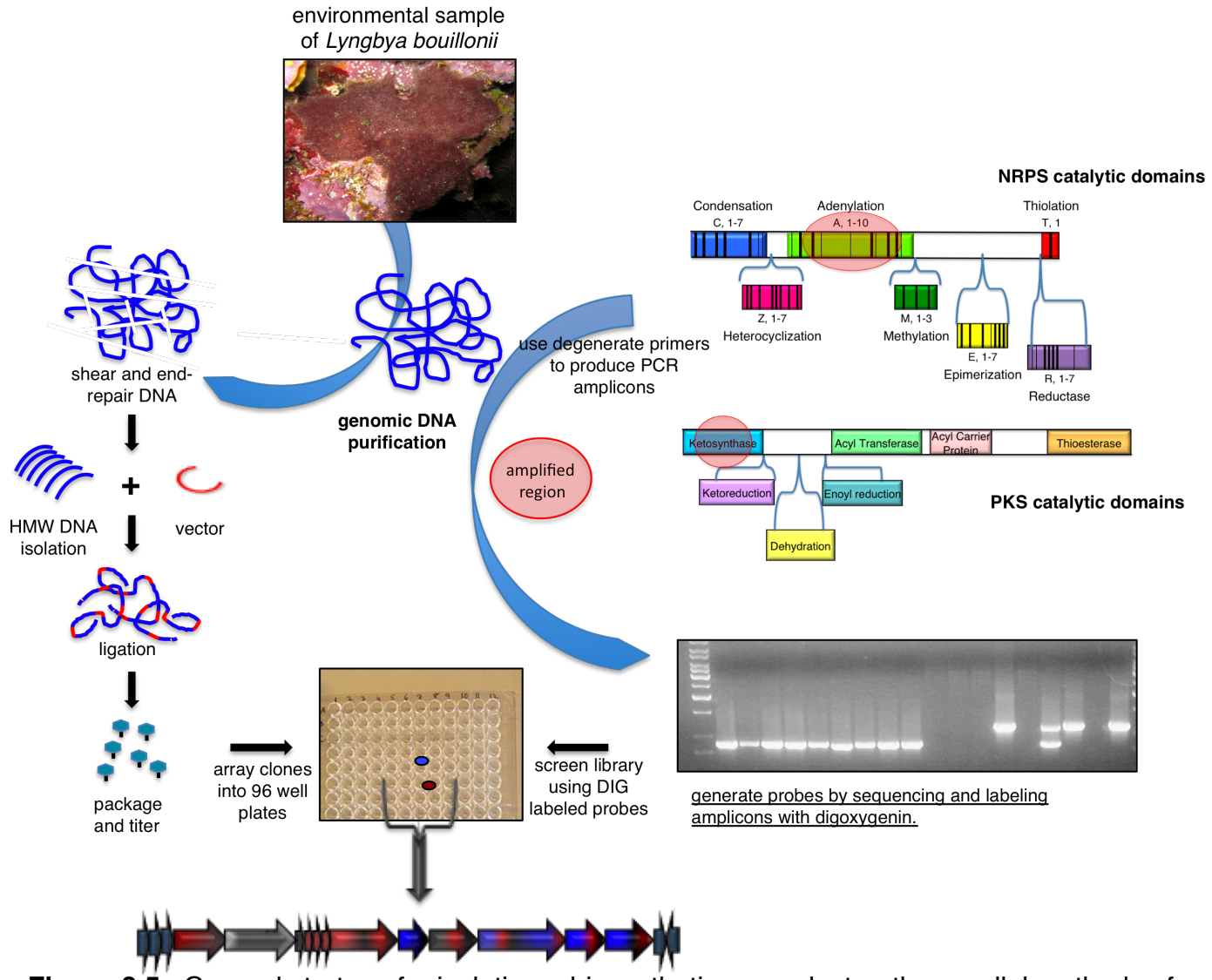


Figure 3.5. General strategy for isolating a biosynthetic gene cluster; the parallel methods of generating a genomic library and amplifying degenerate probes were used in this study.

3.3.3 Designing General Probes for Genome Screening

To isolate a secondary metabolite gene cluster from *L. bouillonii* that specifies the biosynthesis of apratoxin A, laingolide A or madangolide, a gDNA fosmid library was created. Degenerate primers (PKSf 5'-MGIGARGCIHWISMIATGGAYCCICARCAIMG -3') and (PKSf 5'-GGRTCICCIARISWIGTICCIGTICCITG) were designed for KS domains around conserved regions (Figure 3.6). These primers were used to amplify several KS domains from genomic DNA preparations. The approximately 700 base pair (bp) product was cloned into pGEM-T Easy vector and transformed into *E. coli* TOP 10 cells and grown overnight on agar plates. Ten randomly selected clones and were sequenced to verify the insert was in fact a KS domain. This effort yielded seven unique KS sequences. Basic Local Alignment Search Tools (BLAST) analysis of these sequences showed high nucleotide similarity to known polyketide synthases such as curL (88% similarity, 76% identity) and jamK (81% similarity, 67% identity). This suggested that at least seven unique KS gene loci were present in the genomic DNA preparations. However, because we are unable to predict which KS gene belongs to the apratoxin pathway, the KS amplicons were mixed together to generate a "mixed pool KS" probe. The "mixed pool" of KS fragments were then labeled with digoxigenin (DIG) and used to screen the 6X clone library by hybridization and Southern blot analysis.

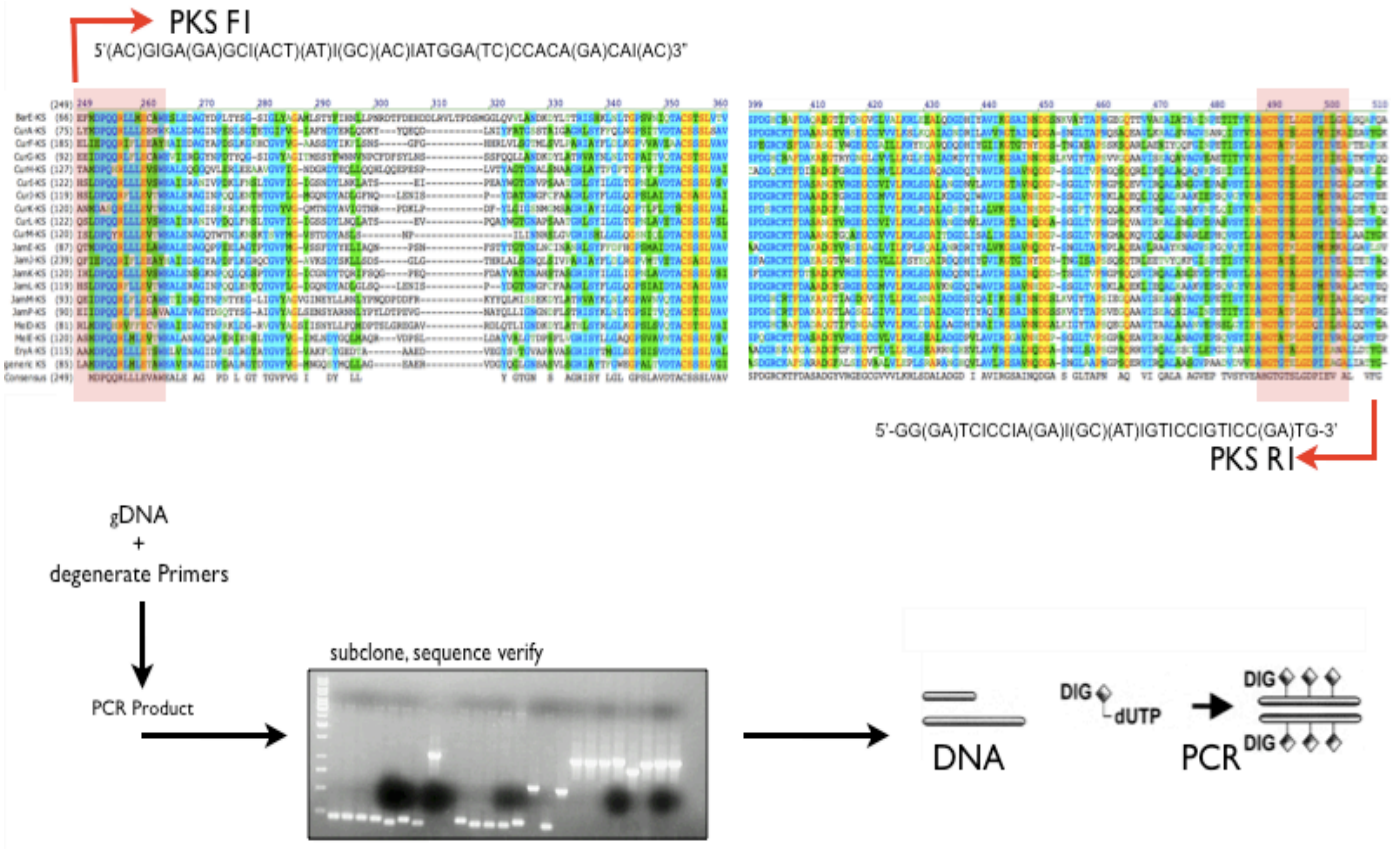


Figure 3.6. Degenerate primer design around conserved regions of the KS domain. Amplicons were subcloned and sequence verified. PCR products were then pooled, labeled with DIG and used as a molecular probe against a genomic DNA library.

Although KS domains are useful markers in identifying a general PKS pathway, they do not provide much information in terms of specific moieties introduced into the nascent compound. The adenylation domain in an NRPS system is the substrate specificity gatekeeper of an assembly line module. It selects anaryl or amino acid, activates it by formation of an aminoacyl-AMP and transfers it to the thiolate of the adjacent primed thiolation domain which captures the aminoacyl species by thioester formation. The A domain substrate specificity is determined by the side chains of ten distinctamino acid residues forming the substrate binding pocket. These ten residues were defined as the substrate specificity-conferring code of NRPS A domains (Stachelhaus, Mootz et al. 1999). By a database of known specificity-conferring codes with characterized substrate specificity, substrates of uncharacterized A domains can be predicted by bioinformatic analysis tools (Rausch, Weber et al. 2005). Apratoxin A contains a modified cysteine, *N*-methyl tyrosine, *N*-methyl alanine, *N*-methyl isoleucine and proline. Using the degenerate primers (A2f1 5'- GCNGGYGGYGCNTAYGTNCC – 3' and A8r1 5'-CCNCGDATYTTNACYTG -3') designed around the A-2 and A-8 conserved regions of the gene (Figure 3.5), four unique sequences were amplified and confirmed by sequencing sub-clones. Bioinformatically, these sequences showed specificity for the amino acids isoleucine, valine, threonine and tyrosine. Results from screening the library with labeled KS domain probes and sequence verification revealed four clones that harbored KS domains.

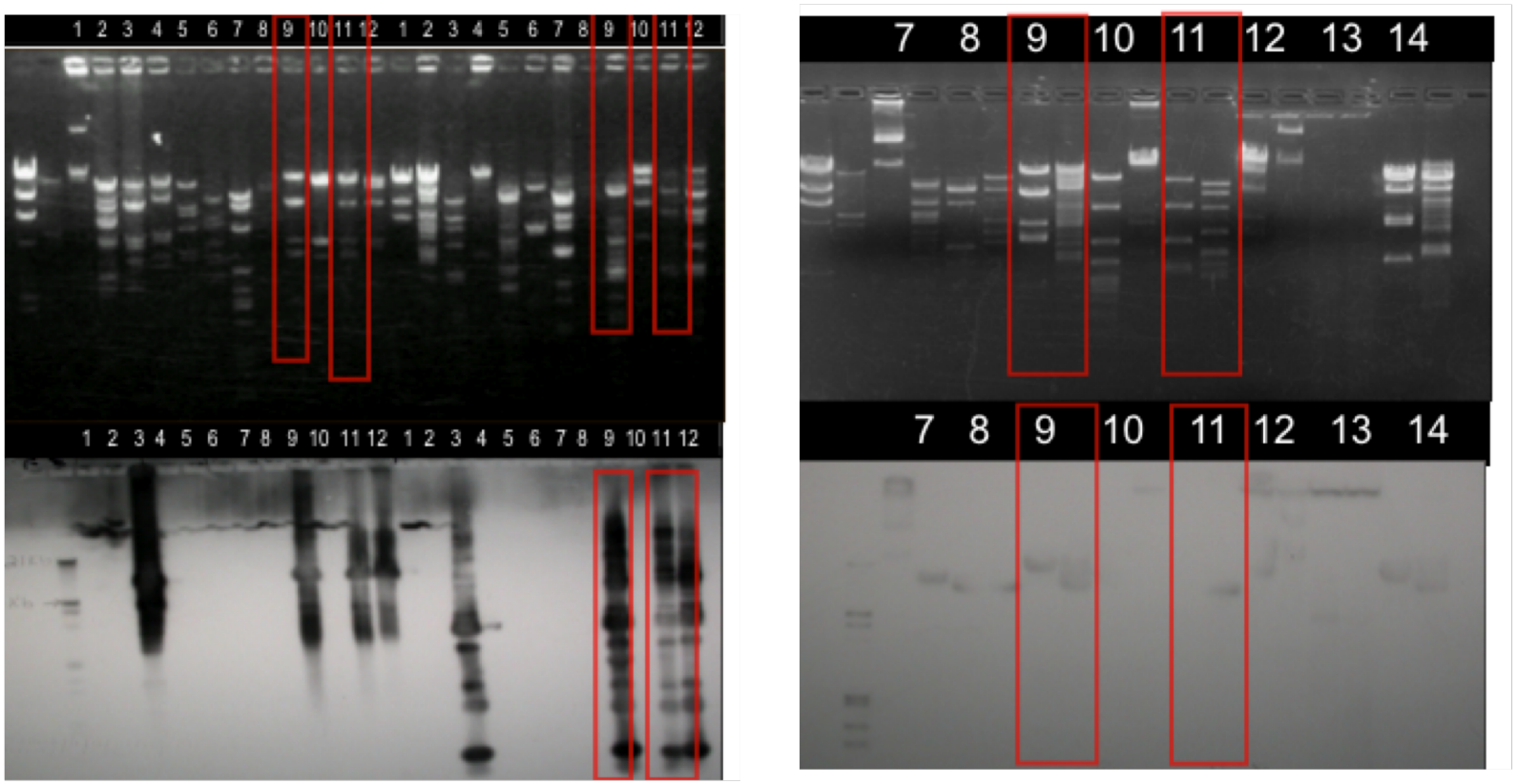


Figure 3.7. Genomic DNA library screen using the mixed pool KS (left) and NRPS (right) revealed two fosmids 278, and 364 (lanes 9 and 11) harboring both biosynthetic gene elements. EtBr stained gel of the candidate fosmids (top panel) cut with EcoRI and Hind III. Southern hybridization (bottom panel) of the probes to the digested fosmids.

3.3.4 Targeting the HMGCS Gene

A distinctive biochemical transformation present in numerous PKS secondary metabolites of cyanobacteria is a branching carbon attached to C-1 carbons of acetate subunits, so called ß-alkylation events. These transformations have been shown to contain an HMG-CoA synthase like (HCS) gene at the core of an extensive cassette of genes, embedded within a PKS pathway, which provide variable tailoring functions to this branching carbon. The genetic architecture and biochemistry of these variably modified ß-branching events have been reported for several natural products, including curacin A (Chang *et al.*, 2004), jamaicamide A (Edwards *et al.,* 2004), mupirocin (El-Sayed, Hothersall, Cooper, Stephens, Simpson, & Thomas, 2003), bacillaene (Calderone, Kowtoniuk, Kelleher, Walsh, & Dorrestein, 2006, Butcher *et al.*, 2007), pederin (Piel, 2002), and myxovirescin (Simunovich, Zapp, Rachid, Krug, Meiser, & Muller, 2006). Each of these clusters have stand alone homologs of a highly conserved HCS, as well as a set of genes encoding one or more ACPs, a mutant KS with a Cys-to-Ser active site substitution (KSs) and two homologs of the enoyl-CoA hydratase (ECH_1 and ECH_2) family. It is hypothesized that these gene products condense acetyl-CoA with the ß-ketoacyl-S-ACP intermediate of a growing polyketide chain to eventually result in the attachment of a branching C-2 of acetate carbon, at various levels of oxidation and functionalization, to a C-1 site. Based on the structure of apratoxin A, we hypothesized that an HCS cassette was responsible for introduction of the saturated methyl group at C-1. The underlying biochemistry is also predicted to be present in the langolide and

madangolide pathways (Figure 3.8). This highly conserved and distinctive motif was thus used as a molecular marker for gDNA library screening.

The next step was to use the degenerate HCS primers to screen those four clones for the HMG-CoA synthase like region. Two of the four clones produced a product of predicted length. Subcloning, sequence verification and multiple sequence alignments of the primary amino acid structure showed that the products from each clone were 99% identical to one another, and 97% identical to the HCS domains in the jamaicamide A and curacin A pathways across approximately 300 amino acids (Figure 3.9). This led to the conclusion that these two fosmids contained both KS domains and the HCS domain, which suggested the one of the pathways of interest was successfully targeted. Further restriction enzyme and Southern blot analysis of the two "positive" fosmids (278 and 364) revealed that they were overlapping by approximately 28 kb. Therefore, based on end-sequencing data from each of the vectors, one fosmid (278) was chosen to move forward with commercial sequencing.

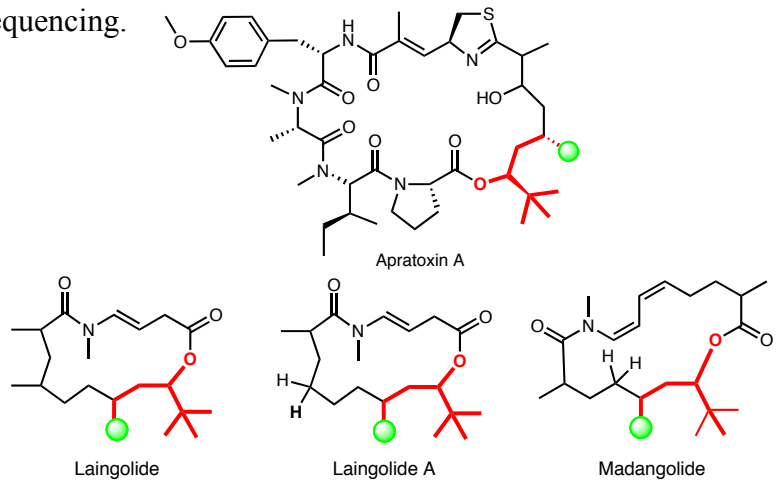
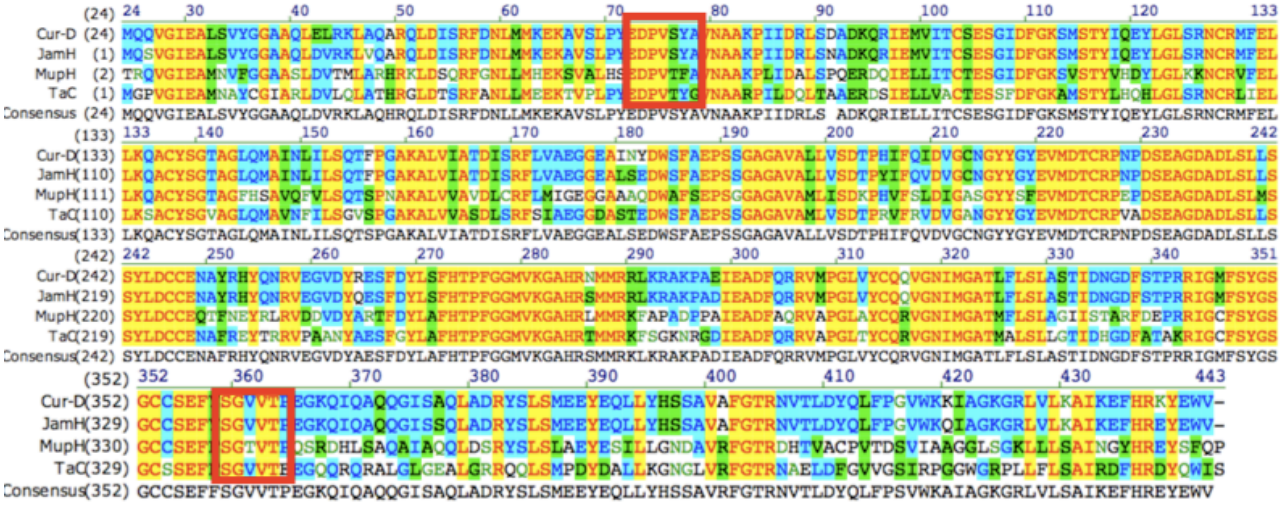


Figure 3.8. Compounds from *L. bouillonii* that have similar structural motifs (red bonds) predicted to be encoded by homologous genes. Predicted C-2 methyl group (green sphere) produced by the HCS gene cassette.



CurD, probable HMG-CoA synthase; **JamH**, HMG-CoA synthase-like protein; **MupH**, HMG-CoA synthase; **TaC**, 3-hydroxy-3-methylglutaryl CoA synthase.

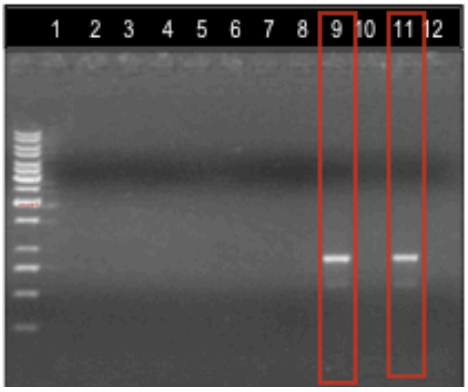
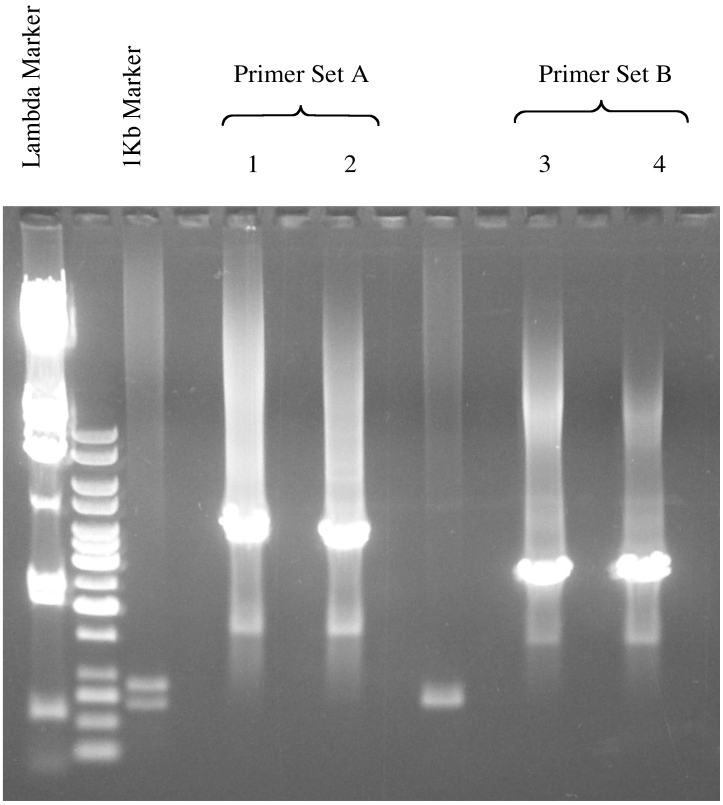


Figure 3.9. Multiple sequence alignment of the HCS domains for CurD, JamH, MpuH, and TaC revealed conserved amino acid sequences (yellow highlights). Degenerate primers were designed around two of these regions (red boxes) and used to screen a subset of fosmids thought to contain PKS and NRPS elements. Results from the screen show two fosmids 278, and 364) (gel lanes 9 and 10) harbor an HCS gene locus.

3.3.5 Long-range PCR Amplification to Map the Overlapping fosmid (364)

End-sequencing and restriction digest analysis revealed that fosmid 364 overlapped fosmid 278 by approximately 28 kbp (Figure 3.11). To gain access to the non-overlapping sequence data, a new PCR technique for full length, locus specific amplification was used. Long range PCR allows the production of amplicons much larger than those achieved with conventional *Taq* polymerases. Up to 27 kb fragments are possible from good quality genomic DNA, although 10 - 20 kb fragments are routinely achievable, given the appropriate conditions. This method relies on a thermostable DNA polymerase with high processivity (i.e. 5'-3' polymerase activity) and with 3'-5' proofreading abilities. By designing specific primers to the 3' end of fosmid 278 and using vector primers, the amplification of an additional 4.5 kb of the pathway was achieved. The amplicons were used as template for six rounds of gene walking resulting in the extension of the biosynthetic pathway by 1.5 modules, adding one full module with the complete host of reductive enzymes. The following module is a partial PKS module consisting of a KS and AT domain.



Primer Set A : PccF/278F1

Primer Set B : PccF/278R1(rc)

Figure 3.10. Long-range PCR amplification from fosmid 364 using two different primer sets. These amplicons were used as template for gene walking.

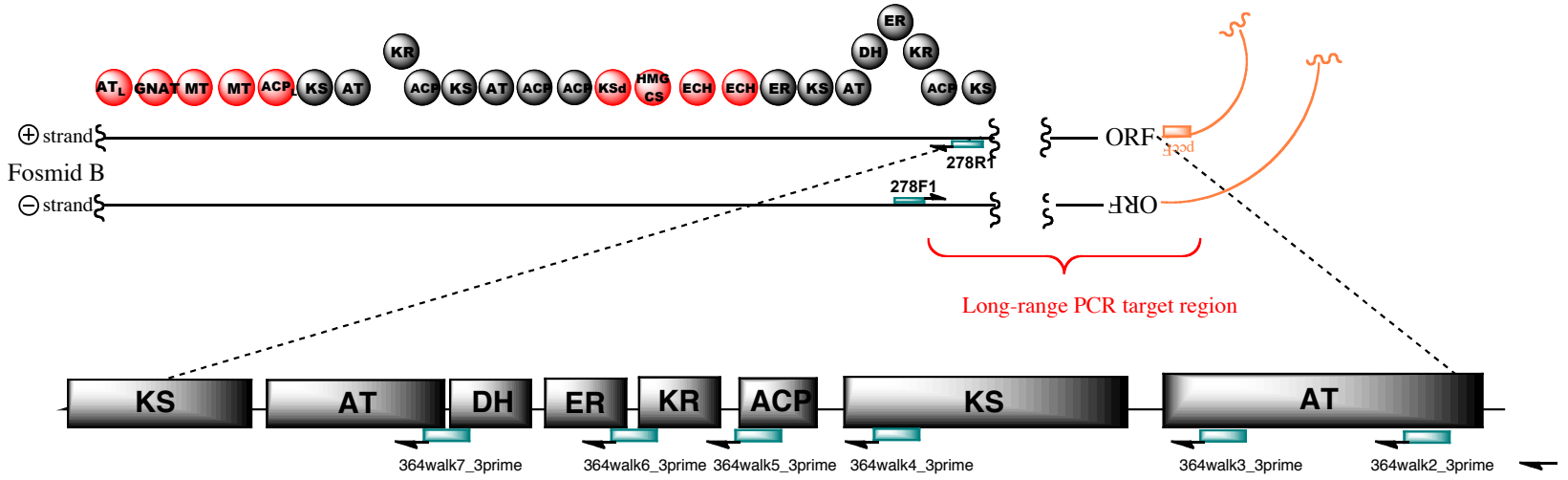


Figure 3.11. Long range PCR amplification and gene walking on fosmid 364 resulted in the extension of the biosynthetic pathway by 4.5 kb, adding one full module with the complete host of reductive enzymes. The following module is a partial PKS module consisting of a KS and AT domain.

3.3.6 Bioinformatic Analysis and Gene Annotation

The roughly 42 kb DNA insert of fosmid-278 and 364 contained a contiguous stretch of eleven large open reading frames (ORFS) approximately 28 kb long with an overall G+C content of 40%. This is consistent with the G+C content encountered in cyanobacterial genomes and other secondary metabolite gene clusters (Chang *et al*., 2002, Chang *et al*., 2004; Edwards & Gerwick, 2004). This indicated this biosynthetic gene cluster warranted further analyses. The general domain architecture was annotated using numerous web-based tools to manipulate and translate the DNA into ORFs (see appendix) and submitting them to both BLAST and BLASTp database searches. To identify conserved active site amino acids, pair-wise and multiple sequence alignment tools were used to compare the unknown ORFs with known catalytic domains. Based on these studies, the resulting annotated pathway was achieved (Figure 3.12). The biosynthetic gene cluster is a type I modular PKS system comprised of a loading module, two extension modules, an HCS cassette (highly homologous to the *jam* and *cur* HCS genes) third and fourth extension module containing the full complement of reductive domains, and finally, an incomplete fifth module where only a KS and AT domain is present. This indicates that I have only part of a biosynthetic gene cluster. Nevertheless, information gained from the bioinformatics analysis of the catalytic domains and the prediction of a partial structure has led to some informative conclusions. Based on the number of modules and the architecture of the catalytic domains within each module, I can predict a partial structure. Assuming co-linearity of the primary

structure in the assembly of, comparison to the known compounds isolated from *L. bouillonii* rules out the production of apratoxin A. However, this structure is still consistent with the production of laingolide, laingolide A or madangolide. Another possibility is that I have isolated an orphan pathway; that is, one for which there is currently no compound to assign to this pathway. This scenario can possibly lead to the opportunity of predicting the chemical structure of the new entity and then to isolating the new compound.

Although the 3' end of this biosynthetic pathway has not yet been isolated, there are areas of high homology to unique domains in both the curacin A and jamaicamide A pathways. For example, the loading module of *bouX* exhibits high homology to the N-terminal end of curA, of which the biochemistry was evaluated (Gu, Geders et al. 2007), and the HCS cassette has highly similar gene architecture to the HCS cassettes in both the *jam* and *cur* pathways. Investigation by comparative analysis of these catalytic domains within the context of the curacin A and jamaicamide A structures could lead to new insights into the biochemical underpinnings of these new catalytic domains.

Furthermore, the bouX annotated cluster harbors regions that are predicted to be homologous to the apratoxin cluster. For example, the putative product of the *bou*X loading module is a *t*-butyl functionality, similar to the starting unit of apratoxin A. Thus the *bou*X cluster sequence was used as a reference data in a later attempt to isolate the apratoxin biosynthetic gene cluster(Chapter 4).

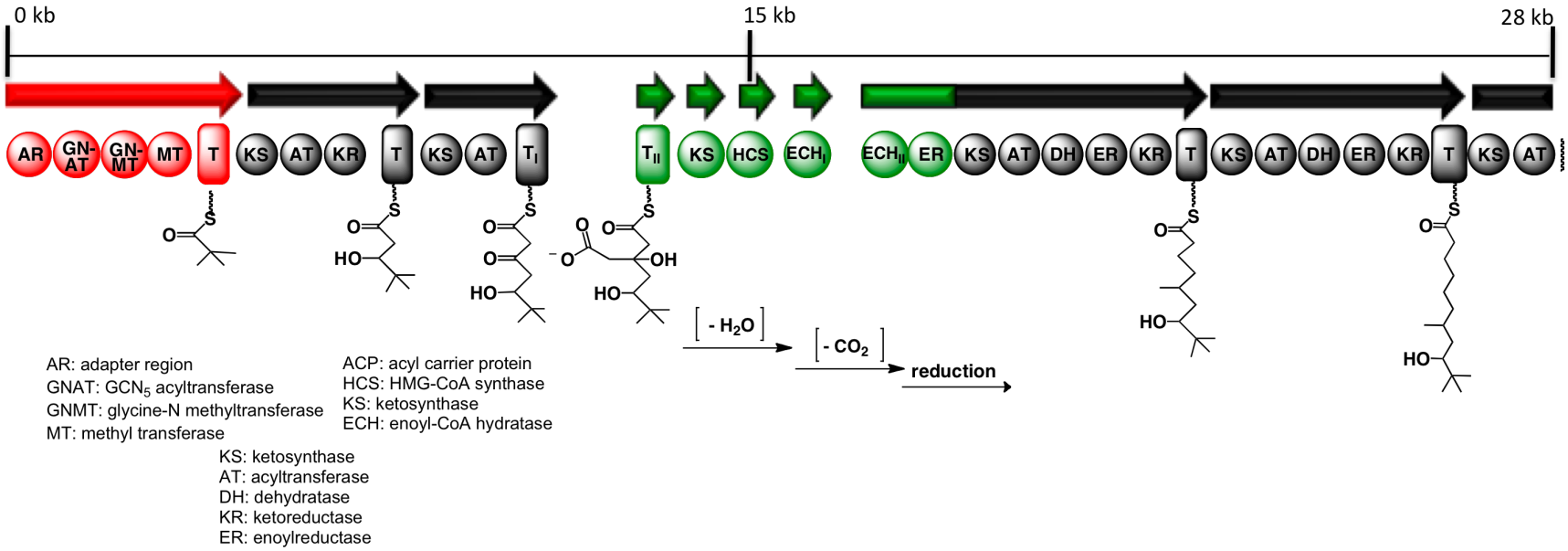


Figure 3.12. Gene annotation of the partial biosynthetic pathway, *bouX*, that is predicted to be responsible for the production of laingolide, laingolide A, madangolide, or an as yet unidentified compound.

3.4 Experimental

Bacterial Strains and Growth Conditions. *Escherichia coli* strain EPI300 was used in this study as a host for fosmid library construction and amplification of fosmid clones. One Shot® TOP10 chemically competent *E. coli* containing pGEM-T easy was grown overnight in LB (Luria-Bertani) medium with ampicillin at a final concentration of 100 mg/ml. The EPI300 strain of *E. coli* harboring fosmid vector was grown overnight in LB medium with chloramphenicol at a final concentration of 12 mg/ml.

Genomic DNA isolation and library construction. DNA from *Lyngbya bouillonii* collection sample PNG/08/03/2001-10 used for PCR amplification was isolated with a modified genomic DNA protocol from the Wizard Kit (Promega, Madison, WI). High molecular weight (HMW) DNA from cultured PNG/08/03/2001-10 was used for library construction, as previously described. The HMW-DNA was end repaired, size selected, and ligated into the copy control fosmid vector pCC1FOS using protocols provided with CopyControl Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, WI). Plasmid preparations were carried out using commercial kits (Qiagen, Santa Clarita, CA) and (FosmidMAX™ Epicentre Biotechnologies, Madison, WI). Other basic DNA manipulations, such as restriction digests and ligations, were done using standard methods (ref Sambrook, 1989).

PCR Cloning of KS, A, and HMGCS Gene Homologs. PCR amplification of probe fragments used in this study was performed with Taq DNA-polymerase (Promega) with the manufacturer's suggested concentration of template and primers in an Eppendorf Mastercycler gradient system. Conditions used were as follows: denaturation, 30 s at 94°C; annealing, 30 s at 48°C; extension, 60 s at 72°C; 30 cycles. Degenerate primers were designed based on two conserved sequences found in HMGCS-like and associated genes from PKS pathways (forward primer 5'-CTNCCNTAYGAYGAYCCCGT-3' and reverse primer 5'-NCKRTGNGCNCCYTTNACCAT-3'). These primers were used to amplify a 900 bp fragment from *L. bouillonii* genomic DNA. The amplicons were cloned into pGEM-T, were DNA sequenced. For amplification of the -ketosynthase domain (KS) fragments from the *L. bouillonii* genomic DNA, previously designed primers KS1Up: 5'-MGI GAR GCI HWI SMI ATG GAY CCI CAR CAI MG-3' KSD1: 5'-GGR TCI CCI ARI SWI GTI CCI GTI CCR TG-3' were used. These primers were used to amplify an approximately 700 bp fragment. Using the degenerate primers (A2f1 5'- GCNGGYGGYGCNTAYGTNCC – 3' and A8r1 5'-CCNCGDATYTTNACYTG -3') designed around the A-2 and A-8 conserved regions of the gene. The amplicons were cloned into pGEM-T, were DNA sequenced.

3.5 References

Berman, F. W., W. H. Gerwick, et al. (1999). "Antillatoxin and kalkitoxin, ichthyotoxins from the tropical cyanobacterium Lyngbya majuscula, induce distinct temporal patterns of NMDA receptor-mediated neurotoxicity." Toxicon **37**(11): 1645-8.

Chang, Z., P. Flatt, et al. (2002). "The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit." Gene **296**(1-2): 235-47.

Chang, Z., N. Sitachitta, et al. (2004). "Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **67**(8): 1356-67.

Edwards, D. J. and W. H. Gerwick (2004). "Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase." J Am Chem Soc **126**(37): 11432-3.

Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*." Chem Biol **11**(6): 817-33.

Flatt, P. M., S. J. O'Connell, et al. (2006). "Characterization of the initial enzymatic steps of barbamide biosynthesis." J Nat Prod **69**(6): 938-44.

Gallimore, W. A., D. L. Galario, et al. (2000). "Two complex proline esters from the sea hare *Stylocheilus longicauda*." J Nat Prod **63**(7): 1022-6.

Gu, L., T. W. Geders, et al. (2007). "GNAT-like strategy for polyketide chain initiation." Science **318**(5852): 970-4.

Gu, L., B. Wang, et al. (2009). "Metamorphic enzyme assembly in polyketide diversification." Nature **459**(7247): 731-5.

Jaspars, M. and L. A. Lawton (1998). "Cyanobacteria - a novel source of pharmaceuticals." Current Opinion in Drug Discovery & Development **1**: 8.

Klein, D., J.-C. Braekman, et al. (1999). "Lyngbyapeptin A, a modified tetrapeptide from *Lyngbya bouillonii* (Cyanophyceae)." Tetrahedron Letters **40**(4): 695-696.

Klein, D., J.-C. Braekman, et al. (1996). "Laingolide, a novel 15-membered macrolide from *Lyngbya bouillonii* (cyanophyceae)." Tetrahedron Letters **37**(42): 7519-7520.

Klein, D., J. C. Braekman, et al. (1999). "Madangolide and laingolide A, two novel macrolides from *Lyngbya bouillonii* (Cyanobacteria)." Journal of Natural Products **62**(6): 934-936.

Klein, D., J. C. Braekman, et al. (1997). "Lyngbyaloside, a novel 2,3,4-tri-*O*-methyl-6-deoxy-alpha-mannopyranoside macrolide from *Lyngbya bouillonii* (Cyanobacteria)." Journal of Natural Products **60**(10): 1057-1059.

Luesch, H., S. K. Chanda, et al. (2006). "A functional genomics approach to the mode of action of apratoxin A." Nat Chem Biol **2**(3): 158-67.

Luesch, H., P. G. Williams, et al. (2002). "Ulongamides A-F, new beta-amino acid-containing cyclodepsipeptides from Palauan collections of the marine cyanobacterium *Lyngbya* sp." J Nat Prod **65**(7): 996-1000.

Luesch, H., W. Y. Yoshida, et al. (1999). "Lyngbyastatin 2 and norlyngbyastatin 2, analogues of dolastatin G and nordolastatin G from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **62**(12): 1702-6.

Luesch, H., W. Y. Yoshida, et al. (2000). "Isolation and structure of the cytotoxin lyngbyabellin B and absolute configuration of lyngbyapeptin A from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **63**(10): 1437-9.

Luesch, H., W. Y. Yoshida, et al. (2002). "Structurally diverse new alkaloids from Palauan collections of the apratoxin-producing marine cyanobacterium *Lyngbya* sp." Tetrahedron **58**(39): 7959-7966.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural products as sources of new drugs over the period 1981-2002." Journal of Natural Products **66**: 1022-1037.

Ramaswamy, A. V., C. M. Sorrels, et al. (2007). "Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **70**(12): 1977-86.

Rausch, C., T. Weber, et al. (2005). "Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs)." Nucleic Acids Res **33**(18): 5799-808.

Stachelhaus, T., H. D. Mootz, et al. (1999). "The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases." Chem Biol **6**(8): 493-505.

Vaillancourt, F. H., E. Yeh, et al. (2006). "Nature's inventory of halogenation catalysts: oxidative strategies predominate." Chem Rev **106**(8): 3364-78.

Chapter 4. Single-cell Genome Amplification Advances Discovery of the Elusive Apratoxin A Gene Cluster.

4.1 Abstract

Filamentous marine cyanobacteria are an extraordinarily rich source of structurally novel and biomedically relevant natural products. To understand the biosynthetic origin of these valuable natural products, as well as to produce increased supplies and analog molecules, it is necessary to access the clustered biosynthetic genes which encode for the assembly enzymes. Complicating this pursuit is the universal presence of heterotrophic bacteria that abundantly occupy the interstitial spaces of filamentous cyanobacteria, both those obtained from the environment or those grown in uni-cyanobacterial culture. The presence of these heterotrophic bacteria make impossible the preparation of pure cyanobacterial genomic DNA. Additionally, high similarity and redundancy in genetic elements across disparate secondary metabolite biosynthetic pathways renders current gene cluster targeting strategies imprecise and inefficient. To address these two major impediments, we have developed a general strategy which utilizes single-cell genomic sequencing techniques that are based on multiple displacement amplification (MDA) and metagenomic library screening. This approach is portable and can be optimized to target natural product pathways in a diversity of other highly complex symbiont communities. Here, we report application of this approach to the identification of the apratoxin A biosynthetic cluster, a potent cancer cell cytotoxic agent with high promise for medicinal applications. The roughly 54 kb biosynthetic gene cluster is

composed of 12 open reading frames and has a type I modular mixed polyketide synthase/nonribosomal peptide synthatase (PKS/NRPS) organization and features loading and off-loading domain architecture never previously described. Moreover, this work represents the first successful isolation of a complete biosynthetic gene cluster from *Lyngbya bouillonii*, a tropical marine cyanobacterium renowned for its production of diverse bioactive secondary metabolites.

4.2 Introduction

Understanding of the mechanistic chemistry underlying the biosynthesis of bacterial polyketide and non-ribosomal peptide natural products has been greatly enhanced through access to the fundamental biosynthetic genes; the DEBS gene cluster, encoding the production of the parent aglycone to the broad spectrum antibiotic erythromycin, provides the prototypical example (Caffrey, Bevitt et al. 1992). Fortunately, in most cases, the biosynthetic genes are clustered in these prokaryote genomes, a feature which facilitates their discovery and characterization (Edwards, Marquez et al. 2004; Ramaswamy, Sorrels et al. 2007). The current methods for locating biosynthetic gene clusters typically use homology-based hybridization or PCR screening of genomic libraries (Chang, Sitachitta et al. 2004), or bioinformatic approaches with sequenced genomes (Cheng, Yang et al. 2007; Rounge, Rohrlack et al. 2009). However, with exotic and less well-studied organisms that are obtained from environmental field collections, such as marine cyanobacteria, they are typically found growing as complex microbial consortia, and

genomic information is lacking. In these cases, the current gene targeting approaches are rendered inefficient or inappropriate. Here we expand upon previous genome screening approaches to accelerate the identification of a biosynthetic gene cluster from an environmental sample of the marine cyanobacterium *Lyngbya bouillonii*. To this end, we strategically utilized recently developed methods that enable DNA sequencing from individual bacterial cells (Raghunathan, Ferguson et al. 2005; Lasken 2007) to facilitate subsequent location of a biosynthetic gene cluster in a metagenomic library.

Marine cyanobacteria have emerged as one of the most productive sources of highly bioactive and structurally novel natural product drug leads. The spectrum of their biological activities is extremely broad, and notably includes antitubulin and antiactin agents, neurotoxins, and antibiotic substances (Grindberg, Shuman et al. 2007; Gerwick, Coates et al. 2008). While a majority of these natural products derive from what has been described as a single species, *Lyngbya majuscula* (Tidgewell, Clark et al. 2010), the Indo-Pacific species *L. bouillonii* is also a rich source of new natural products; to date some thirty natural products spanning eight different chemical classes have been isolated from this latter species, thus revealing its exceptional biosynthetic vigor. Among these is the potent cancer cell cytotoxin, apratoxin A (**1**), which has been isolated from *L. bouillonii* inhabiting shallow coral reef environments surrounding the islands of Guam, Palau (Luesch, Yoshida et al. 2001; Luesch, Yoshida et al. 2002) and Papua New Guinea (Gutierrez, Suyama et al. 2008). Apratoxin A is a fascinating structure composed of a polyketide fused with a

modified pentapeptide to form a cyclic lipopeptide and possessing a unique tertiary butyl group at one terminus. Moreover, apratoxin A shows an extremely promising profile of selective cytotoxicity to cancer cells grown on solid agar media (Valeriote, unpublished data). Recently, the mechanism of cytotoxic action of apratoxin A has been described as involving the reversible inhibition of a secretory pathway for several cancer-associated receptors by interfering with their cotranslational translocation (Liu, Law et al. 2009). As a novel structure working by a new mechanism to potently and selectively kill cancer cells, it becomes of interest to understand the underlying genetic encoding for its biosynthesis as well as the chemical mechanism for creation of some of its distinctive structural features.

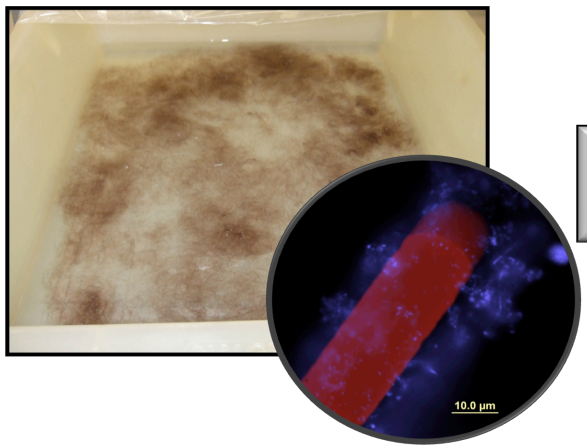
Unfortunately, this filamentous cyanobacterium typically grows with a number of gamma- proteobacteria deeply imbedded in its sheath and other interstitial spaces surrounding the cyanobacterial cells (Figure 4.1) (Hube, Heyduck-Soller et al. 2009). Thus, it is extremely difficult to obtain axenic laboratory cultures of this species, and to date, this has not been achieved. However, by isolating a single *L. bouillonii* cell from this microbial consortium and submitting it to multiple displacement amplification (MDA) (Dean, Nelson et al. 2001; Dean, Hosono et al. 2002; Hosono, Faruqi et al. 2003), the DNA was expanded more than a billion fold, thus generating the microgram quantities of pure *L. bouillonii* DNA required for genome sequencing. The amplified DNA was subjected to 454 pyrosequencing (Marcy, Ouverney et al. 2007) and portions of the apratoxin biosynthetic gene cluster identified through bioinformatics. This information was used to design specific

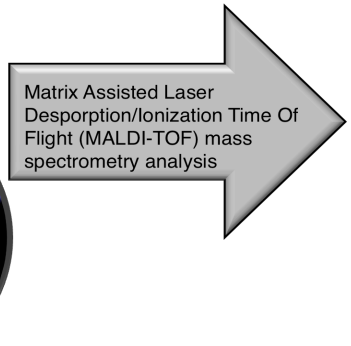
primers for the efficient identification of fosmids containing portions of the apratoxin A pathway genes from a genomic library, ultimately leading to complete characterization of the gene cluster.

4.3 Results

4.3.1 Confirmation of Apratoxin A production by MALDI-MS and i-MALDI.

Uni-cyanobacterial cultures of the IndoPacific tropical marine cyanobacterium *Lyngbya bouillonii* were established by the careful collection of portions of colonies using SCUBA and subsequent manipulations during transport from Papua New Guinea. Once acclimated to laboratory culture, this cyanobacterium grows robustly although with a relatively slow doubling time (ca. 12 days). Randomly selected filaments from a 5 L culture were extracted with 2:1 CH_2Cl_2/MeOH and the extract placed on a Matrix Assisted Laser Desorption Ionization (MALDI) plate. MALDI-TOF analysis indicated that apratoxin A (*m/z* 840.5) was a major metabolite (Figure 4.2), and was accompanied by several apratoxin analogs with different levels of methylation (e.g. apratoxin B and C) (Luesch, Yoshida et al. 2001). Subsequently, cultured single filaments of *L. bouillonii* were harvested and analyzed directly by imaging MALDI (i-MALDI), and confirmed the presence of **1** in these cultures (Erhard, von Dohren et al. 1997) (Esquenazi, Coates et al. 2008).





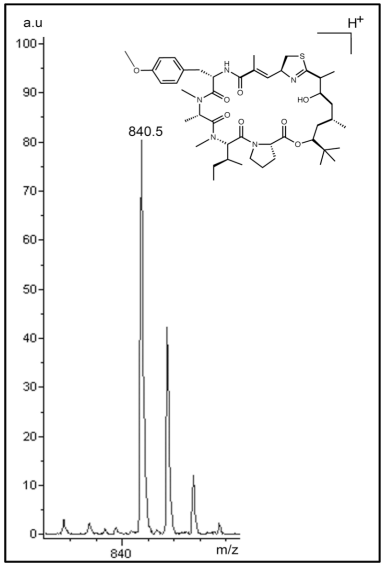


Figure 4.1. 6 liter laboratory grown culture of *Lyngbya bouillonii* originally collected in Papua New Guinea. Inset: DAPI stained and epi-fluorescent

Figure 4.2. MALDI-TOF-imaging of the intact marine cyanobacterium *Lyngbya bouillonii* PNG filament indicates biosynthetic production of apratoxin A.

Light and fluorescent microscopy with and without DAPI staining revealed that this *L. bouillonii* was uni-cyanobacterial, but that the sheath material harbored a rich population of heterotrophic bacteria (Figure 4.1). Recent studies of a related cyanobacterium, *Oscillatoria brevis*, revealed that more than 50 different strains of heterotrophic bacteria can be associated with a cultured filamentous cyanobacteria (Hube, Heyduck-Soller et al. 2009). Unfortunately, antibiotic treatment of these cultures resulted in rapid bleaching and death of the cyanobacterial filaments, and thus, these cultures are best characterized as microbial consortia. Therefore, to obtain pure cyanobacterial DNA for genome sequencing, we isolated single cells from short filaments of *L. bouillonii* strain PNG/08/03/2001 using a micromanipulator and finely drawn glass micropipette. Repetitive gentle pressing of the micropipette tip against the filaments ultimately dislodged single cells, free of their intact sheath. Four such single cells were sequentially recovered (Figure 4.3), cycled through several washing steps, and then placed separately into a one-pot lysis/reaction buffer mixture.

The DNA of each of these four cells was subjected to whole genome amplification using the bacteriophage ϕ29 DNA polymerase. By adding random exonuclease-resistant primers to this highly processive multiple displacement amplification (MDA), the genome of each cell was amplified in a single day, under isothermal conditions, to approximately by 5 μg (Dean, Hosono et al. 2002). To evaluate for intra-cellular genetic variability, as well as to assess genome coverage since MDA is known to show primer bias (Marcy, Ishoey et al. 2007), the quality of

each amplified DNA was analyzed. The 16S rDNA gene was PCR amplified from each MDA-DNA reaction using the primers 27F/Cyano809R, and the products were sub-cloned and sequenced. The 16S rDNA sequence from each amplified cell was aligned and compared to the reference *L. bouillonii* sequence. Phylogenetic analyses showed no significant variation between the neighboring cells from the same filament and that all sequences were closely related to *L. bouillonii* (Figure 4.4).

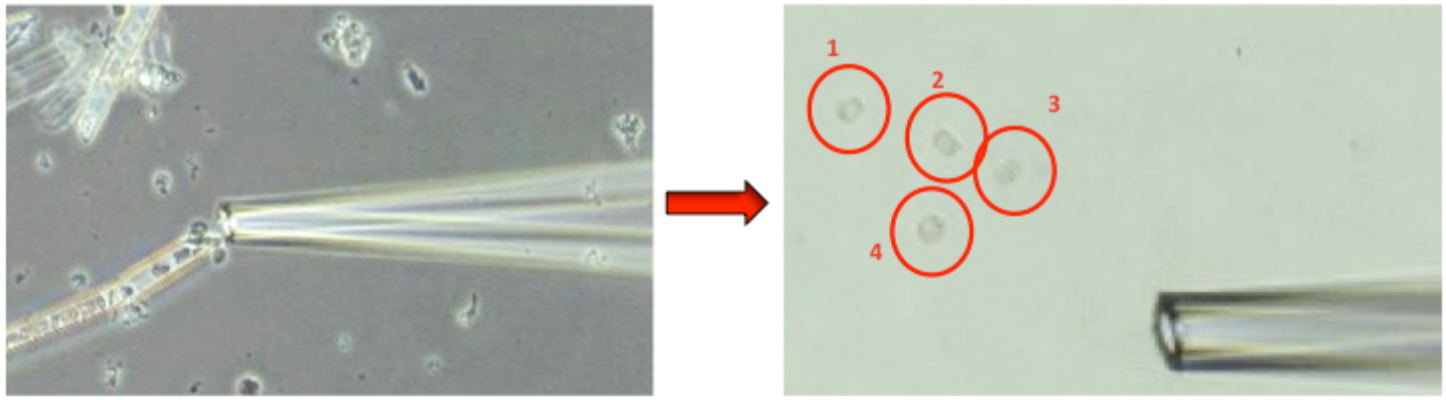


Figure 4.3. Neighboring cells captured from one filament and washed in sterile PBS-buffer prior to individual whole genome amplification by multiple displacement amplification (MDA).

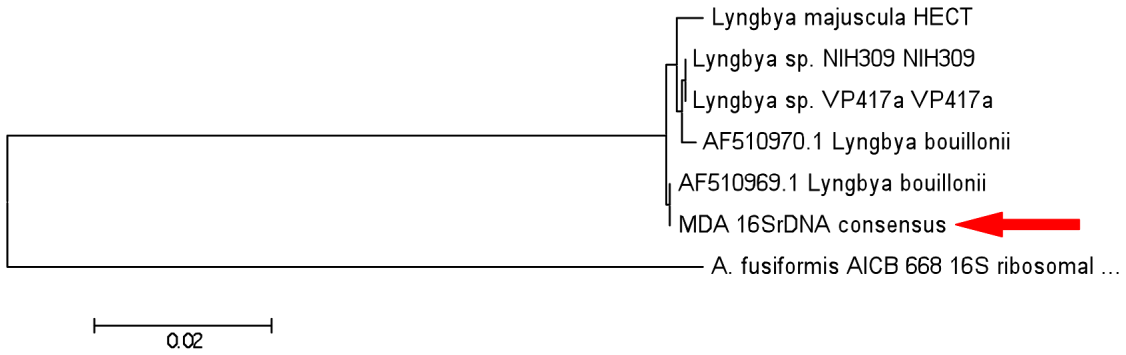


Figure 4.4. Phylogenetic analysis of the 16S rDNA consensus sequence from the individual single cell MDAs.

4.3.2 Targeting the HMGCS gene.

A distinctive biochemical transformation present in numerous PKS secondary metabolites of cyanobacteria is a branching carbon attached to C-1 carbons of acetate subunits, so called ß-alkylation events (Calderone, Iwig et al. 2007; Gu, Wang et al. 2009) These transformations have been shown to contain an HMG-CoA synthase-like (HCS) gene at the core of an extensive cassette of genes, embedded within a PKS pathway, which provide variable tailoring functions to this branching carbon. The genetic architecture and biochemistry of these variably modified ß-branching events have been reported for several natural products, including curacin A (Chang, Sitachitta et al. 2004), jamaicamide A (Edwards, Marquez et al. 2004), mupirocin (El-Sayed, Hothersall et al. 2003), bacillaene (Calderone, Kowtoniuk et al. 2006; Butcher, Schroeder et al. 2007), pederin (Piel, Wen et al. 2004), and myxovirescin (Simunovic, Zapp et al. 2006). Each of these clusters have stand alone homologs of a highly conserved HCS (70% to 98% identity at the amino acid level), as well as a set of genes encoding one or more ACPs, a mutant KS with a Cys-to-Ser active site substitution (KSs) and two homologs of the enoyl-CoA hydratase (ECH_1 and ECH_2) family. It is hypothesized that these gene products condense acetyl-CoA with the ß-ketoacyl-S-ACP intermediate of a growing polyketide chain to eventually result in the attachment of a branching C-2 of acetate carbon, at various levels of oxidation and functionalization,

to a C-1 site. Based on the structure of apratoxin A, we hypothesized that a HMGCS cassette was responsible for introduction of the C-1 saturated methyl group on the polyketide region. This highly conserved and distinctive motif was thus used as a primary molecular marker for *in silico* contig screening, and subsequent *in vitro* metagenome library screening (Figure 4.5).

However, before beginning genome sequencing and assembly, we sought to establish that the MDA-generated DNA possessed at least one HCS homolog, as well as other signature NRPS/PKS biosynthetic genes. First, degenerate PCR primers were designed based on two conserved sequences found in the HMGCS-like gene (forward primer 5'- CTNCCNTAYGAYGAYCCCGT-3' and reverse primer 5'-NCKRTGNGCNCCYTTNACCAT-3'). For amplification of ketosynthase domain (KS) fragments from *L. bouillonii* genomic DNA, the previously designed primers KS1Up: 5'-MGI GAR GCI HWI SMI ATG GAY CCI CAR CAI MG-3' KSD1: 5'-GGR TCI CCI ARI SWI GTI CCI GTI CCR TG-3' were used. These primer sets were successful in amplifying a 900 bp and a 700 bp fragment from the MDA template, respectively. The PCR products were sub-cloned and subsequent DNA sequence analysis confirmed the presence of an HCS gene (~86% identity to JamH and CurD) and several KS genes in the genome.

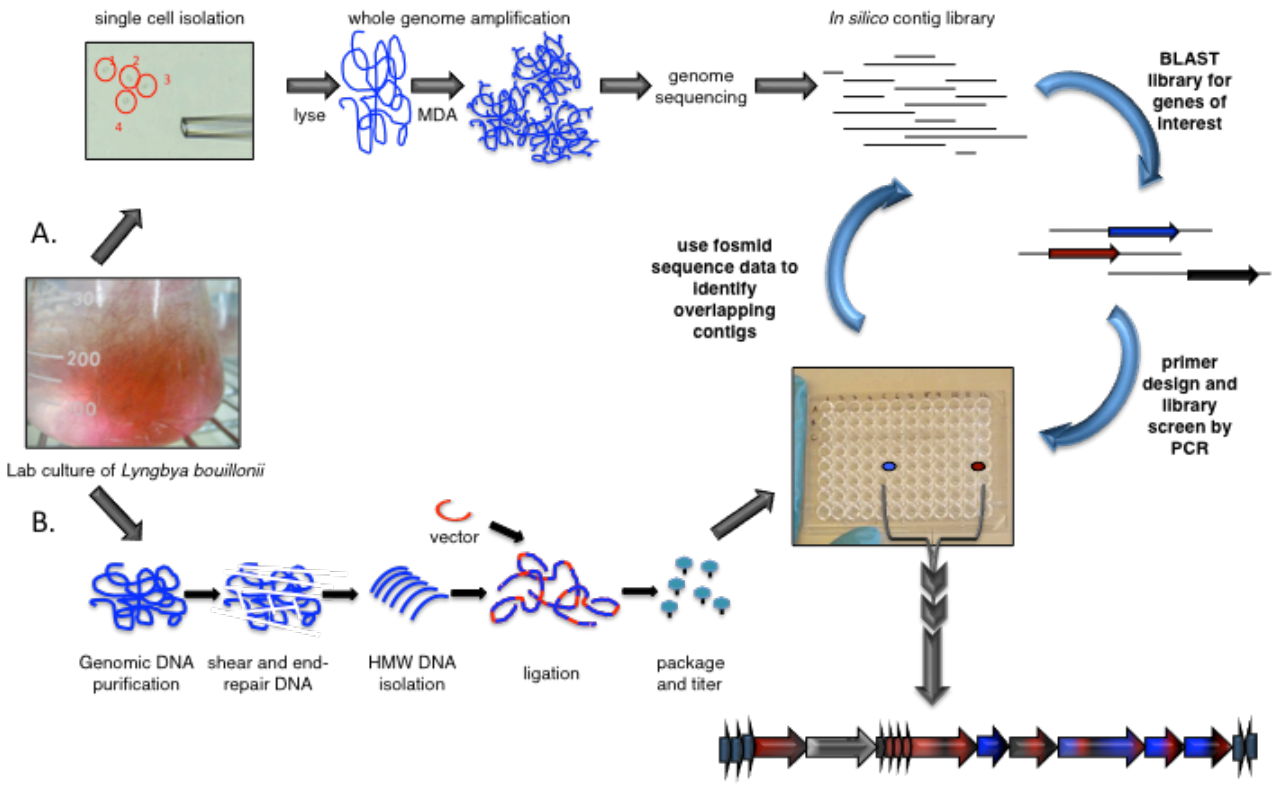


Figure 4.5. Parallel strategies were used to isolate the apratoxin biosynthetic gene cluster. Path A) utilizing micromanipulation techniques to isolate single cells as template source for multiple displacement amplification (MDA) and sequencing of the whole genome. B) Metagenomic DNA used to create and screen a library by PCR.

4.3.3 Genome Sequencing and Assembly.

The MDA amplified *L. bouillonii* DNA was sequenced using the following strategy. One half of a 454 FLX plate was allocated to a single cell MDA and the other half plate was a combination of four individually amplified cells. The rationale behind this strategy was to average potential coverage bias in the second half plate and be able to compare the effect on assembly and annotation statistics. Sets of reads were assembled into long contigs using a hybrid of 454 Newbler and EULER-SR assembly engines, as described below. Combining these two engines was necessary in order to recover sequences that were of low amplification or from regions hard to sequence.

De novo assembly was first achieved using the 454 Newbler assembler. Resulting data consisted of 3502 contigs of size larger than 500 bp, covering 6.6 Mb with the largest contig being 26 Kb equal to 2.2 Kb. Initial screening of HCS signatures in the assembly generated by 454 Newbler revealed that most belonged to short contigs (< 2Kb), and these limited further analysis. However, as discussed below, one 2.05 Kb contig (contig 04978) contained several of the domains recognizable in the HCS cassette (e.g. HMG, ECH_1, ECH_2, Figure 4.6) and was instrumental to locating fosmids containing portions of the apratoxin A (**70**) cluster from a metagenomic library.

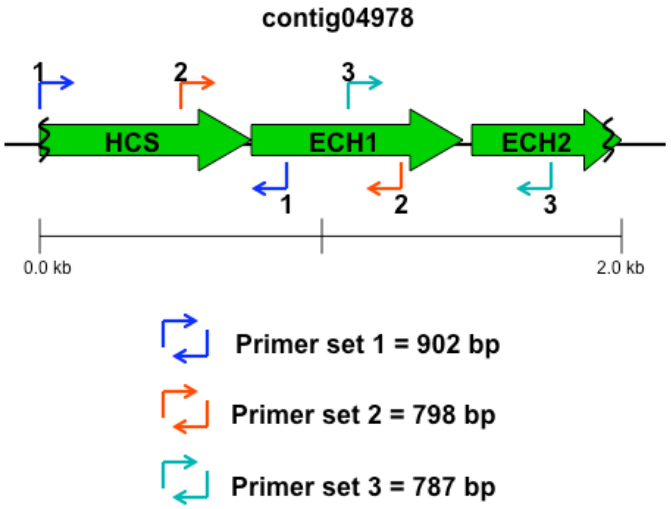
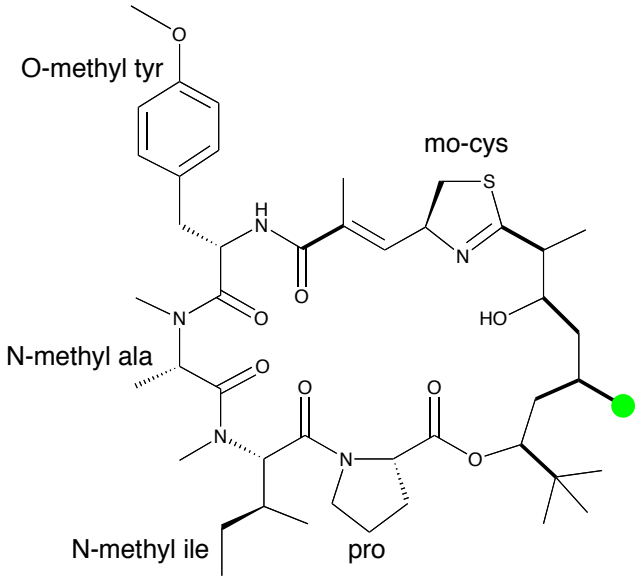


Figure 4.6. *In silico* contig containing the familiar HCS-ECH$_1$-ECH$_2$ catalytic motif. Specific primers were designed (arrows) and used to PCR screen the *in vivo* clone library.



Apratoxin A (**70**)

● predicted methyl group from C-2 of acetate through an HMGCS-like reaction

Figure 4.7. Structure of apratoxin A.

Reasons why the Newbler engine was unable to produce longer contigs included bias introduced during the amplification and errors in the reads themselves. To overcome these limitations we explored methods to extend and merge the initially determined contigs. First, to recover sequence regions that were error prone or of low coverage, we used the EULER-SR assembler which corrects reads prior to assembly and enables a reduction in the 'trustable coverage threshold' for low coverage regions. The goal of this second effort was to assemble as much novel sequence as possible which was not assembled by 454 Newbler. Thus, we ran EULER-SR with parameters favorable for assembling low coverage regions (assuming that 454 Newbler assembled them well). The assembly produced by EULER-SR resulted in about 5000 contigs that were larger than 500 bp with the largest contig being 12.5 Kb.

We next extended and combined contigs generated by 454 Newbler with contigs produced by EULER-SR. Contigs were merged if they had a common seed of size larger than 30 bp with high alignment score. Ultimately, 34 contigs between 10 Kb and 42 Kb, including one contig of 30 Kb which had most of the signature HCS sequences, and mapped to the upstream and roughly first half of the gene cluster from parallel fosmid sequencing efforts, as described below.

As noted above, the initial Newbler assembled contig library was screened by BLAST using an HCS consensus sequence so as to recover all such sequences in the contig library, this survey identified a single 2,050 bp contig (contig 04978) containing the recognizable HMG, ECH_1, ECH_2 domain architecture (Figure 4.7). To

explore if there were other HMGCS genes in this *L. bouillonii* genome, we used the degenerate primers HMG F1/HMGR1 to PCR amplify a 900 bp fragment from the purified metagenomic DNA. Sub-cloning and sequencing of 20 amplicons revealed that the genome contained a single unique HMGCS sequence that was 100% identical to contig 04978. The laboratory production of apratoxin A combined with this finding gave confidence that the genome contained the biosynthetic cluster and that this *in silico* recovered HCS motif was a part of the pathway.

4.3.4 Isolation of the Entire Mixed-PKS/NRPS, HCS-containing Locus from a Metagenomic DNA Library

The identification of a single contig containing the -HMG, ECH_1, ECH_2- motif was essential in identifying the rest of the apratoxin pathway. Our dual library screening approach required the construction of an *in vitro* metagenomic DNA (meta-gDNA) fosmid library. To generate this, high molecular weight (HMW) meta-gDNA was isolated from the laboratory cultured *Lyngbya bouillonii* strain. Approximately 1 μg of DNA was end repaired, ligated into a fosmid vector, packaged into phage particles and adsorbed onto a strain of engineered *E. coli* (EPI 300) cells. These efforts resulted in 1,642 colonies with approximately 40-45 Kb of genomic DNA in each fosmid, representing a nearly 6X coverage of the estimated 10 Mbp genome. The colonies were arrayed into 96 well plates and screened by a previously developed pooling strategy (Hrvatin and Piel 2007).

Using 3 unique sets of specific primers spanning various portions of the HMG, ECH_1, and ECH_2 motifs on contig 04978, we used PCR to screen the *in vitro* metagenomic library for fosmids containing these domains (Figure 4.6). Six fosmids with varying degrees of overlap to each other were identified from this screen. End sequencing of these six fosmids revealed that the DNA inserts harbored biosynthetic elements consistent with the apratoxin biosynthetic pathway, and hence, two of these were selected for full insert shotgun sequencing on the basis of deduced size and pathway coverage. Fosmid Apr1 contained a 38.8 Kb insert which possessed 36.7 Kb of the putative pathway as well as 2.1 Kb of upstream region. The second sequenced fosmid, Apr2 was composed of 35.08 Kb of insert DNA, all of which was deduced as part of the apratoxin pathway, extending from the HMGCS cassette domain, which was the original fosmid selection factor, through three additional extension modules containing PKS and NRPS elements ending at the condensation domain of module 8 (that codes for the condensation of isoleucine and the nascent growing compound). Unfortunately, neither fosmid provided coverage of the 3'-terminus of the cluster, estimated at this point as comprising an additional 5% of the biosynthetic pathway. A triple primer set PCR screen using both intra- and intergenic sequences from the 3' end of fosmid Apr2 was successful in identifying a third fosmid, Apr3, which contained additional sections of the cluster. Unfortunately, it was quickly realized that this fosmid was chimeric from end sequencing which showed overlapping chromatograms. Due to this complexity, as well as the high similarly observed in tandem NRPS modules which likely arise

through gene duplication events (Jenke-Kodama, Sandmann et al. 2005; Rahman, Hothersall et al. 2005), a very careful specific primer walking approach was taken to sequence relevant portions of this fosmid. In total, 21 rounds of primer design/gene walking were required, and extended the cluster an additional 5.4 kp. Luckily, partial confirmation of portions of this tricky sequence was obtained from a contig acquired from the genome sequencing effort (contig 03772). Additionally, a second contig (00221) was found to overlap sections of the 3' end of fosmid Apr3, and extended the cluster by a further 3264 bp. Finally, a third PCR screen of the metagenomic library using the 3' terminal sequence of contig 00221 revealed an overlapping fosmid which was also subjected to shotgun sequencing. This provided the remaining 4000 bp to the presumed end of the cluster, and a further 12.1 Kbp into DNA stretches clearly not encoding for secondary metabolites (sugar transport proteins and transposases). Thus, from data obtained from 4 separate contigs and partial or complete sequencing of 4 fosmids, the 57.4 Kb sequence of the putative gene cluster for apratoxin A was obtained (Figure 4.8).

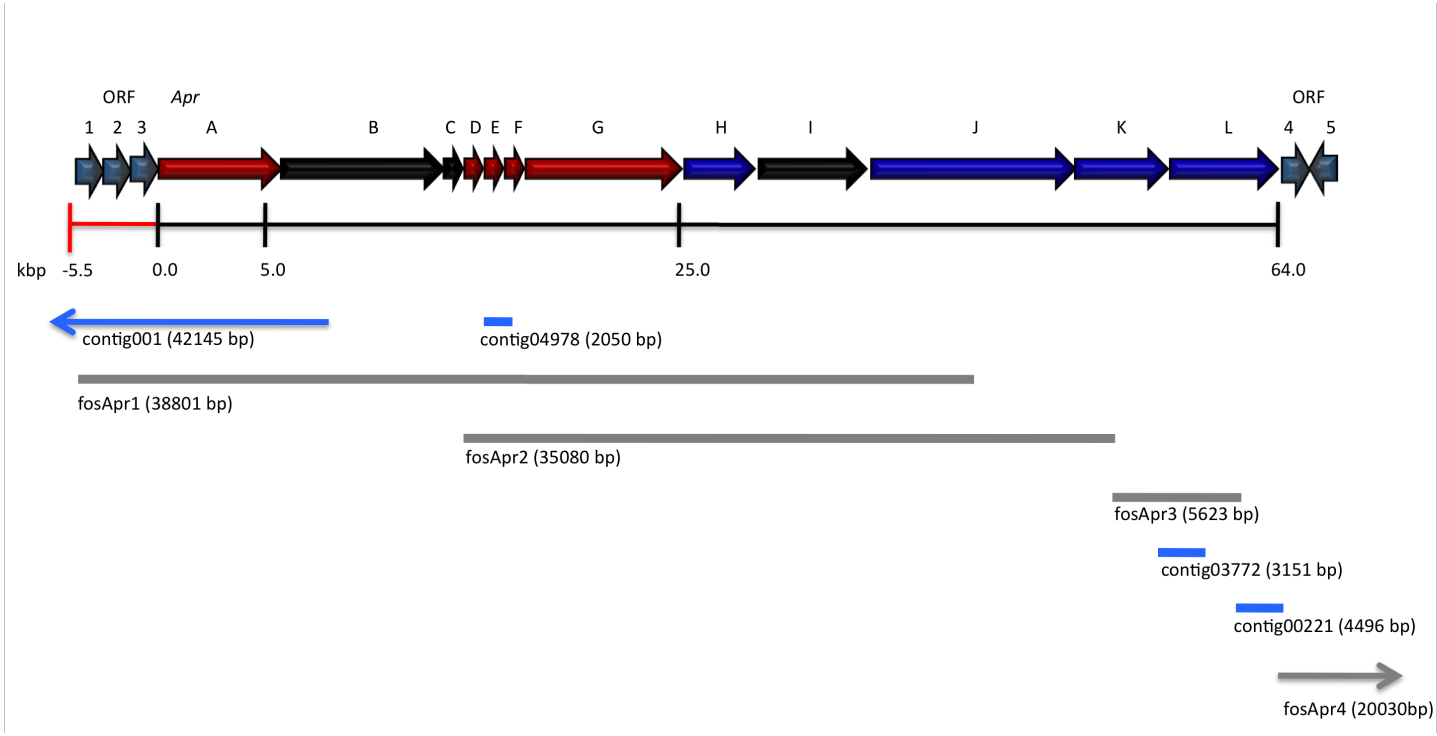


Figure 4.8. Map and gene arrangement of the ~64 kbp *apr* gene cluster from *L. bouillonii* PNG. Inserts of four fosmids (grey) and four contigs (light blue) containing the *jam* gene cluster are shown (size of molecules indicated in parentheses).

4.3.5 Gene Cluster Architecture and Proposed Biosynthesis is Consistent with the Structure of Apratoxin A

The 57,421 bp pathway has type I modular mixed PKS/NRPS organization containing 12 open reading frames including a PKS-type loading module and nine extension modules (four PKS and 4 NRPS) (Figure 4.9). The core modules are flanked by putative transposases and coding regions for hypothetical proteins, presumably delimiting the boundaries of the biosynthetic cluster (Table 4.1). AprB codes for two PKS extension modules with a tandem di-thiolation domain at the C terminus. Following AprB is a series of genes homologous to the familiar HCS-like gene cassettes involved β-branching polyketide backbones. This includes the PCR targeted contig 04978. The motif –T, KS, HCS, ECH_1, ECH_2-ER- (Apr C, AprD, AprE, AprF, and the N terminus AprG) is predicted to add C-2 of acetate to the growing β-keto intermediate. Final elaboration of the pendant methyl group with the fully saturated α-β, C-C bond involves dehydration, decarboxylation and reduction presumably *via* AprF (ECH_1), AprG (ECH_2-ER). The C-terminal region of AprG codes for the addition of acetate and for the subsequent reduction of the resulting β-keto carbon. The C-MT domain is envisioned to add a methyl group *via* SAM to the C-2 position of the nascent compound. The adenylation (A) domain in AprH shows predicted specificity for the amino acid cysteine. The presence of a cyclization (Cy) domain is consistent with the formation of the modified cysteine in apratoxin A. With respect to biosynthesis, the next five

modules (one PKS and four NRPS) are thought to follow their co-linear arrangement and are thus responsible for the successive addition of a modified acetate unit and the *O*-methyl tyrosine, *N*-methyl alanine, *N*-methyl isoleucine, and proline moieties, respectively.

4.3.6 Non-canonical Domains are Predicted to Catalyze Unique Mechanistic Reactions

In addition to the canonical hybrid PKS/NRPS genetic architecture, the putative apratoxin A pathway exhibits sequence similarity to some catalytic domains less common among natural product pathways. The loading module contains regions of high identity (~67%) to a family of GCN_5-related transferases; methyltransferase (GNMT) and acyltransferase (GNAT). The apratoxin A starter functionality is predicted to be the tri-methylated group at the terminal end of the molecule. This could arise by one of several ways. The *t*-butyl group may come from an unusual *t*-pentanoic acid starter, or alternatively from acetate, propionate, or *iso*-butyrate undergoing one to three SAM-derived methylations. The bioinformatic analysis of the catalytic domains could shed light on the natural biosynthetic route.

The homologous CurA GNAT was recently shown to catalyze both the decarboxylation of malonyl-coenzyme A (malonyl-CoA) to acetyl-CoA and to direct S-acetyltransfer from acetyl-CoA to load an adjacent ACP domain (Gu, Geders et al. 2007). By analogy, the GNAT domain could catalyze the decarboxylation of

Table 4.1. Deduced functions of the open reading frames in the *apr* pathway.

protein	size (aa)	catalytic domains	proposed function	sequence similarity	identity/similarity
ORF1	90	transposase	transposase	*Cyanothece* sp. ATCC 51142	34%, 49%
ORF2	45	hypothetical protein	unknown	*Nostoc* sp. PCC7120	57%, 65%
ORF3	61	transposase	transposase	Lyngbya sp. PCC8106	77%, 86%
AprA	1139	AR, MT, AT, MT, ACP	2, 2,-dimethyl propionyl synthase	BryX, *Candidatus E. sertula*	41%, 58%
AprB	2828	KS, AT, KR, T, KS, AT, ACP, ACP	PKS	CurI, *L. majuscula*	58%, 74%
AprC	84	ACP	acyl carrier protein	CurB, *L. majuscula*	79%, 89%
AprD	408	KS	keto-acyl synthase	JamG, *L. majuscula*	74%, 85%
AprE	418	HMGCS-like	Hmg-CoA sythase	JamH, *L. majuscula*	84%, 92%
AprF	259	ECH1	dehydration	JamI, *L. majuscula*	79%, 89%
AprG	2943	ECH2, ER, KS, AT, MT, KR, ACP	PKS	JamJ, *L. majuscula*	67%, 79%
AprH	1177	Cy, A_{cys}, PCP	NRPS	CurF, *L. majuscula*	77%, 86%
AprI	1765	KS, AT, DH, MT, KR, ACP	PKS	JamJ, *L. majuscula*	58%, 73%
AprJ	2955	C, A_{tyr}, MT, PCP, C, A_{ala}, MT, PCP	NRPS	CrpC, *Nostoc* sp. ATCC 53789	56%, 71%
AprK	1264	C, A_{ile}, MT, PCP	NRPS	BarG, *Lyngbya majuscula*	58%, 73%
AprL	2126	C, ACP_{pro}, MT, PCP, C, A	NRPS, Heterocyclization	NosA, *Nostoc* sp. GSV224	45%, 61%
ORF4	77	hypothetical protein	unknown	N/A	N/A
ORF5	38	transposase	transposase	HctC, *Lyngbya majuscula*	84%, 89%

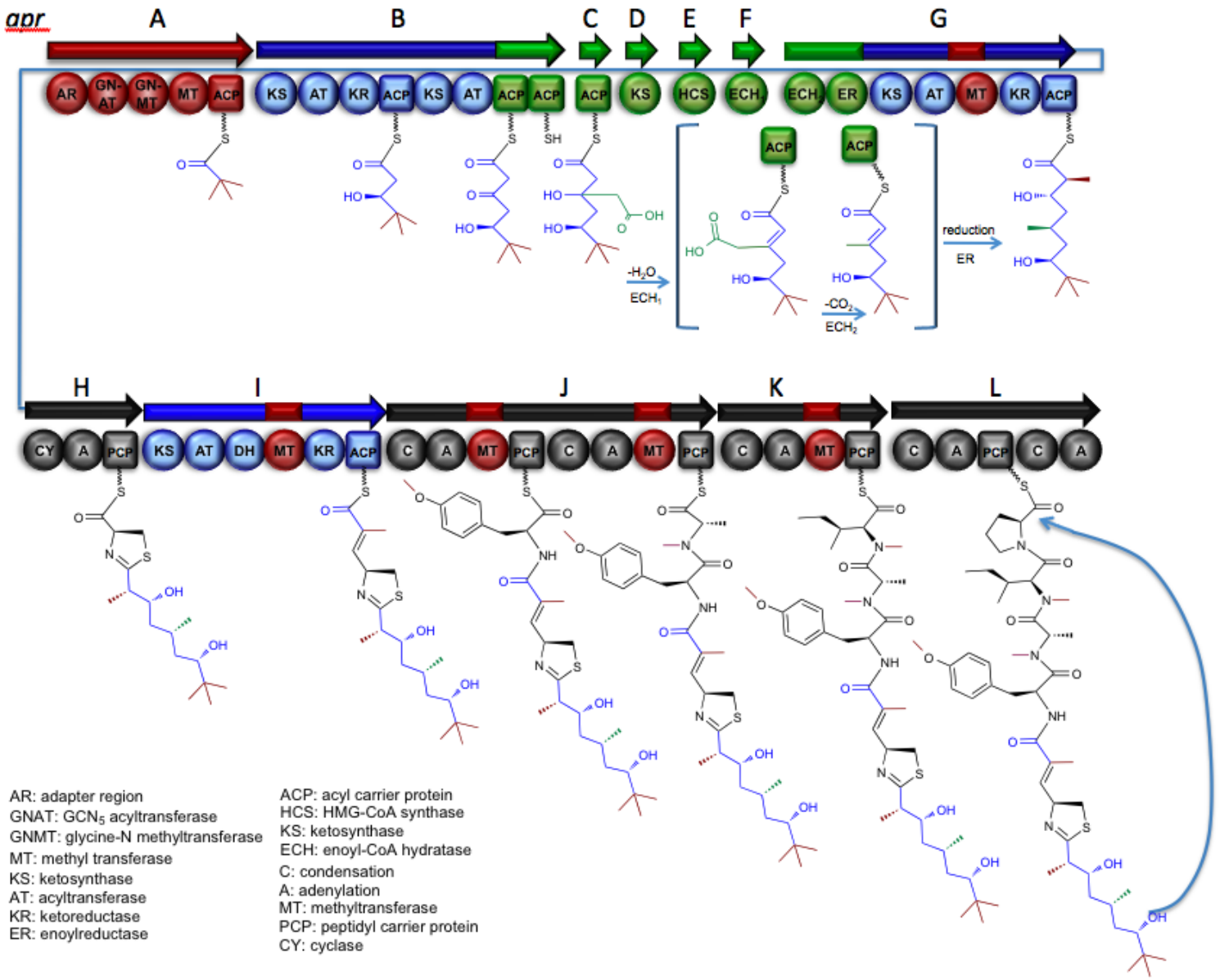


Figure 4.9. Proposed biosynthesisi of apratoxin A. The mechanism of macrocyclization and thoiester release is unclear.

malonyl-CoA (or methylmalonyl-CoA). However, in the apratoxin pathway, the glycine *N*-methyltransferase (GNMT- a SAM dependent methyltransferase known to methylate the nitrogen atom of glycine) (Takata, Huang et al. 2003) would then utilize the newly liberated electrons in a substitution reaction with SAM to methylate the starter unit. The GNAT domain appears to then transfer the methylated substrate to the adjacent holo-ACP where it could undergo a round (or successive rounds) of methylation (Figure 4.10).

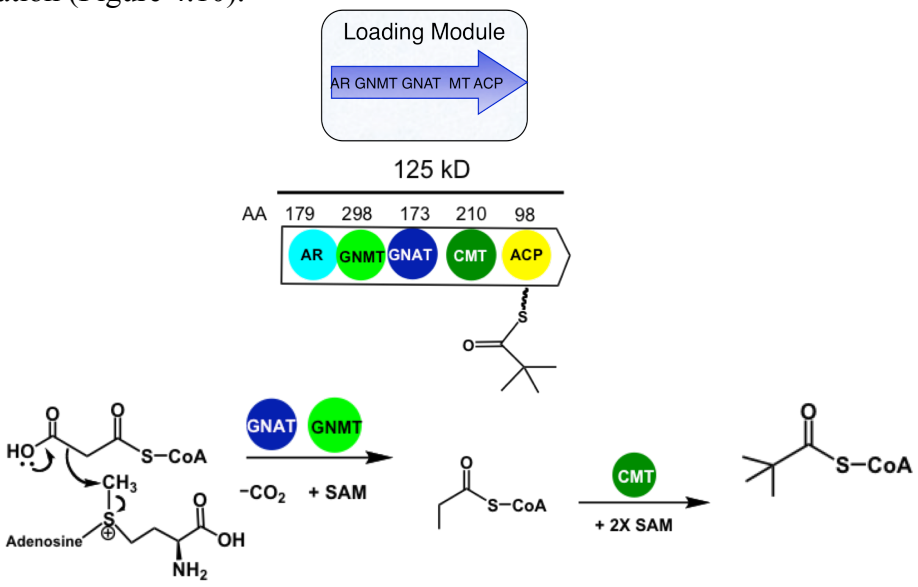


Figure 4.10. Generic mechanistic proposal for the formation of the tri-methylated starter unit.

The final module of AprL is unusual. The condensation domain appears to be catalytically competent, whereas the functionality of the A domain is unclear. Primary sequence analysis indicates this domain is selective towards activating hydroxy-containing amino acids. However, the universally conserved active site

lysine is absent. This module also lacks the obligatory ACP. Coupled to the above observations, conspicuously absent from the apratoxin pathway is the thioesterase (TE) domain that is typically fused to the final module of the assembly line and is required for release of the natural product. The condensation domain could conceivably catalyze the dehydration and macrocyclization of the compound. Another enzyme or domain dedicated to liberation of the tethered intermediate is not readily apparent. Therefore, the precise role of AprL A domain in the context of apratoxin biosynthesis remains unclear until further analysis establishes its biochemical function.

4.3.7 Homologous regions of the *apr*A and *bou*X pathways are highly identical

Two mechanistically intriguing regions of the *apr* pathway, the loading module and the HMG-CoA synthase cassette, have high nucleotide similarity to the homologous genes in the *bou*X pathway. A pairwise sequence alignment of the *apr*A and *bou*X loading modules showed that the two pathways are 97.7 % identical. The 2.3 % difference arises from single nucleotide polymorphisms dispersed across the ~ 3,800 bp locus, and not from major insertions or deletions (Figure 4.11 top). Likewise, a similar analysis of the domains belonging to the HMG-CoA synthase cassette revealed a region of high identity (> 95%) (Figure 4.11 bottom). This pattern suggests that the biosynthetic logic whereby an acetate addition and methyl group retention at a C1 position on the carbon backbone of both apratoxin A and the product of *bou*X is conserved. Also, the two pathways are greater than 90%

identical for the ECH_1, ECH_2 and ER, which suggests a similar modification of the acetate unit to produce a secondary methyl group in each case, thus supporting the hypothesis that the *bou*X pathway is consistent with the production of laingolide or madangolide.

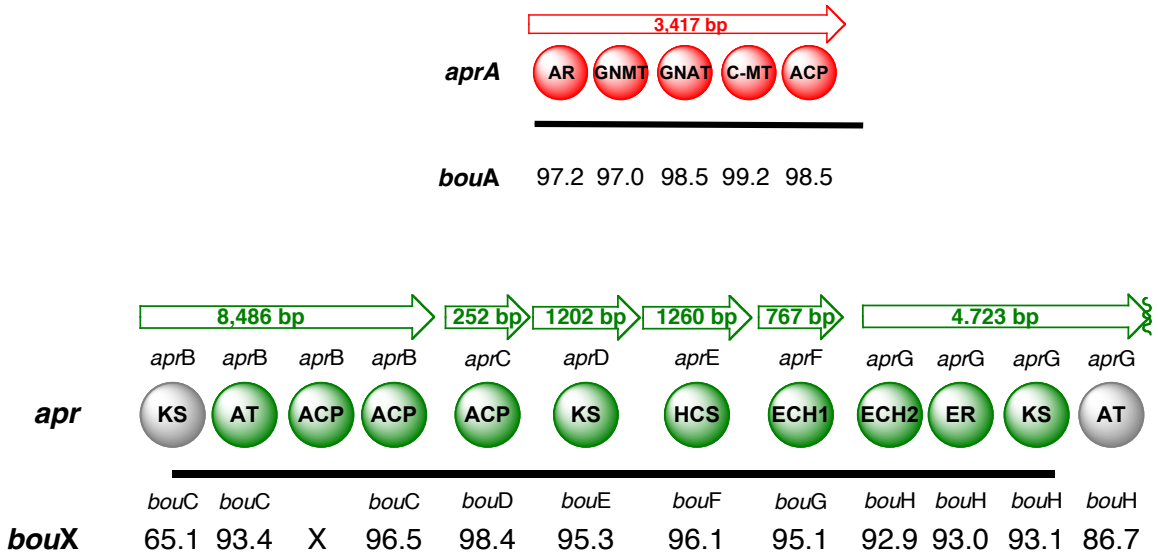


Figure 4.11. Pairwise sequence alignment of the domains in the HMG-CoA synthase cassettes of bouX and the apratoxin cluster (apr). Percent nucleotide identities are shown. **top)** The modules if the loading domains are highly homologous with greater than 97 percent identity. **bottom)** The region of high sequence identity (> 95%, green) flanked by canonical PKS domains, suggests highly conserved mechanistic biochemistry.

4.4. Discussion

The collective chemical profiling efforts of *Lyngbya bouillonii* over the past 20 years from several research groups have yielded eight distinct classes of metabolites, consisting of some 36 different compounds. Here, we report the first successful isolation of a putatively complete biosynthetic gene cluster from this highly productive organism. The apratoxin A gene cluster is of particular interest due in part to the bioactivity and mode of action of the molecule produced. Importantly, the loading domain architecture, β-branching manifold and the offloading and cyclization of the mature compound all represent new mechanistic chemistries that will shed light on and add to the current body of knowledge within the field of NRPS/PKS biosynthesis.

The significance of this work lies in the methodology utilized to surmount two major impediments. One is that due to the close association of filamentous cyanobacteria with interstitial-like heterotrophic bacteria, it is currently nearly impossible to obtain pure cyanobacterial genomic DNA preparations. This occludes any effort to create a non-metagenomic DNA library. The other difficulty is in the similarity of biosynthetic NRPS and PKS gene elements shared amongst the compounds isolated. This blurs gene cluster targeting efforts as previous attempts have failed (unpublished data and Chapter 3). The utility of single cell isolation and whole genome amplification (of non-metagenomic DNA) was both integral and necessary in isolating this gene cluster. While some sequence is lost due to

amplification bias or damage to the single genome copy, as much as 90% of the genome has been recovered from single cell sequencing (Ishoey, Woyke et al. 2008). Ultimately, it was the interplay between the MDA-generated genome and traditional library screening methods that yielded these successful results. Further, this approach can be adapted and applied to other mixed assemblages. It has enabled direct sequencing from environmental bacteria without the need to develop culture methods (Raghunathan, Ferguson et al. 2005; Kvist, Ahring et al. 2007; Marcy, Ishoey et al. 2007; Mussmann, Hu et al. 2007).

4.5. Experimental

***Lyngbya bouillonii* collection and culture.** The *Lyngbya bouillonii* collection was made in Baru, Papua New Guinea at the GPS location: 5°40.473 S and 146° 32.819 E and given the collection number PNG/08/03/2001-10. Samples of the cyanobacterium were hand collected at depths of 30 – 60 ft using SCUBA. Topside, filaments were separated and temporarily placed in growth vials containing local seawater. Following transfer to the laboratory *L. bouillonii* was isolated from contaminating cyanobacteria and other microalgae using previously described techniques (Sitachitta, M·rquez et al. 2000). The samples have been maintained in SWBG11 in a controlled temperature room with 16 hr light/ 8 hr dark cycles.

Intact cell MALDI-TOF MS. MALDI Matrix and sample preparation. Per 1 ml: 70 mg of a mixture of 1:1 a-cyano-hydroxycinnamic and DHB (Universal MALDI matrix from Sigma), 750 µl Acetonitrile, 248 ul milliQ H_20, 2 ul TFA.

Using sterile tweezers, 3-4 filaments (roughly 5-10 μg wet weight) of *L. bouillonii* was placed in a 1.5 ml eppendorf tube. About 1 μl of MALDI matrix solution per 1 μg biomass was added to the tube. After 20-30 seconds, 1 μl of this matrix extract was deposited on a well of a Bruker Microflex MSP 96 Stainless Steel Target Plate. After each spot has dried, the plate was placed in the Bruker Microflex MALDI-TOF equipped with flexControl 3.0.

MALDI-TOF Settings. The instrument and program settings for these experiments were as follows: *General*: Flex-Control Method- RP_pepmix.par. *Processing*: Flexanalysis Method- none, Biotools MS method- none. Laser Power: 30-38 % **Sample Carrier**: nothing **Spectrometer**: On, Ion Source 1- 19.0 0mV, Ion Source 2- 16.40 mV, Lens- 9.45 mV, Reflector 20.00, Pulsed Ion Extraction- 190 ns, Polarity- Positive. Matrix Suppression: Deflection, Suppress up to: *m/z* 500. **Detection**: Mass Range- 500-1600, Detector Gain- Reflector 3.7X. Sample Rate- 2.00 GS/s, Mode- low range, Electronic Gain-Enhanced, 100 mV. Real time Smooth- Off. Spectrometer, Size: 81040, Delay 42968. **Processing Method**: Factory method RP_2465. **Setup**: Mass Range- Low. Laser Frequency- 20 Hz, Autoteaching- off. Instrument Specific Settings: Digitizer- Trigger Level- 2000 mV, Digital Off Linear- 127 cnt, Digital Off, Reflector- 127 cnt. Detector Gain Voltage Offset, Linear- 1300 V, Reflector- 1400 V. Laser Attenuator, Offset -12 %, Range- 30 %, Electronic Gain Button Definitions, Regular: 100 mv (offset lin) 100 mV (offset ref) 200mV/full scale. Enh: 51 mV (offset lin), 51 mV (offset ref) 100 mV/full scale. Highest: 25 mV (offsetlin) 25 mV (offset ref) 50 mV/full scale.

Calibration: Calibration was accomplished using angiotensin II as an external standard. Zoom Range +-1.0%, Peak Assignment Tolerance- User Defined-500 ppm.

DAPI staining of *L. bouillonii.* Individual filaments of *L. bouillonii* previously cultured statically in fresh BG-11 for 30 days were mounted using VECTASHIELD® Mounting Medium with DAPI (Vector Labs). The filaments were visualized on an Olympus IX51 using a DAPI/Hoeschst/AMCA Filter cube (Ex: 350/50, Em: 460/50). Images were captured using an Olympus DP70 camera.

Bacterial Strains and Growth Conditions. *Escherichia coli* strain EPI300 was used in this study as a host for fosmid library construction and amplification of fosmid clones. One Shot® TOP10 chemically competent *E. coli* containing pGEM-T easy was grown overnight in LB (Luria-Bertani) medium with ampicillin at a final concentration of 100 mg/ml. The EPI300 strain of *E. coli* harboring fosmid vector was grown overnight in LB medium with chloramphenicol at a final concentration of 12 mg/ml.

Metagenomic DNA isolation and library construction. DNA from *Lyngbya bouillonii* laboratory culture sample PNG/08/03/2001-10 used for PCR amplification was isolated with a modified genomic DNA protocol from the Wizard Kit (Promega, Madison, WI). High molecular weight (HMW) DNA from cultured PNG/08/03/2001-10 was used for library construction, as previously described. The HMW-DNA was end repaired, size selected, and ligated into the copy control fosmid vector pCC1FOS using protocols provided with CopyControl Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, WI). Plasmid preparations

were carried out using commercial kits (Qiagen, Santa Clarita, CA) and (FosmidMAX™ Epicentre Biotechnologies, Madison, WI). Other basic DNA manipulations, such as restriction digests and ligations, were done using standard methods (Sambrook, Fritsch et al. 1989).

Single cell isolation and multiple displacement amplification. Filaments from *Lyngbya bouillonii* laboratory culture sample PNG/08/03/2001-10 were handled using sterile tweezers and lightly homogenized in sterile growth medium. The homogenate was kept on ice until micromanipulation. Individual cells from a single filament were isolated using mechanical micromanipulation as described previously (Ishoey, Woyke et al. 2008). The micromanipulator (TransferMan NK2; CellTram Vario, Eppendorf) was connected to an inverted microscope (Olympus IX70) and the sample was observed with a 40x objective. The separation of cells from the sheath was achieved using a capillary of approximate 20 μm inner diameter to first push the four neighboring cells out of the sheath followed by capture and washing in sterile PBS-buffer. Finally, the cells were transferred individually to 0.5 μl lysis buffer (400 mM KOH and EDTA) in a 200 μl PCR tube (Eppendorf). Non-template control reactions were setup by transfer of approximately the same amount of washing buffer to tubes containing lysis solution. The tubes were kept on ice until all cells for the experiment were collected. Prior to whole genome amplification, the tubes were incubated at 50°C for 10 minutes in a thermocycler (Eppendorf MasterCycler, Eppendorf). Whole genome amplification (Repli-g UltraFast kit, Qiagen) was carried out in a final volume of 5 μl by addition of 4.5 μl master mix to

all tubes. The incubation time was 8 hours at 30°C and the reaction was terminated by incubation at 65°C for 3 minutes (Eppendorf MasterCycler, Eppendorf). The quality of the amplified DNA was assessed by direct sequencing of 16S rDNA PCR product using primers 27f (5'-TTA GAG TTT GAT CCT GGC TCA G-3') and 1492r (5'-CGG TTA CCT TGT TAC GAC TT-3').

Genome sequencing and assembly. Sequencing of amplified single cell MDA from one single cell and a combination of four individually amplified cells was done by 454 FLX pyrosequencing at the Joint Technology Center (J. Craig Venter Institute, Rockville MD). Approximately 5µg of MDA product was used for 454 FLX library construction according to the recommended procedures of 454 Life Sciences.

The sequences from each one half plate were *de-novo* assembled using the Newbler assembly software supplied by 454 Life Sciences. Our dataset consists of 567,000 unpaired reads generated by 454 FLX platform from DNA of five individually amplified cells of *L. bouillonii*. The DNA from a single cell was amplified using multiple displacement amplification (MDA). The amplification allows genomic sequencing without culturing. The average read length is 250 bp and an expected genome size is 7 to 10 Mb. All reads are longer than 50 bp and only 0.1% of reads are shorter than 100 bp. Each read has a 4 nucleotide TAG prefix, which was used to pool DNA, and about 80% of reads have a large sequence (approx 30 nt) of primers.

***In vivo* clone -and *in silico* contig library screening.** The *Lyngbya bouillonii* gDNA fosmid library, comprised of 1,632 colonies containing approximately 40 kb of insert DNA, were arrayed in 17- 96 well plates. Amplified fosmids from 96 wells were pooled using a multiwell pipet (volume 100 μl/well), and diluted 1:1 with LB broth. The matrix of 17- 96 well plates was therefore reduced to 17 total pools representing the complete library. These pools were used as templates for PCR analysis as described below. Further, each positive plate 12 wells across a row or 8 wells down a column were pooled reducing the 96 well matrix to 16 total pools representing one plate. The pooled fosmids, and the 96 well plate (resealed with fresh sealing tape), were stored at -4°C. At all steps, extra care was taken to avoid cross contamination of samples.

The assembly previously described yielded approximately 5000 contigs > 500 bp. The resulting contigs were uploaded to the personal sequence database (PSD) hosted by Oreon State University (Givan, Sullivan et al. 2007). Multiple database and sequence analysis functions, including: database searching by keyword and BLAST sequence similarity searching protein domain identification and multiple sequence alignments were performed via the PSD for initial contig and domain identification.

PCR Cloning of HMGCS and PKS Gene Homologs. PCR amplification of probe fragments used in this study was performed with Taq DNA-polymerase (Promega) with the manufacturer's suggested concentration of template and primers in an Eppendorf Mastercycler gradient system. Conditions used were as follows: denaturation, 30 s at 94°C; annealing, 30 s at 48°C; extension, 60 s at 72°C; 30 cycles. Degenerate primers were designed based on two conserved sequences found in HMGCS-like and associated genes from PKS pathways (forward primer 5'-

CTNCCNTAYGAYGAYCCCGT-3' and reverse primer 5'-NCKRTGNGCNCCYTTNACCAT-3'). These primers were used to amplify a 900 bp fragment from *L. bouillonii* genomic DNA. The amplicons were cloned into pGEM-T and were DNA sequenced. For amplification of the -ketosynthase domain (KS) fragments from the *L. bouillonii* genomic DNA, previously designed primers KS1Up: 5'-MGI GAR GCI HWI SMI ATG GAY CCI CAR CAI MG-3' KSD1: 5'-GGR TCI CCI ARI SWI GTI CCI GTI CCR TG-3' were used. These primers were used to amplify an approximately 700 bp fragment. The amplicons were cloned into pGEM-T and were DNA sequenced.

Fosmid sequencing. Sanger Sequencing Methods**:** Owing to its high accuracy, versatility and ability to obtain, de novo, and complete DNA sequence information, the Sanger sequencing method has been considered the definitive approach for DNA sequencing. Currently, the Sanger-based whole genome shotgun (WGS) sequencing process, excluding the library construction step, is fully supported by automation and managed as a production "pipeline' with varying degrees of integration between different steps that are fully supported by automated equipment and a laboratory information management system (LIMS). The methods for WGS sequencing are outlined below:

Library construction: Genomic DNA is nebulized to produce fragments with a distribution of approximately 1-25 kB. The DNA fragments are then subject to a process of blunt ending using *Bal* 31. A fill in reaction followed to ensure that the

DNA is completely nick free. The ends are adapterized using *Bst* XI adapters and are then ligated into the in-house pHOS2 vector and transformed into GC10 cells.

The fosmid libraries are constructed using about 1 μg of DNA that is sheared using bead beating to generate cuts in the DNA. The staggered ends or nicks are repaired by filling with dNTPs. A size selection process follows on a pulse field electrophoresis system with lambda ladder to select for 39-40Kb fragments. This DNA is then recovered from a gel and ligated to the blunt ended pCC1FOS vector and packaged into lambda packaging extracts and incubated with the host cells and plated out to get the clones with the insert.

Clone picking and Inoculation: The libraries are transformed, and cells are plated onto large format (16 cm x 16 cm) agar plates. The grown colonies are picked for template preparation using the Qbot or QPix colony-picking robots (Genetix) and inoculated into 384-well blocks containing liquid media and incubated overnight with shaking.

DNA template preparation: The principle method that we used for our template production is the ALPI, a modified alkaline lysis plasmid isolation method. High purity plasmid DNA is prepared from liquid bacterial growth using the DNA purification robotic workstation custom built for the JTC by Thermo CRS. The ALPI method is based on the alkaline lysis miniprep (Sambrook, Fritsch et al. 1989), modified for high-throughput processing in 384 well microtiter plates.

Sequencing reactions: Sequencing protocols are based on the di-deoxy sequencing method (Sanger, Nicklen et al. 1977). To obtain paired sequence reads

from opposite ends of each clone insert, two 384 well cycle sequencing reaction plates are prepared from each plate of template DNA. Sequencing reactions are carried out using Big Dye Terminator chemistry version 3.1 Cycle Sequencing Ready Reaction Kits (Applied Biosystems) and standard M13 or other designated forward and reverse primers. The plates are then transferred to one of the AB3730xl DNA Analyzers for electrophoresis. The current polymers and software allow for 12 runs per day on an AB 3730xl with a set-up time of less than one hour and an achieved average trimmed sequence read length from shotgun libraries of > 800 bp on an AB 3730xl.

4.6 References

Butcher, R. A., F. C. Schroeder, et al. (2007). "The identification of bacillaene, the product of the pksX megacomplex in *Bacillus subtilis*." Proc Natl Acad Sci U S A **104**(5): 1506-9.

Caffrey, P., D. J. Bevitt, et al. (1992). "Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from *Saccharopolyspora erythraea*." FEBS Lett **304**(2-3): 225-8.

Calderone, C. T., D. F. Iwig, et al. (2007). "Incorporation of nonmethyl branches by isoprenoid-like logic: multiple beta-alkylation events in the biosynthesis of myxovirescin A1." Chem Biol **14**(7): 835-46.

Calderone, C. T., W. E. Kowtoniuk, et al. (2006). "Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*." Proc Natl Acad Sci U S A **103**(24): 8977-82.

Chang, Z., N. Sitachitta, et al. (2004). "Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **67**(8): 1356-67.

Cheng, Y. Q., M. Yang, et al. (2007). "Characterization of a gene cluster responsible for the biosynthesis of anticancer agent FK228 in *Chromobacterium violaceum* No. 968." Appl Environ Microbiol **73**(11): 3460-9.

Dean, F. B., S. Hosono, et al. (2002). "Comprehensive human genome amplification using multiple displacement amplification." Proc Natl Acad Sci U S A **99**(8): 5261-6.

Dean, F. B., J. R. Nelson, et al. (2001). "Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification." Genome Res **11**(6): 1095-9.

Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*." Chem Biol **11**(6): 817-33.

El-Sayed, A. K., J. Hothersall, et al. (2003). "Characterization of the mupirocin biosynthesis gene cluster from *Pseudomonas fluorescens* NCIMB 10586." Chem Biol **10**(5): 419-30.

Erhard, M., H. von Dohren, et al. (1997). "Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry." Nat Biotechnol **15**(9): 906-9.

Esquenazi, E., C. Coates, et al. (2008). "Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging." Mol Biosyst **4**(6): 562-70.

Gerwick, W. H., R. C. Coates, et al. (2008). "Giant marine cyanobacteria produce exciting potential pharmaceuticals." Microbe **3**(6): 8.

Givan, S. A., C. M. Sullivan, et al. (2007). "The personal sequence database: a suite of tools to create and maintain web-accessible sequence databases." BMC Bioinformatics **8**: 479.

Grindberg, R. V., C. F. Shuman, et al. (2007). Neurotoxic Alkaloids from Cyanobacteria. Modern Alkaloids. P. O. T.-S. Prof. Ernesto Fattorusso**:** 139-170.

Gu, L., T. W. Geders, et al. (2007). "GNAT-like strategy for polyketide chain initiation." Science **318**(5852): 970-4.

Gu, L., B. Wang, et al. (2009). "Metamorphic enzyme assembly in polyketide diversification." Nature **459**(7247): 731-5.

Gutierrez, M., T. L. Suyama, et al. (2008). "Apratoxin D, a potent cytotoxic cyclodepsipeptide from papua new guinea collections of the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida*." J Nat Prod **71**(6): 1099-103.

Hosono, S., A. F. Faruqi, et al. (2003). "Unbiased whole-genome amplification directly from clinical samples." Genome Res **13**(5): 954-64.

Hrvatin, S. and J. Piel (2007). "Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools." J Microbiol Methods **68**(2): 434-6.

Hube, A. E., B. Heyduck-Soller, et al. (2009). "Phylogenetic classification of heterotrophic bacteria associated with filamentous marine cyanobacteria in culture." Syst Appl Microbiol **32**(4): 256-65.

Ishoey, T., T. Woyke, et al. (2008). "Genomic sequencing of single microbial cells from environmental samples." Curr Opin Microbiol **11**(3): 198-204.

Jenke-Kodama, H., A. Sandmann, et al. (2005). "Evolutionary implications of bacterial polyketide synthases." Mol Biol Evol **22**(10): 2027-39.

Kvist, T., B. K. Ahring, et al. (2007). "Specific single-cell isolation and genomic amplification of uncultured microorganisms." Appl Microbiol Biotechnol **74**(4): 926-35.

Lasken, R. S. (2007). "Single-cell genomic sequencing using multiple displacement amplification." Curr Opin Microbiol **10**(5): 510-6.

Liu, Y., B. K. Law, et al. (2009). "Apratoxin A reversibly inhibits the secretory pathway by preventing cotranslational translocation." Mol Pharmacol.

Luesch, H., W. Y. Yoshida, et al. (2002). "New apratoxins of marine cyanobacterial origin from Guam and Palau." Bioorg Med Chem **10**(6): 1973-8.

Luesch, H., W. Y. Yoshida, et al. (2001). "Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula*." J Am Chem Soc **123**(23): 5418-23.

Marcy, Y., T. Ishoey, et al. (2007). "Nanoliter reactors improve multiple displacement amplification of genomes from single cells." PLoS Genet **3**(9): 1702-8.

Marcy, Y., C. Ouverney, et al. (2007). "Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth." Proc Natl Acad Sci U S A **104**(29): 11889-94.

Mussmann, M., F. Z. Hu, et al. (2007). "Insights into the genome of large sulfur bacteria revealed by analysis of single filaments." PLoS Biol **5**(9): e230.

Piel, J., G. Wen, et al. (2004). "Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase." Chembiochem **5**(1): 93-8.

Raghunathan, A., H. R. Ferguson, Jr., et al. (2005). "Genomic DNA amplification from a single bacterium." Appl Environ Microbiol **71**(6): 3342-7.

Rahman, A. S., J. Hothersall, et al. (2005). "Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series." J Biol Chem **280**(8): 6399-408.

Ramaswamy, A. V., C. M. Sorrels, et al. (2007). "Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **70**(12): 1977-86.

Rounge, T. B., T. Rohrlack, et al. (2009). "A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a Planktothrix rubescens strain." BMC Genomics **10**: 396.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning: a laboratory manual.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.

Simunovic, V., J. Zapp, et al. (2006). "Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases." Chembiochem **7**(8): 1206-20.

Sitachitta, N., B. L. M·rquez, et al. (2000). "Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium *Lyngbya majuscula*." Tetrahedron **56**(46): 9103-9113.

Takata, Y., Y. Huang, et al. (2003). "Catalytic mechanism of glycine N-methyltransferase." Biochemistry **42**(28): 8394-402.

Tidgewell, K., B. R. Clark, et al. (2010). The natural products chemistry of cyanobacteria.

Chapter 5. Biochemical Investigations of β-alkylation in the Jamaicamide Pathway

5.1 Abstract

Of considerable interest recently is the biosynthetic logic whereby β-alkylation occurs within PKS systems to produce methyl- and branched chain polyketides. New ß-alkylation events involving an HMG-CoA synthase (HCS)-like gene and associated tailoring domains embedded within PKS/NRPS pathways have been discovered. To date, several biosynthetic gene clusters for the production of secondary metabolites containing that type of genetic architecture have been reported, such as curacin A (Chang, Sitachitta et al. 2004) and jamaicamide A, B and C (Edwards, Marquez et al. 2004) in *Lyngbya majuscula*, mupirocin in *Pseudomonas fluorescens* (El-Sayed, Hothersall et al. 2003), and bacillaene in *Bacillus subtilis* (Calderone, Kowtoniuk et al. 2006; Butcher, Schroeder et al. 2007) (Figure 5.1). Each have been annotated as having stand alone homologs of the HCS and a set of genes encoding one or more ACPs, a mutant KS with a cys-to-ser active site substitution (KSs) and two homologs of the enoyl-CoA hydratase (ECH_1 and ECH_2) family. It is known that these gene products condense acetyl-CoA with the ß-ketoacyl-S-ACP intermediate of the growing polyketide to eventually result in the addition of the C-2 acetate carbon in the structure. Furthermore, elaboration upon that carbon has yielded diverse functionalities including an α-β unsaturated pendant methyl group in bacillaene and mupirocin, a fully saturated methyl moiety in apratoxin A, a cyclopropyl ring in curacin A, and a vinylic chloride in jamaicamide

A. The biochemical investigations of the genes and enzymes that prescribe the addition of the C2 of acetate onto the polyketide backbone of jamaicamide A with final elaboration of a vinyl chloride group is presented. The focus of this system involves elucidating the activity the HCS-ECH_1-ECH_2-ER motif. Preliminary results reveal differential mechanistic activity of ECH_1 and ECH_2 as well as a highly unexpected chlorinated intermediate.

5.2 Introduction

Marine cyanobacteria exhibit a high production of bioactive and structurally diverse natural products. A number of these secondary metabolites or their derivatives are lead compounds in drug development programs aimed at providing new therapies to treat cancer, bacterial infections, inflammatory responses, and in crop protection to kill harmful microbial pathogens and insects. Isolation and structural analysis of marine and terrestrial cyanobacterial natural products has provided access to a large number of mixed non-ribosomal peptide synthetase/polyketide synthase (NRPS/PKS) systems. The corresponding metabolic systems are comprised of an intriguing set of complex multifunctional proteins that along with allied enzymes generate structurally complex molecules via a modular multi-step process. A full understanding of the molecular mechanisms, catalytic activities, kinetic properties, and substrate specificities within cyanobacterial biosynthetic pathways is just beginning to unfold. The results of this study contribute to the understanding of the specific β-alkylating events in the jamaicamide

pathway. That is that Jam ECH_1 and ECH_2 exhibit full activity with the native, non-chlorinated substrate. However, when presented with the corresponding curacin substrate, the activity of Jam ECH_2 was lost. These results are significant in that speculation of an evolutionary divergence of mechanistic activity is reasonable. This chapter is submitted as a sister chapter to our recently published manuscript on the differential biochemical activity of the Cur and Jam pathways using the native, chlorinated curacin substrates (Gu, Wang et al. 2009).

5.2 Results and Discussion

The pksX cluster, responsible for the production of bacillaene exhibits the characteristic HMG-CoA synthase-like manifold. Calderone *et al.* (2006) observed the following, 1) an upstreatm AT domain (PksC) selectively loaded malonyl-CoA (over acetyl-CoA and methylmalonyl-CoA) and transferred malonyl to the carrier protein, AcpK, and then the PksF-mediated (KSq) decarboxylation 2) the resulting acetate functionality, in a PksG-mediated reaction (HCS), underwent an aldol condensation with acetoacetyl-S-PksL to yield HMG-S-PksL (Figure 5.2A). The metabolic reactions for the downstream enzymes in the curacin A pathway were also reported on. Gu *et al.* (2006) demonstrated that the CurE/CurF ECH_1-ECH_2 enzyme pair catalyzes the successive dehydration and decarboxylation of (S)-HMG-ACP to generate a 3-methylcrotonyl-ACP intermediate for subsequent formation of the cyclopropane ring (Figure 5.2B).

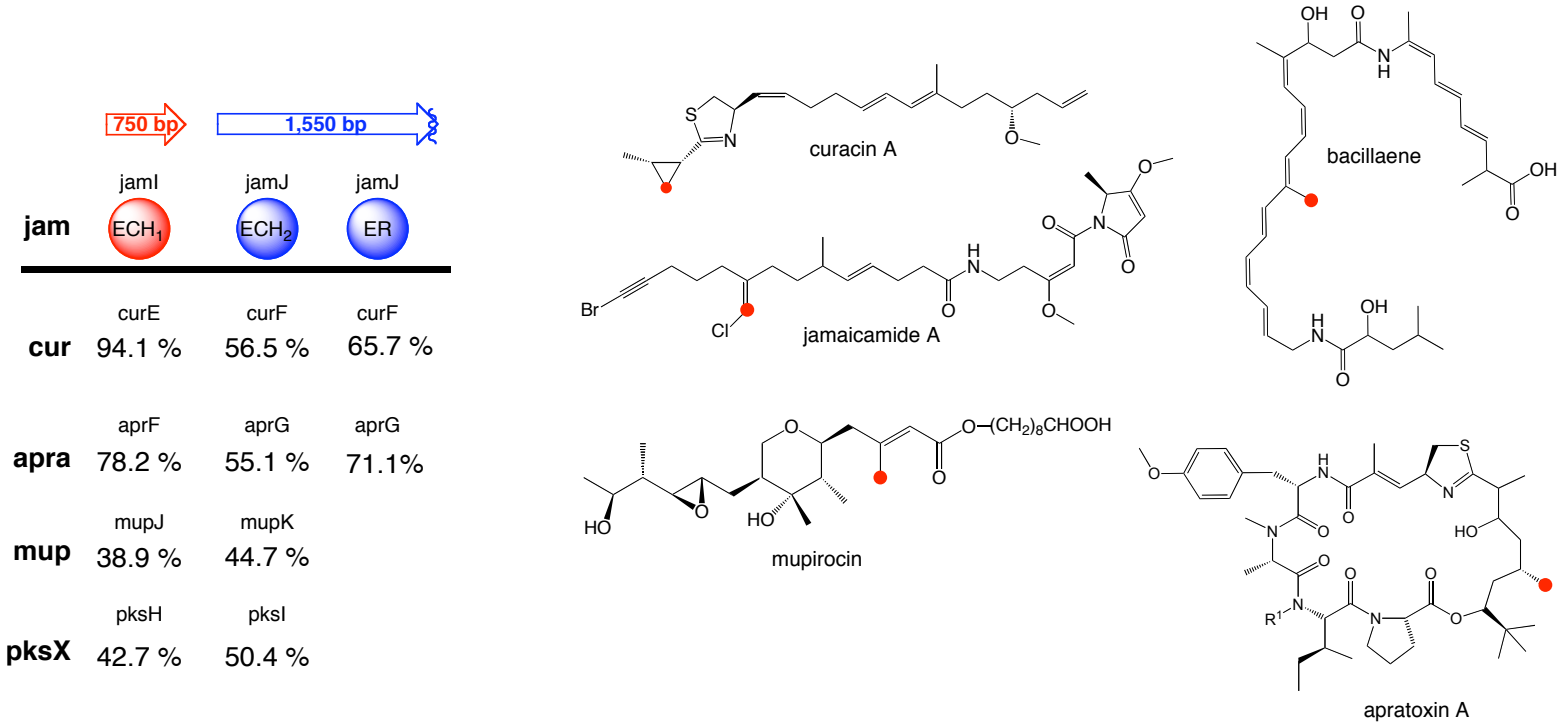


Figure 5.1. Secondary metabolite gene clusters that harbor the HCS manifold. Homologous HCS–like cassettes introduce a carbon from C-2 of acetate (indicated by the red dot) into their respective molecules. Further elaboration by tailoring domains ECH_1, ECH_2, and ER creates diverse functionalities

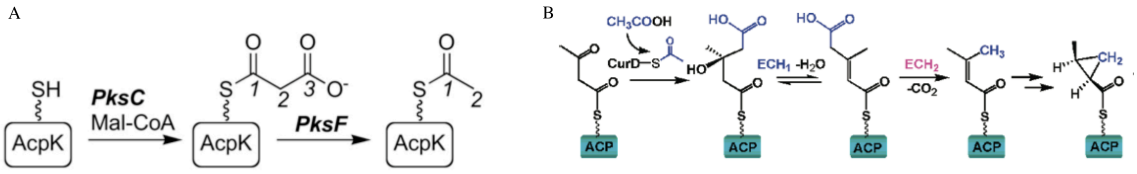


Figure 5.2. A. PksX scheme for loading of malonyl and decarboxylation of malonyl-ACP to acetate, which is then transferred to the HCS domain (Calderone *et al.*, 2006). B. Curacin A scheme for the activity of ECH_1 and ECH_2 for subsequent formation of the cyclopropane ring (Gu, *et al.* 2006)

The Jam pathway β-alkylating gene manifold includes *jamF*, *jamG*, *jamH*, *jamI* and the N terminus of *jamJ*. These gene products were hypothesized to prescribe the addition of the C-2 of acetate onto the polyketide backbone of jamaicamide A with final elaboration of a vinyl chloride group. The corresponding genes in the Cur pathway include *curA*, *curB*, *curC*, *curD*, *curE* and the N terminus of *curF*. Sequence identities of the Jam and Cur homologous genes were extraordinarily high (90-97%). However, the N-terminal regions of jamJ and curF (containing the homologous ECH2s and ERs) were considerably lower (~ 60%). This observation suggested that the underlying mechanistic biochemical activity of the parallel gene cassettes diverged and that the ECH2s were key branch point determinants. Therefore we sought to characterize the biochemical activity of these tailoring enzymes. The focus involved elucidating the activity of the ECH_1-ECH_2-ER (JamI-JamJ) motif (Figure 5.3). We hypothesized that these enzymes were involved in the addition of a C-2 carbon from acetate *via* condensation of acetyl-HCS (**71**) with 8-bromo-3-oxooct-7-y noyl -ACP (**72**) to form

(*R* or *S*)-(**73**) with subsequent dehydration by ECH_1 to form (R or S **74**), and decarboxylation by ECH_2 to form one of two possible intermediates (7**5** or **76**) that differ only in regiochemistry (Figure 5.4). Preliminary results revealed differential mechanistic activity of the Jam and Cur ECH_1^s and ECH_2^s.

5.3.1 Cloning, Overexpression and Characterization of JamF, JamI and JamJ

JamF (ACP), and JamI (ECH_1) were amplified using primers designed to the native start and stop codons, with modified 5' and 3' additions for the *Nde*I and *Xho*I restriction sites. PCR amplicons were subcloned into the pET20b(+) vector with a C-terminal 6X His-tag. Vectors were transformed into the *E.coli* BL21 (DE3) cells. Optimal protein overexpression was obtained at 25° C and induction with 0.4 mM IPTG that yielded the expression of 9 kD (ACP) and 30 kD (ECH_1) proteins that eluted from the soluble fraction was successful.

The N-terminal region of JamJ contains the ECH_2-ER di-domain. The ECH_2 mono-domain and the ECH_2-ER di-domain were successfully amplified using the primers targeting the native start codon and a reverse primer with an engineered stop codon. Proteins were subcloned and overexpressed as previously described. The 30 kD (ECH_2) and 70 kD (ECH_2-ER) soluble proteins were successfully isolated (Figures 5.5 and 5.6).

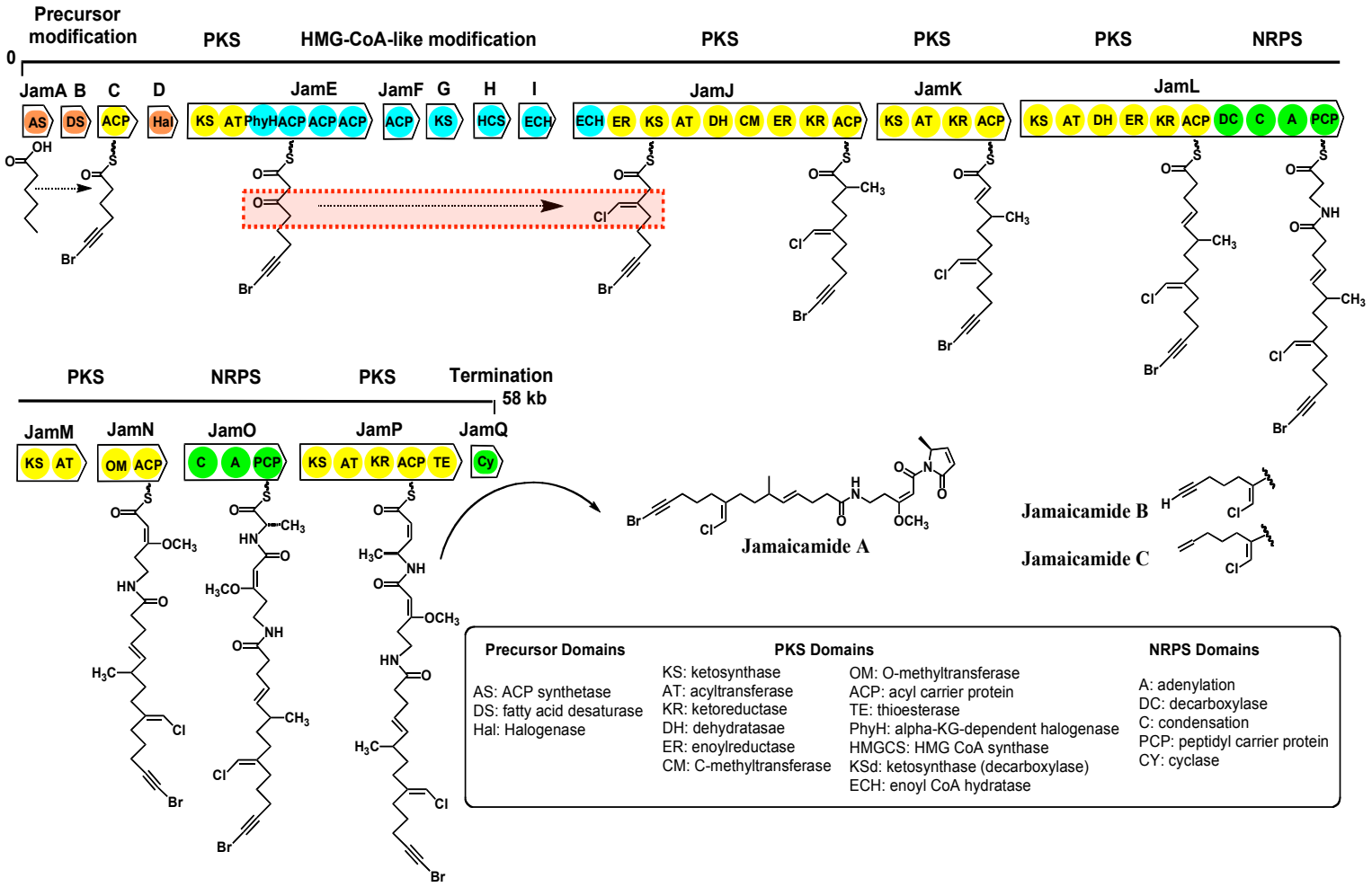


Figure 5.3. Biosynthetic pathway for jamaicamide A, B, and C. The domains associated with HCS-like modifications (red shaded area) were biochemically investigated.

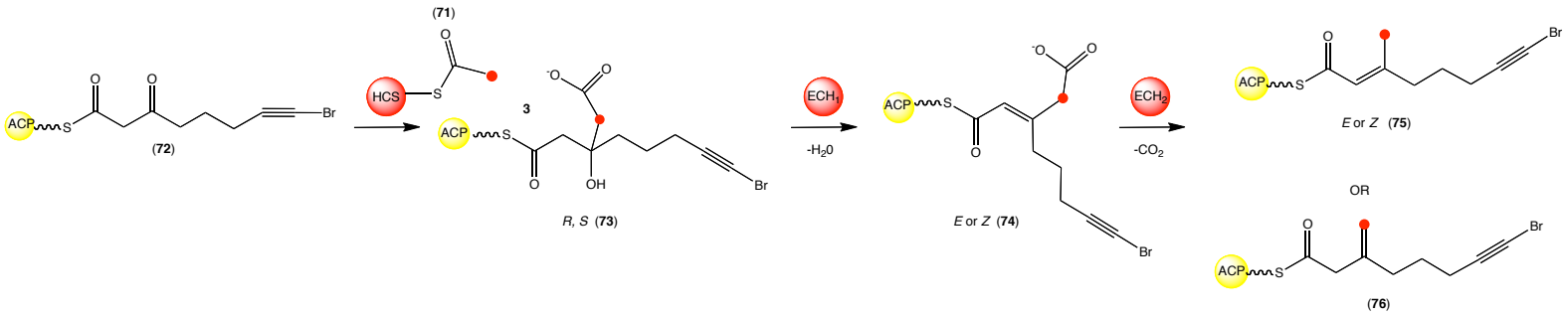


Figure 5.4. Proposed mechanisms of the HCS and associated enzymes in the Jam biosynthetic pathway. HCS introduces a carbon (indicated by red dot) from C-2 of acetate to the polyketide backbone. JamI-ECH$_1$ is responsible for the dehydration of (73)-ACP to the α-β unsaturated (74)-ACP product. ECH$_2$ decarboxylates the substrate to leave the α-β unsaturated methyl or exomethylene functionality.

We attempted to investigate the mono-domain, ER, separately. Overexpression was successful, however the protein remained insoluble as protein aggregates or inclusion bodies. Some generic and commonly encountered reasons for this problem relate to the biophysical parameters of the specific protein such as size, degree of hydrophobicity, charge and identity of the amino acids (Wilkinson and Harrison 1991). It is unclear precisely why our ECH_2 construct was problematic, but one possible explanation is that the protein was not folding correctly. Despite further optimization attempts, such as lowering the incubation temperature from 25 – 18° C, and growing for a period of 18-24 hrs, or changing sonication pulse sequences during cell lysis, we remained unsuccessful. Therefore, we used the 30 kD ECH_2-ER di-domain in the reactions.

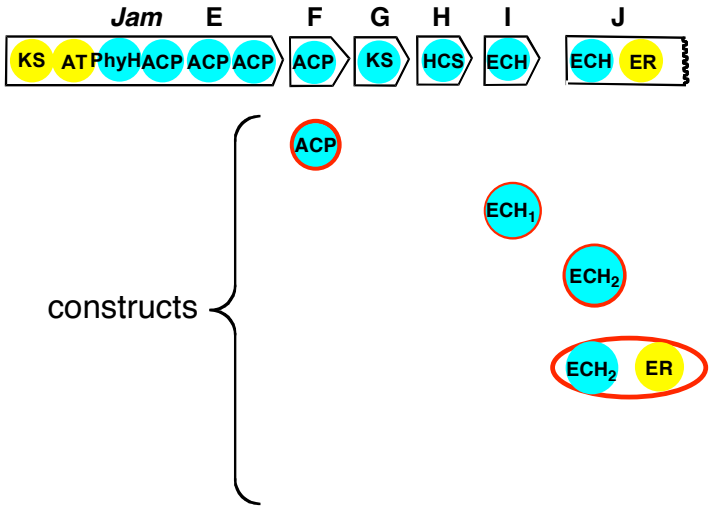


Figure 5.5. Constructs for the cloning and overexpression of JamF, JamK and JamJ. Red circle indicates mono- or di- domain construct.

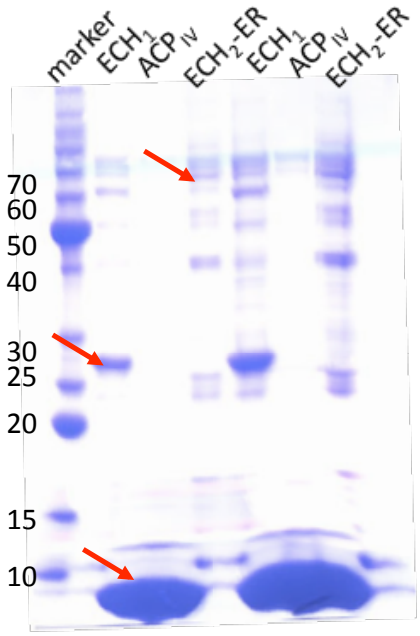


Figure 5.6. Coomassie stained gel showing the verexpression and purification of JamF (ACP), JamI (ECH_1), and JamJ (ECH_2-ER di-domains) (red arrows).

We sought to evaluate the activities of Jam ECH_1 and ECH_2. Based on previous data of the cur enzymes (Gu, Wang et al. 2009) we hypothesized that the ECH_1 catalyzes the dehydration of *R* or *S* (**73**)-ACP (or equivalent substrate) to two possible geometric isomers of (**74**). To quickly determine whether the isolated proteins exhibited biochemical activity, we first used a simplified substrate (and incidentally the same substrate native to the curacin pathway) (R,S)-HMG. Apo JamF was first co-incubated with (R-S)-HMG-CoA and *sfp,* the phosphopantetheinyl transferase enzyme, to yield (R,S)-HMG-holo-JamF. Reactions were set up with the purified protein, ECH_1, and the substrates (*R, S*)-(HMG-ACP). Using FT-MS to assay for the conversion to the dehydration product, we observed an 18 Da loss, consistent with the loss of H_2O leaving

an unsaturated C=C bond (Figure 5.7). The regiochemical nature (*E* or *Z* configuration) of the product however, was not evaluated.

Next, we tested the ability of ECH_2 to catalyze decarboxylation of the ACP-bound substrate from the previous experiment. Products from the ECH_1 reaction were incubated with the purified ECH_2 protein followed by FT-MS for detection of the reaction product. We expected the reaction product in this experiment to be (**75** or **76**), depending on which carbon the enzyme protonates (α or γ). However, the characteristic 62 Da loss was not observed resulting in the conclusion that Jam ECH_2 did not catalyze decarboxylation of the native cur substrate (Figure 5.7).

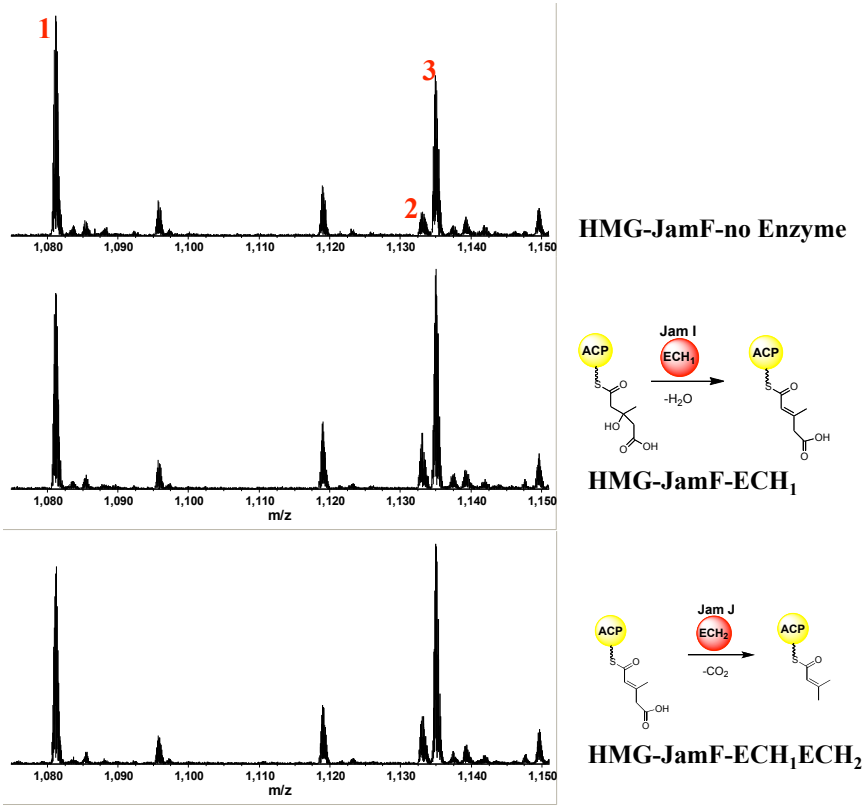


Figure 5.7. Partial FTICR mass spectra for the Jam ECH_1 and ECH_2 reactions with the simplified (R,S)-HMG-JamF substrate. The reactions were incubated for 4 hours.

Using the simplified (R, S)-HMG-ACP substrates to test the bioactivity of ECH_1 and ECH_2 was helpful in optimizing reaction conditions and determining the best way to assay for these enzymes. From these experiments, we found that ECH_1 was active, but ECH_2 was not. To gain full understanding of these enzymes, we tested them with the native, non-chlorinated, jam substrate (**73**)-ACP. The apo-ACP domains were transformed to the holo-ACP by co-incubation with *sfp* and the synthesized substrate, 8-bromo-3-oxooct-7-ynoyl-CoA to produce the β-keto intermediate. As in the previous experiments, a mass change of -18 Da was detected in the ECH_1 reaction (Figure 5.8) indicating successful dehydration of (**73)** to generate (**74**), the presumed precursor for ECH_2. Further, the ECH_2 product peak at -64 Da, is consistent with decarboxylation of the substrate.

We next assessed whether Jam ER could catalyze the saturation of (**74** or **75**). One-pot reactions were used to confirm the products of ECH_1/ECH_2 and ECH_2-ER coupled reactions. After incubation the substrates with purified JamECH_2-ER di-domain protein, we were able to detect a mass shift of +2 using infrared multiphoton dissociation (IRMPD) methods. The mass change of 475.11 to 477.12 indicates a 2 proton addition across the C=C double bond to yield the saturated product (Figure 5.9).

The enzymology of the jamaicamide proteins, ECH_1, ECH_2 and ER has revealed differential catalytic activity of the second ECH domain between the jam and cur pathways. That is that jam ECH_1 and ECH_2 exhibit full activity with the native substrate. However, when presented with the corresponding curacin substrate, the

activity of Jam ECH_2 is lost. The results of this study contribute to the understanding of the specific β-alkylating events in the jamaicamide pathway. These results are significant in that speculation of an evolutionary divergence of mechanistic activity is reasonable. This provides both insight into the molecular genetic understanding of homologous gene cassettes and it provides the foundation for further investigation into the evolutionary history of secondary metabolite biosynthetic pathways.

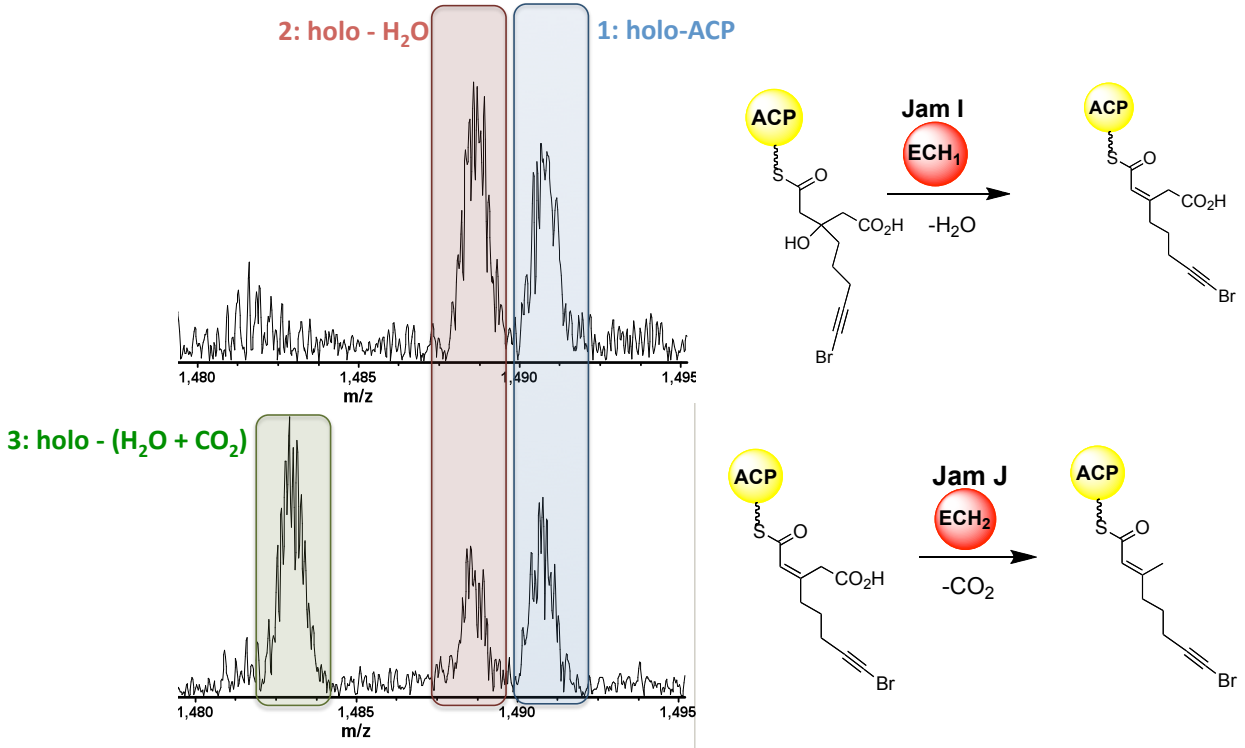


Figure 5.8. Partial FT- mass spectra for the Jam ECH_1 and ECH_2 reactions with the non-chlorinated jamaicamide substrate.

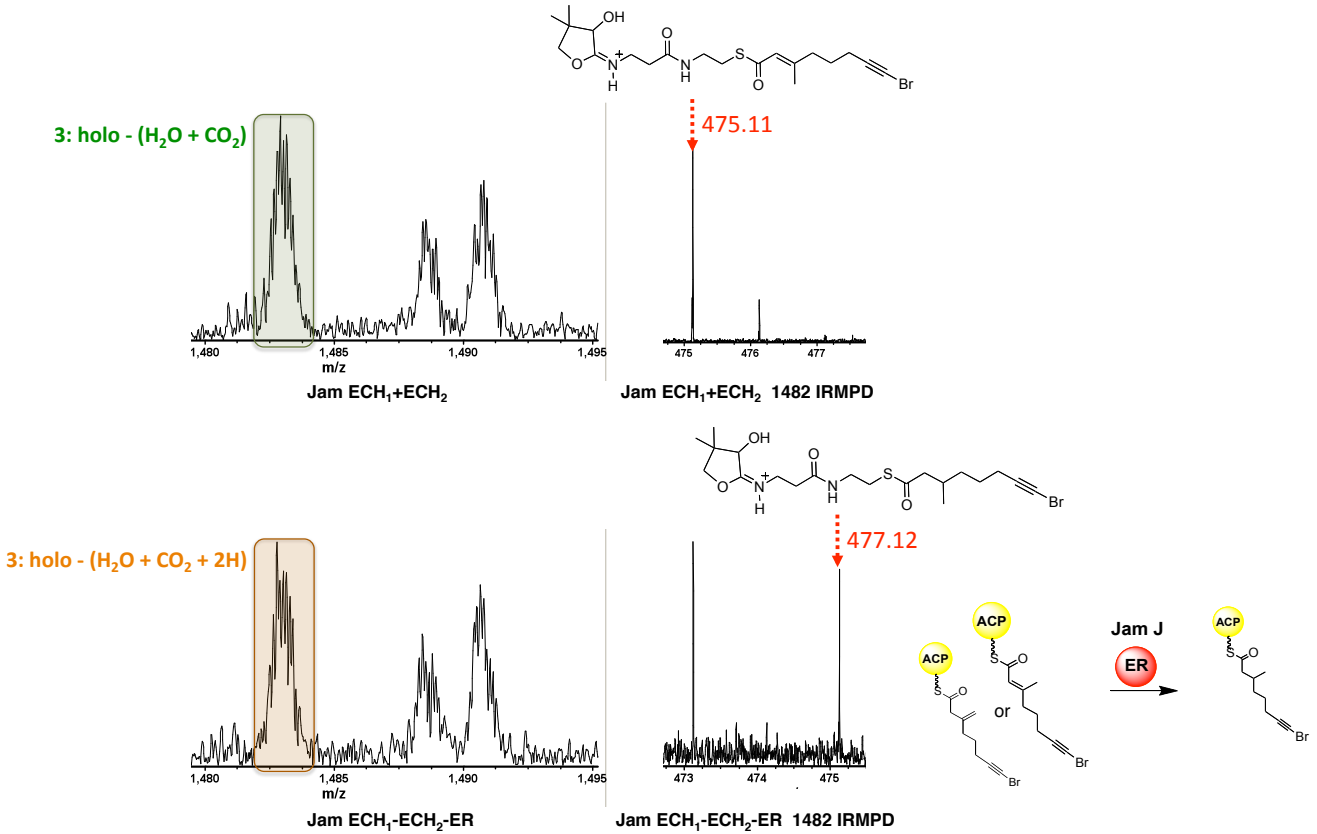


Figure 5.9. Partial FTICR mass spectra for the Jam ER reaction with the jamaicamide substrate and analysis of the holo - (H_2O + CO_2) peak using IRMPD.

5.4 Experimental

Bacterial strains, media and culture conditions. *Escherichia coli* DH5α (Invitrogen) was used for DNA propagation. *Escherichia coli* BL21 (DE3) transformed with the derivatives of pET20b (+) (Novagen) were used for protein overexpression in Luria-Bertani (LB) medium. Ampicillin (100 μg/mL) and carbenecillin (100 μg/mL) were used for the corresponding plasmid construct resistance marker selection in *E. coli* cultures.

Plasmid construction. JamF ACP, JamI ECH_1, JamJ ECH_2 and JamJ ER genes were amplified from fosmid pJam3 (Edwards, Marquez et al. 2004) and inserted into pET20b plasmid using *Nde*I/*Xho*I restriction sites. The following primers were used to PCR amplify the genes of interest.

JamI-ECH1-for CATATGTATTACCAAACCCTAAAA (*Nde*I)
JamI-ECH1-rev CTCGAGGCTTTGCCATGGATATAAC (*Xho*I)
JamJ-ECH2-for CATATGGCAAAGCTGAACTTGAATC (*Nde*I)
JamJ-ECH2-rev CTCGAGCTGCTGGAAGGTTTTTTC (*Xho*I)
JamJ-ER-rev CTCGAGTTCTTTTGTCGATTCTGGTTC (*Xho*I)

Protein purification. Protein purifications were performed at 4˚C. In general, the first step Ni-affinity purifications for all His-tagged proteins were performed under the same conditions. *E. coli* cells were harvested by centrifugation (5,000 g, 20 min, 4˚C), resuspended in ice cold lysis buffer A (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol) and disrupted by sonication on ice. The cell debris was removed by centrifugation at 15,000 g for 50 min. The supernatant was gently removed and loaded onto the 5 ml HisTrap column (GE

Healthcare) preequilibrated with lysis buffer A. The resin was washed successively with ~10 column volumes of washing buffer B (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 20 mM imidazole, 10% glycerol) to remove nonspecifically bound contaminants. Bound proteins were eluted with imidazole by a linear gradient of the elution buffer C (50 mM PBS buffer, pH 8.0, 300 mM NaCl, 250 mM imidazole, 20% glycerol). The fractions of the eluate were checked for purity by SDS-PAGE, pooled and concentrated using Amicon Ultra-15 (30 kDa, 10 kDa or 5 kDa) centrifugal devices (Millipore).

ECH_1 and ECH_2 activity assays. ECH_1 and ECH_2 assays were performed as previously described (Gu, Jia et al. 2006). Briefly, ~50 μM of the substrate was added with 1 μM ECH_1 or ECH_1/ECH_2 in 50 mM Tris-HCl buffer (pH 7.5), and incubated at 30˚C. After quenched by addition of 10% formic acid, the reactions were analyzed by reverse-phase HPLC using Jupiter C4 column. FTICR-MS and IRMPD were applied to detect the products.

ER activity assays. The non-chlorinated substrate was generated by Jam ECH_1/ECH_2. Typically, ER reactions were performed by incubating ~50 μM the substrate with 1 μM ECH_2-ER and 0.5 mM NADPH in 50 mM Tris-HCl buffer (pH 7.5) at 30˚C.

5.5 References

Butcher, R. A., F. C. Schroeder, et al. (2007). "The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis." Proc Natl Acad Sci U S A **104**(5): 1506-9.

Calderone, C. T., W. E. Kowtoniuk, et al. (2006). "Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of Bacillus subtilis." Proc Natl Acad Sci U S A **103**(24): 8977-82.

Chang, Z., N. Sitachitta, et al. (2004). "Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium Lyngbya majuscula." J Nat Prod **67**(8): 1356-67.

Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula." Chem Biol **11**(6): 817-33.

El-Sayed, A. K., J. Hothersall, et al. (2003). "Characterization of the mupirocin biosynthesis gene cluster from Pseudomonas fluorescens NCIMB 10586." Chem Biol **10**(5): 419-30.

Gu, L., J. Jia, et al. (2006). "Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair." J Am Chem Soc **128**(28): 9014-5.

Gu, L., B. Wang, et al. (2009). "Metamorphic enzyme assembly in polyketide diversification." Nature **459**(7247): 731-5.

Wilkinson, D. L. and R. G. Harrison (1991). "Predicting the solubility of recombinant proteins in Escherichia coli." Biotechnology (N Y) **9**(5): 443-8.

Chapter 6. General Conclusions

Natural products represent an unparalleled source of molecular diversity in drug discovery (Shu, 1998). In fact, one study was successful in quantifying a correlation between natural products, drug compounds and combinatorial chemical entities using multivariate analyses (Feher and Schmidt 2003). They found that properties (such as molecular weight, degrees of unsaturation, number or rotatable bonds, etc.) of drugs today more closely resemble the chemical entities found in nature, as opposed to those produced by combinatorial methods.

Moreover, in the areas of cancer and infectious disease, 60% and 75% of new drugs, respectively, originated from natural sources between 1981 and 2002 (Newman, Cragg et al. 2003). Between 2001 and 2005, 23 new drugs derived from natural products were introduced for the treatment of disorders such as bacterial and fungal infections, cancer, diabetes, Alzheimer's disease, and the list goes on (Lam 2007).

Marine cyanobacteria in particular, have a deep history of producing a multitude of secondary metabolites. Nearly 800 compounds have been reported from these organisms (marinLit, 2009). Yet there remains an urgent need to identify new chemotypes as leads for effective drug development in many therapeutic areas (Newman *et al.,* 2007). To understand the biosynthetic origin of these valuable natural products, as well as to produce increased supplies and analog molecules, it is

necessary to access the clustered biosynthetic genes that encode for the assembly enzymes.

Significant advances in the field of natural products drug discovery have been made on this front. Much biotechnological advancement in molecular genetic applications in recent years – ranging from single-cell genomics, meta- and comparative genomics, proteomics, microarray analyses, to MALDI mass spectrometry and analytical spectroscopy – have aided in these efforts. In doing so, the mechanistic chemistry underlying the biosynthesis of bacterial polyketide and non-ribosomal peptide natural products has been greatly enhanced.

To this end, the work I've reported in this dissertation contributes original scientific advancements on several fundamental levels. First, the work I completed on the isolation of apratoxin A yielded sufficient natural product for bioactivity testing. This resulted in the discovery that apratoxin A has a unique and medically relevant profile of cancer cell toxicity. Next, the gene cluster isolation I reported on in chapters 3 and 4 has resulted in newly discovered biosynthetic pathways with predicted new and interesting mechanistic chemistries. In particular, the glycine N-methyltransferases, catalytic enzymes in the HCS manifold, and unique cyclization domain, all represent tailoring enzymes responsible for structural refinement of natural product skeletons and each provides opportunities for natural product diversification. Furthermore, in my efforts to isolate the "elusive" apratoxin A biosynthetic cluster I developed and successfully utilized new strategies in

overcoming the difficulty of metagenomic contaminants. This approach is portable and can be optimized to target natural product pathways in a diversity of other highly complex symbiont communities. Finally, the enzymology of the jamaicamide proteins, ECH_1, ECH_2 and ER has revealed differential catalytic activity of the second ECH domain between the jam and cur pathways. That is that Jam ECH_1 and ECH_2 exhibit full activity with the native substrate. However, when presented with the corresponding curacin substrate, the activity of Jam ECH_2 is lost. The results of this study contribute to the understanding of the specific β-alkylating events in the jamaicamide pathway. These results are significant in that speculation of an evolutionary divergence of mechanistic activity is reasonable. This provides both insight into the molecular genetic understanding of homologous gene cassettes and it provides the foundation for further investigation into the evolutionary history of secondary metabolite biosynthetic pathways. For example, the *jam*, *cur*, *bou*X and *apr* HMG coA synthase-like gene cassettes, code for enzymes that create diverse functionalities. One could compare these gene pathways and individual motifs to answer fundamental questions such as; do these genes propagate by horizontal gene transfer? By what specific mechanisms might they be transferred? How do they change to create new functionalities and thus new molecules?

In addition to these formal chapters I have provided considerable thought, insight and physical efforts toward the genomic sequencing of both the 3L strain of *Lyngbya majuscula* and the apratoxin producing strain of *L. bouillonii*. It is my hope that the increasing trend in chemical, genomic and biosynthetic advancements in

natural products discovery will continue and that the work that I have contributed promotes that trend.

Historically, the field of natural products has not been thought of as technologically limited. However, with the advent of genome sequencing and annotation, we've come to realize that the biosynthetic potential of these organisms far exceeds what we can currently detect using modern methodologies. As more sophisticated means of genome sequencing is developed and more is understood about regulation, biosynthetic capacity, gene transfer and evolution, these technologies will allow us to tap into the vast resources of these biosynthetically robust organisms (Li and Vederas 2009).

6.1 References

Feher, M. and J. M. Schmidt (2003). "Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry." J Chem Inf Comput Sci **43**(1): 218-27.

Lam, K. S. (2007). "New aspects of natural products in drug discovery." Trends in Microbiology **15**(6): 279-289.

Li, J. W. and J. C. Vederas (2009). "Drug discovery and natural products: end of an era or an endless frontier?" Science **325**(5937): 161-5.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural Products as Sources of New Drugs over the Period 1981−2002." Journal of Natural Products **66**(7): 1022-1037.

BIBLIOGRAPHY

Ahmed, A. (2007). "Digoxin and reduction in mortality and hospitalization in geriatric heart failure: importance of low doses and low serum concentrations." J Gerontol A Biol Sci Med Sci **62**(3): 323-9.

Apeldoorn, M. E. V., H. P. V. Egmond, et al. (2007). "Toxins of cyanobacteria." Molecular Nutrition & Food Research **51**(1): 7-60.

Bai, R., R. B. Bates, et al. (2002). "Lyngbyastatin 1 and Ibu-epilyngbyastatin 1: synthesis, stereochemistry, and NMR line broadening." J Nat Prod **65**(12): 1824-9.

Barchi, J. J., R. E. Moore, et al. (1984). "Acutiphycin and 20,21-didehydroacutiphycin, new antineoplastic agents from the cyanophyte *Oscillatoria acutissima*." Journal of the American Chemical Society **106**(26): 8193-8197.

Beerling, D. J. (2009). "Coevolution of photosynthetic organisms and the environment." Geobiology **7**(2): 97-9.

Bergmann, W. and D. C. Burke (1955). "Contributions to the study of marine products. XXXIX. The nucleosides of sponges. III. Spongothymidine and spongouridine. ." The Journal of Organic Chemistry **20**(11): 1501-1507.

Bergmann, W. and R. J. Feeney (1950). "The isolation of a new thymine pentoside from sponges. ." Journal of the American Chemical Society **72**(6): 2809-2810.

Bergmann, W. and R. J. Feeney (1951). "Contributions to the study of marine products. XXXII. The nucleosides of sponges. I. ." The Journal of Organic Chemistry **16**(6): 981-987.

Berman, F. W., W. H. Gerwick, et al. (1999). "Antillatoxin and kalkitoxin, ichthyotoxins from the tropical cyanobacterium Lyngbya majuscula, induce distinct temporal patterns of NMDA receptor-mediated neurotoxicity." Toxicon **37**(11): 1645-8.

Blunt, J. W., B. R. Copp, et al. (2009). "Marine natural products." Nat Prod Rep **26**(2): 170-244.

Boyer, S. L., V. R. Flechtner, et al. (2001). "Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria." Mol Biol Evol **18**(6): 1057-69.

Brun, Y. V. and L. J. Shimkets (2000). Prokaryotic Development, ASM Press.

Burja, A. M., B. Banaigs, et al. (2001). "Marine cyanobacteria-a prolific source of natural products." Tetrahedron **57**(46): 9347-9377.

Butcher, R. A., F. C. Schroeder, et al. (2007). "The identification of bacillaene, the product of the PksX megacomplex in Bacillus subtilis." Proc Natl Acad Sci U S A **104**(5): 1506-9.

Butler, M. S. (2008). "Natural products to drugs: natural product-derived compounds in clinical trials." Natural Product Reports **25**(3): 475-516.

Caffrey, P. (2003). "Conserved amino acid residues correlating with ketoreductase stereospecificity in modular polyketide synthases." Chembiochem **4**(7): 654-7.

Caffrey, P., D. J. Bevitt, et al. (1992). "Identification of DEBS 1, DEBS 2 and DEBS 3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from *Saccharopolyspora erythraea*." FEBS Lett **304**(2-3): 225-8.

Calderone, C. T., D. F. Iwig, et al. (2007). "Incorporation of nonmethyl branches by isoprenoid-like logic: multiple beta-alkylation events in the biosynthesis of myxovirescin A1." Chem Biol **14**(7): 835-46.

Calderone, C. T., W. E. Kowtoniuk, et al. (2006). "Convergence of isoprene and polyketide biosynthetic machinery: isoprenyl-S-carrier proteins in the pksX pathway of *Bacillus subtilis*." Proc Natl Acad Sci U S A **103**(24): 8977-82.

Cane, D. E. and C. T. Walsh (1999). "The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases." Chem Biol **6**(12): R319-25.

Carter, N. J. and S. J. Keam (2007). "Trabectedin: a review of its use in the management of soft tissue sarcoma and ovarian cancer." Drugs **67**(15): 2257-76.

Celmer, W. D. (1965). "Macrolide Stereochemistry. 3. A Configurational Model for Macrolide Antibiotics." J Am Chem Soc **87**: 1801-2.

Chan, Y. A., A. M. Podevels, et al. (2009). "Biosynthesis of polyketide synthase extender units." Nat Prod Rep **26**(1): 90-114.

Chang, Z., P. Flatt, et al. (2002). "The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit." Gene **296**(1-2): 235-47.

Chang, Z., N. Sitachitta, et al. (2004). "Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **67**(8): 1356-67.

Chen, J. and C. J. Forsyth (2003). "Synthesis of the apratoxin 2,4-disubstituted thiazoline via an intramolecular aza-Wittig reaction." Org Lett **5**(8): 1281-3.

Chen, J. and C. J. Forsyth (2003). "Total synthesis of apratoxin A." J Am Chem Soc **125**(29): 8734-5.

Chen, J. and C. J. Forsyth (2004). "Total synthesis of the marine cyanobacterial cyclodepsipeptide apratoxin A." Proc Natl Acad Sci U S A **101**(33): 12067-72.

Cheng, Y. Q., M. Yang, et al. (2007). "Characterization of a gene cluster responsible for the biosynthesis of anticancer agent FK228 in *Chromobacterium violaceum* No. 968." Appl Environ Microbiol **73**(11): 3460-9.

Chin, Y. W., M. J. Balunas, et al. (2006). "Drug discovery from natural sources." The AAPS Journal **8**(2): 239-253.

Copp, J. N. and B. A. Neilan (2006). "The phosphopantetheinyl transferase superfamily: phylogenetic analysis and functional implications in cyanobacteria." Appl Environ Microbiol **72**(4): 2298-305.

Cortes, J., S. F. Haydock, et al. (1990). "An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of Saccharopolyspora erythraea." Nature **348**(6297): 176-178.

Dean, F. B., S. Hosono, et al. (2002). "Comprehensive human genome amplification using multiple displacement amplification." Proc Natl Acad Sci U S A **99**(8): 5261-6.

Dean, F. B., J. R. Nelson, et al. (2001). "Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification." Genome Res **11**(6): 1095-9.

Doi, T., Y. Numajiri, et al. (2006). "Total synthesis of apratoxin A." Org Lett **8**(3): 531-4.

Donadio, S., M. J. Staver, et al. (1991). "Modular organization of genes required for complex polyketide biosynthesis." Science **252**(5006): 675-679.

Duitman, E. H., L. W. Hamoen, et al. (1999). "The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase." Proc Natl Acad Sci U S A **96**(23): 13294-9.

Edwards, D. J. and W. H. Gerwick (2004). "Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase." J Am Chem Soc **126**(37): 11432-3.

Edwards, D. J., B. L. Marquez, et al. (2004). "Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium Lyngbya majuscula." Chem Biol **11**(6): 817-33.

El-Sayed, A. K., J. Hothersall, et al. (2003). "Characterization of the mupirocin biosynthesis gene cluster from Pseudomonas fluorescens NCIMB 10586." Chem Biol **10**(5): 419-30.

Erhard, M., H. von Dohren, et al. (1997). "Rapid typing and elucidation of new secondary metabolites of intact cyanobacteria using MALDI-TOF mass spectrometry." Nat Biotechnol **15**(9): 906-9.

Esquenazi, E., C. Coates, et al. (2008). "Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges via MALDI-TOF imaging." Mol Biosyst **4**(6): 562-70.

Farquhar, J., H. Bao, et al. (2000). "Atmospheric influence of earth's earliest sulfur cycle." Science **289**(5480): 756-758.

Feher, M. and J. M. Schmidt (2003). "Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry." J Chem Inf Comput Sci **43**(1): 218-27.

Finking, R. and M. A. Marahiel (2004). "Biosynthesis of nonribosomal peptides." Annu Rev Microbiol **58**: 453-88.

Fischbach, M. A. and C. T. Walsh (2006). "Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms." Chem Rev **106**(8): 3468-96.

Flatt, P. M., S. J. O'Connell, et al. (2006). "Characterization of the initial enzymatic steps of barbamide biosynthesis." J Nat Prod **69**(6): 938-44.

Gallimore, W. A., D. L. Galario, et al. (2000). "Two complex proline esters from the sea hare *Stylocheilus longicauda*." J Nat Prod **63**(7): 1022-6.

Gehring, A. M., E. DeMoll, et al. (1998). "Iron acquisition in plague: modular logic in enzymatic biogenesis of yersiniabactin by *Yersinia pestis*." **5**(10): 573-586.

Gehring, A. M., I. Mori, et al. (1998). "Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF." Biochemistry **37**(8): 2648-2659.

Gerwick, W. H., R. C. Coates, et al. (2008). "Giant marine cyanobacteria produce exciting potential pharmaceuticals." Microbe **3**(6): 8.

Gerwick, W. H., C. Mrozek, et al. (1989). "Novel cytotoxic peptides from the tropical marine cyanobacterium *Hormothamnion enteromorphoides*. 1. Discovery, isolation and initial chemical and biological characterization of the hormothamnins from wild and cultured material." Experientia **45**(2): 115-21.

Gerwick, W. H., L. T. Tan, et al. (2001). "Nitrogen-containing metabolites from marine cyanobacteria." Alkaloids Chem Biol **57**: 75-184.

Givan, S. A., C. M. Sullivan, et al. (2007). "The personal sequence database: a suite of tools to create and maintain web-accessible sequence databases." BMC Bioinformatics **8**: 479.

Graber, M. A. and W. H. Gerwick (1998). "Kalkipyrone, a toxic gamma-pyrone from an assemblage of the marine cyanobacteria *Lyngbya majuscula* and *Tolypothrix* sp." J Nat Prod **61**(5): 677-80.

Grindberg, R. V., C. F. Shuman, et al. (2007). Neurotoxic Alkaloids from Cyanobacteria. Modern Alkaloids. P. O. T.-S. Prof. Ernesto Fattorusso**:** 139-170.

Gu, L., T. W. Geders, et al. (2007). "GNAT-like strategy for polyketide chain initiation." Science **318**(5852): 970-4.

Gu, L., J. Jia, et al. (2006). "Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: functional identification of a mechanistically diverse enzyme pair." J Am Chem Soc **128**(28): 9014-5.

Gu, L., B. Wang, et al. (2009). "Metamorphic enzyme assembly in polyketide diversification." Nature **459**(7247): 731-5.

Gunasekera, S. P., C. Ross, et al. (2008). "Dragonamides C and D, linear lipopeptides from the marine cyanobacterium brown *Lyngbya polychroa*." J Nat Prod **71**(5): 887-90.

Gutierrez, M., T. L. Suyama, et al. (2008). "Apratoxin D, a potent cytotoxic cyclodepsipeptide from papua new guinea collections of the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida*." J Nat Prod **71**(6): 1099-103.

Hamann, M. T. and P. J. Scheuer (1993). "Kahalalide F: a bioactive depsipeptide from the sacoglossan mollusk *Elysia rufescens* and the green alga *Bryopsis* sp." Journal of the American Chemical Society **115**(13): 5825-5826.

Harrigan, G. G., W. Y. Yoshida, et al. (1998). "Isolation, structure determination, and biological activity of dolastatin 12 and lyngbyastatin 1 from *Lyngbya majuscula*/*Schizothrix calcicola* cyanobacterial assemblages." J Nat Prod **61**(10): 1221-5.

Haydock, S. F., J. F. Aparicio, et al. (1995). "Divergent sequence motifs correlated with the substrate specificity of (methyl)malonyl-CoA: acyl carrier protein transacylase domains in modular polyketide synthases." FEBS Lett **374**(2): 246-8.

Hegazi, M. F., R. T. Borchard, et al. (1976). "Letter: SN2-like transition state for methyl transfer catalyzed by catechol-O-methyl-transferase." J Am Chem Soc **98**(10): 3048-9.

Hoffmann, L. and V. Demoulin (1991). "Marine Cyanophyceae of Papua New Guinea. II. *Lyngbya bouillonii* sp. nov., a remarkable tropical reef-inhabiting blue-green alga." Belg. J. Bot. **124**: 82-88.

Holt, J. G., N. R. Krieg, et al. (1994). Bergey's Manual of Determinative Bacteriology. Baltimore, Md., Williams and Wilkins.

Hopwood, D. A. and D. H. Sherman (1990). "Molecular genetics of polyketides and its comparison to fatty acid biosynthesis." Annu Rev Genet **24**: 37-66.

Hosono, S., A. F. Faruqi, et al. (2003). "Unbiased whole-genome amplification directly from clinical samples." Genome Res **13**(5): 954-64.

Hrvatin, S. and J. Piel (2007). "Rapid isolation of rare clones from highly complex DNA libraries by PCR analysis of liquid gel pools." J Microbiol Methods **68**(2): 434-6.

Hube, A. E., B. Heyduck-Soller, et al. (2009). "Phylogenetic classification of heterotrophic bacteria associated with filamentous marine cyanobacteria in culture." Syst Appl Microbiol **32**(4): 256-65.

Ishoey, T., T. Woyke, et al. (2008). "Genomic sequencing of single microbial cells from environmental samples." Curr Opin Microbiol **11**(3): 198-204.

Jaspars, M. and L. A. Lawton (1998). "Cyanobacteria - a novel source of pharmaceuticals." Current Opinion in Drug Discovery & Development **1**: 8.

Jenke-Kodama, H., A. Sandmann, et al. (2005). "Evolutionary implications of bacterial polyketide synthases." Mol Biol Evol **22**(10): 2027-39.

Jimenez, J. I. and P. J. Scheuer (2001). "New lipopeptides from the caribbean cyanobacterium *Lyngbya majuscula*." J Nat Prod **64**(2): 200-3.

K. Turgay, M. K. M. A. M. (1992). "Four homologous domains in the primary structure of GrsB are related to domains in a superfamily of adenylate-forming enzymes." Molecular Microbiology **6**(4): 529-546.

Kao, C. M., M. McPherson, et al. (1998). "Alcohol stereochemistry in polyketide backbones is controlled by the beta-ketoreductase domains of modular polyketide synthases." Journal of the American Chemical Society **120**(10): 2478-2479.

Kasting, J. F. and J. L. Siefert (2002). "Life and the evolution of earth's atmosphere." Science **296**(5570): 1066-1068.

Khosla, C., R. S. Gokhale, et al. (1999). "Tolerance and specificity of polyketide synthases." Annual Review of Biochemistry **68**(1): 219-253.

Klein, D., J.-C. Braekman, et al. (1999). "Lyngbyapeptin A, a modified tetrapeptide from *Lyngbya bouillonii* (Cyanophyceae)." Tetrahedron Letters **40**(4): 695-696.

Klein, D., J.-C. Braekman, et al. (1996). "Laingolide, a novel 15-membered macrolide from *Lyngbya bouillonii* (cyanophyceae)." Tetrahedron Letters **37**(42): 7519-7520.

Klein, D., J. C. Braekman, et al. (1999). "Madangolide and laingolide A, two novel macrolides from *Lyngbya bouillonii* (Cyanobacteria)." Journal of Natural Products **62**(6): 934-936.

Klein, D., J. C. Braekman, et al. (1997). "Lyngbyaloside, a novel 2,3,4-tri-*O*-methyl-6-deoxy-alpha-mannopyranoside macrolide from *Lyngbya bouillonii* (Cyanobacteria)." Journal of Natural Products **60**(10): 1057-1059.

Kohn, A. J. (1956). "Piscivorous gastropods of the genus *Conus*." Proc Natl Acad Sci U S A **42**(3): 168-71.

Kvist, T., B. K. Ahring, et al. (2007). "Specific single-cell isolation and genomic amplification of uncultured microorganisms." Appl Microbiol Biotechnol **74**(4): 926-35.

Kwan, D. H., Y. Sun, et al. (2008). "Prediction and manipulation of the stereochemistry of enoylreduction in modular polyketide synthases." **15**(11): 1231-1240.

Lai, J. R., A. Koglin, et al. (2006). "Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis." Biochemistry **45**(50): 14869-79.

Lam, K. S. (2007). "New aspects of natural products in drug discovery." Trends in Microbiology **15**(6): 279-289.

Lasken, R. S. (2007). "Single-cell genomic sequencing using multiple displacement amplification." Curr Opin Microbiol **10**(5): 510-6.

Li, J. W. and J. C. Vederas (2009). "Drug discovery and natural products: end of an era or an endless frontier?" Science **325**(5937): 161-5.

Liu, Y., B. K. Law, et al. (2009). "Apratoxin A reversibly inhibits the secretory pathway by preventing cotranslational translocation." Mol Pharmacol.

Luesch, H., S. K. Chanda, et al. (2006). "A functional genomics approach to the mode of action of apratoxin A." Nat Chem Biol **2**(3): 158-67.

Luesch, H., P. G. Williams, et al. (2002). "Ulongamides A-F, new beta-amino acid-containing cyclodepsipeptides from Palauan collections of the marine cyanobacterium *Lyngbya* sp." J Nat Prod **65**(7): 996-1000.

Luesch, H., W. Y. Yoshida, et al. (2002). "Lyngbyaloside B, a new glycoside macrolide from a Palauan marine cyanobacterium, *Lyngbya* sp." J Nat Prod **65**(12): 1945-8.

Luesch, H., W. Y. Yoshida, et al. (1999). "Lyngbyastatin 2 and norlyngbyastatin 2, analogues of dolastatin G and nordolastatin G from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **62**(12): 1702-6.

Luesch, H., W. Y. Yoshida, et al. (2000). "Apramides A-G, novel lipopeptides from the marine cyanobacterium Lyngbya majuscula." J Nat Prod **63**(8): 1106-12.

Luesch, H., W. Y. Yoshida, et al. (2000). "Isolation and structure of the cytotoxin lyngbyabellin B and absolute configuration of lyngbyapeptin A from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **63**(10): 1437-9.

Luesch, H., W. Y. Yoshida, et al. (2002). "New apratoxins of marine cyanobacterial origin from Guam and Palau." Bioorg Med Chem **10**(6): 1973-8.

Luesch, H., W. Y. Yoshida, et al. (2002). "Structurally diverse new alkaloids from Palauan collections of the apratoxin-producing marine cyanobacterium *Lyngbya* sp." Tetrahedron **58**(39): 7959-7966.

Luesch, H., W. Y. Yoshida, et al. (2001). "Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula*." J Am Chem Soc **123**(23): 5418-23.

Luesch, H., W. Y. Yoshida, et al. (2000). "Isolation, structure determination, and biological activity of Lyngbyabellin A from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **63**(5): 611-5.

Ma, D., B. Zou, et al. (2006). "Total synthesis of the cyclodepsipeptide apratoxin A and its analogues and assessment of their biological activities." Chemistry **12**(29): 7615-26.

Marcy, Y., T. Ishoey, et al. (2007). "Nanoliter reactors improve multiple displacement amplification of genomes from single cells." PLoS Genet **3**(9): 1702-8.

Marcy, Y., C. Ouverney, et al. (2007). "Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth." Proc Natl Acad Sci U S A **104**(29): 11889-94.

Marquez, B. L., K. S. Watts, et al. (2002). "Structure and absolute stereochemistry of hectochlorin, a potent stimulator of actin assembly." J Nat Prod **65**(6): 866-71.

Matthew, S., C. Ross, et al. (2007). "Lyngbyastatin 4, a dolastatin 13 analogue with elastase and chymotrypsin inhibitory activity from the marine cyanobacterium *Lyngbya confervoides*." J Nat Prod **70**(1): 124-7.

Matthew, S., P. J. Schupp, et al. (2008). "Apratoxin E, a cytotoxic peptolide from a guamanian collection of the marine cyanobacterium *Lyngbya bouillonii*." J Nat Prod **71**(6): 1113-6.

McCleskey, E. W., A. P. Fox, et al. (1987). "Omega-conotoxin: direct and persistent blockade of specific types of calcium channels in neurons but not muscle." Proc Natl Acad Sci U S A **84**(12): 4327-31.

McDaniel, R., C. M. Kao, et al. (1997). "Engineered intermodular and intramodular polyketide synthase fusions." **4**(9): 667-674.

Mercer, A. C. and M. D. Burkart (2007). "The ubiquitous carrier protein- a window to metabolite biosynthesis." Nat Prod Rep **24**(4): 750-73.

Middeldorf, R. and W. Montigny (1951). "Examination of quinine for accessory cinchona alkaloids." Pharmazie **6**(3): 94-8.

Moldenhauer, J., X. H. Chen Rainer, et al.

Mootz, H. D. and M. A. Marahiel (1997). "The tyrocidine biosynthesis operon of *Bacillus brevis*: complete nucleotide sequence and biochemical characterization of functional internal adenylation domains." J. Bacteriol. **179**(21): 6843-6850.

Mussmann, M., F. Z. Hu, et al. (2007). "Insights into the genome of large sulfur bacteria revealed by analysis of single filaments." PLoS Biol **5**(9): e230.

Mutou, T., T. Kondo, et al. (1996). "Isolation and stereostructures of dolastatin G and nordolastatin G, cytotoxic 35-membered cyclodepsipeptides from the Japanese sea hare *Dolabella auricularia*." J Org Chem **61**(18): 6340-6345.

NA (2006). "Trabectedin: ecteinascidin 743, ecteinascidin-743, ET 743, ET-743, NSC 684766." Drugs R D **7**(5): 317-28.

Newman, D. J. and G. M. Cragg (2004). "Advanced preclinical and clinical trials of natural products and related compounds from marine sources." Curr Med Chem **11**(13): 1693-713.

Newman, D. J. and G. M. Cragg (2004). "Marine natural products and related compounds in clinical and advanced preclinical trials." J Nat Prod **67**(8): 1216-38.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural products as sources of new drugs over the period 1981 - 2002." Journal of Natural Products **66**(7): 1022-1037.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural products as sources of new drugs over the period 1981-2002." Journal of Natural Products **66**: 1022-1037.

Newman, D. J., G. M. Cragg, et al. (2003). "Natural Products as Sources of New Drugs over the Period 1981−2002." Journal of Natural Products **66**(7): 1022-1037.

Nogle, L. M. and W. H. Gerwick (2002). "Somocystinamide A, a novel cytotoxic disulfide dimer from a Fijian marine cyanobacterial mixed assemblage." Org Lett **4**(7): 1095-8.

Nubel, U., F. Garcia-Pichel, et al. (1997). "PCR primers to amplify 16S rRNA genes from cyanobacteria." Appl Environ Microbiol **63**(8): 3327-32.

Olivera, B. M. (2000). omega-Conotoxin MVIIA: from marine snail venom to analgesic drug. Drugs from the Sea. N. Fusetani, Karger**:** 74-85.

Orjala, J. and W. H. Gerwick (1996). "Barbamide, a chlorinated metabolite with molluscicidal activity from the caribbean cyanobacterium *Lyngbya majuscula*." J Nat Prod **59**(4): 427-30.

Pettit, G. R., C. L. Herald, et al. (1982). "Isolation and structure of bryostatin 1." Journal of the American Chemical Society **104**(24): 6846-6848.

Pettit, G. R., Y. Kamano, et al. (1989). "Isolation and structure of the cytostatic linear depsipeptide dolastatin 15." The Journal of Organic Chemistry **54**(26): 6005-6006.

Piel, J., G. Wen, et al. (2004). "Unprecedented diversity of catalytic domains in the first four modules of the putative pederin polyketide synthase." Chembiochem **5**(1): 93-8.

Raghunathan, A., H. R. Ferguson, Jr., et al. (2005). "Genomic DNA amplification from a single bacterium." Appl Environ Microbiol **71**(6): 3342-7.

Rahman, A. S., J. Hothersall, et al. (2005). "Tandemly duplicated acyl carrier proteins, which increase polyketide antibiotic production, can apparently function either in parallel or in series." J Biol Chem **280**(8): 6399-408.

Ramaswamy, A. V., C. M. Sorrels, et al. (2007). "Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*." J Nat Prod **70**(12): 1977-86.

Raskin, I. (1992). "Role of salicylic-acid in plants." Annual Review of Plant Physiology and Plant Molecular Biology **43**: 439-463.

Rausch, C., T. Weber, et al. (2005). "Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs)." Nucleic Acids Res **33**(18): 5799-808.

Reid, R., M. Piagentini, et al. (2003). "A model of structure and catalysis for ketoreductase domains in modular polyketide synthases." Biochemistry **42**(1): 72-79.

Rinehart, K. L., J. B. Gloer, et al. (1981). "Structures of the didemnins, antiviral and cytotoxic depsipeptides from a Caribbean tunicate." Journal of the American Chemical Society **103**(7): 1857-1859.

Rinehart, K. L., T. G. Holt, et al. (1990). "Bioactive compounds from aquatic and terrestrial sources." J Nat Prod **53**(4): 771-92.

Rippka, R. (1988). "Recognition and identification of cyanobacteria." Methods Enzymology **167**: 28-67.

Rippka, R. J., J. Deruelles, et al. (1981). Provisional generic assignments for cyanobacteria in pure culture. The Prokaryotes. H. G. Schlegel. Heidelberg, Springer-Verlag. **1**.

Rounge, T. B., T. Rohrlack, et al. (2009). "A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a Planktothrix rubescens strain." BMC Genomics **10**: 396.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning: a laboratory manual.

Sanger, F., S. Nicklen, et al. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.

Schmidt, B. M., D. M. Ribnicky, et al. (2007). "Revisiting the ancient concept of botanical therapeutics." Nat Chem Biol **3**(7): 360-366.

Schwarzer, D., R. Finking, et al. (2003). "Nonribosomal peptides: from genes to products." Nat Prod Rep **20**(3): 275-87.

Shen, B. (2003). "Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms." Curr Opin Chem Biol **7**(2): 285-95.

Shen, S., P. Zhang, et al. (2009). "Cyclodepsipeptide toxin promotes the degradation of Hsp90 client proteins through chaperone-mediated autophagy." J Cell Biol.

Simunovic, V., J. Zapp, et al. (2006). "Myxovirescin A biosynthesis is directed by hybrid polyketide synthases/nonribosomal peptide synthetase, 3-hydroxy-3-methylglutaryl-CoA synthases, and trans-acting acyltransferases." Chembiochem **7**(8): 1206-20.

Sitachitta, N., B. L. M·rquez, et al. (2000). "Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium *Lyngbya majuscula*." Tetrahedron **56**(46): 9103-9113.

Smith, S. and S. C. Tsai (2007). "The type I fatty acid and polyketide synthases: a tale of two megasynthases." Nat Prod Rep **24**(5): 1041-72.

Sone, H., T. Kondo, et al. (1995). "Dolabellin, a cytotoxic bisthiazole metabolite from the sea hare *Dolabella auricularia*: structural determination and synthesis." The Journal of Organic Chemistry **60**(15): 4774-4781.

Stachelhaus, T., H. D. Mootz, et al. (1999). "The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases." Chem Biol **6**(8): 493-505.

Takata, Y., Y. Huang, et al. (2003). "Catalytic mechanism of glycine N-methyltransferase." Biochemistry **42**(28): 8394-402.

Tan, L. T., B. L. Marquez, et al. (2002). "Lyngbouilloside, a novel glycosidic macrolide from the marine cyanobacterium *Lyngbya bouillonii*." J Nat Prod **65**(6): 925-8.

Taori, K., S. Matthew, et al. (2007). "Lyngbyastatins 5-7, potent elastase inhibitors from Floridian marine cyanobacteria, *Lyngbya* spp." J Nat Prod **70**(10): 1593-600.

Thacker, R. W. and V. J. Paul (2004). "Morphological, chemical, and genetic diversity of tropical marine cyanobacteria *Lyngbya* spp. and *Symploca* spp. (Oscillatoriales)." Appl Environ Microbiol **70**(6): 3305-12.

Thajuddin, N. and G. Subramanian (2005). "Cyanobacterial biodiversity and potential applications in biotechnology." Current Science **89**(1): 47-57.

Tidgewell, K., B. R. Clark, et al. (2010). The natural products chemistry of cyanobacteria.

Uemura, Y. H. a. D. (1986). "Halichondrins—antitumor polyether macrolides from a marine sponge " Pure Appl. Chem. **58**: 701-710.

Vaillancourt, F. H., E. Yeh, et al. (2006). "Nature's inventory of halogenation catalysts: oxidative strategies predominate." Chem Rev **106**(8): 3364-78.

Valeriote, F., C. K. Grieshaber, et al. (2002). "Discovery and development of anticancer agents from plants." J Exp Ther Oncol **2**(4): 228-36.

Valoti, G., M. I. Nicoletti, et al. (1998). "Ecteinascidin-743, a new marine natural product with potent antitumor activity on human ovarian carcinoma xenografts." Clin Cancer Res **4**(8): 1977-83.

Van Dyke, H. B. and R. C. Li (1935). "A study of the standardization of digitalis. Ii. The relationship between laboratory methods of assay and potency as determined by experimental cumulative poisoning and clinical standardization." J Clin Invest **14**(6): 733-7.

Van Wagoner, R. M., A. K. Drummond, et al. (2007). "Biogenetic diversity of cyanobacterial metabolites." Adv Appl Microbiol **61**: 89-217.

Versluis, M., B. Schmitz, et al. (2000). "How snapping shrimp snap: through cavitating bubbles." Science **289**(5487): 2114-2117.

Weber, G., K. Schörgendorfer, et al. (2004). "The peptide synthetase catalyzing cyclosporine production in *Tolypocladium niveum* is encoded by a giant 45.8-kilobase open reading frame " Current Genetics **26**(2).

Wilkinson, D. L. and R. G. Harrison (1991). "Predicting the solubility of recombinant proteins in Escherichia coli." Biotechnology (N Y) **9**(5): 443-8.

Williams, P. G., H. Luesch, et al. (2003). "Continuing studies on the cyanobacterium *Lyngbya* sp.: isolation and structure determination of 15-norlyngbyapeptin A and lyngbyabellin D." J Nat Prod **66**(5): 595-8.

Williams, P. G., R. E. Moore, et al. (2003). "Isolation and structure determination of lyngbyastatin 3, a lyngbyastatin 1 homologue from the marine cyanobacterium *Lyngbya majuscula* Determination of the configuration of the 4-amino-2,2-dimethyl-3-oxopentanoic acid unit in majusculamide C,

dolastatin 12, lyngbyastatin 1, and lyngbyastatin 3 from cyanobacteria." J Nat Prod **66**(10): 1356-63.

Woodard, R. W., M. D. Tsai, et al. (1980). "Stereochemical course of the transmethylation catalyzed by catechol O-methyltransferase." J Biol Chem **255**(19): 9124-7.

Ziegler, H. W., T. H. Beasley, Sr., et al. (1975). "Simultaneous assay for six alkaloids in opium, using high-performance liquid chromatography." J Assoc Off Anal Chem **58**(5): 888-97.