UNIVERSITY OF CALIFORNIA, SAN DIEGO

Regulatory, genetic, and genomic investigations of natural products biosynthesis in marine cyanobacteria

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

in

Oceanography

by

Adam Christopher Jones

Committee in charge:

William Gerwick, Chair
Michael Burkart
Pieter Dorrestein
Lena Gerwick
Bradley Moore
Brian Palenik

2011

The dissertation of Adam Christopher Jones is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2011

DEDICATION

My doctoral dissertation is dedicated to my parents. Thank you so much for all you have done for me and the constant support, guidance, and love you provide.

My doctoral dissertation is dedicated to my wife, Carla. Thank you for being with me on this journey and for making me a better person. I love you.

EPIGRAPH

Our deepest fear is not that we are inadequate. Our deepest fear is that we are powerful beyond measure. It is our light, not our darkness that most frightens us. We ask ourselves, who am I to be brilliant, gorgeous, talented and fabulous? Actually, who are you not to be? You are a child of God. Your playing small doesn't serve the world. There is nothing enlightened about shrinking so that other people won't feel insecure around you. We were born to make manifest the glory of God that is within us. It is not just in some of us: it's in everyone. And when we let our own light shine, we unconsciously give other people permission to do the same. As we are liberated from our own fear, our presence automatically liberates others.

Nelson Mandela, 1918 -

Walk away quietly in any direction and taste the freedom of the mountaineer. Camp out among the grasses and gentians of glacial meadows, in craggy garden nooks full of Nature's darlings. Climb the mountains and get their good tidings. Nature's peace will flow into you as sunshine flows into trees. The winds will blow their own freshness into you, and the storms their energy, while cares will drop off like autumn leaves. As age comes on, one source of enjoyment after another is closed, but Nature's sources never fail.

The grand show is eternal. It is always sunrise somewhere; the dew is never dried all at once; a shower is forever falling; vapor is ever rising. Eternal sunrise, eternal dawn and gloaming, on sea and continents and islands, each in its turn,
as the round Earth rolls.

John Muir, 1838 - 1914

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

A – Adenylation domain

ATP – Adenosine triphosphate

BCA – Bicinchoninic acid

BLAST – Basic local alignment search tool

bp – Base pair

C – Condensation domain

CCA – Complementary chromatic adaptation

cDNA – Complementary DNA

ECH – Enoyl coA-hydratase

EMSA – Electromobility shift assay

ER – Enoyl reductase

FW – Freshwater

GNAT - GCN5-related *N*-acetyltransferase

GST – Glutathione S-transferase

HMG – 3-hydroxy-3-methylglutaryl

kDa – KiloDalton

KS – Ketosynthase

LC-MS – Liquid chromatography – Mass spectrometry

MALDI – Matrix assisted laser desorption ionization

MS – Mass spectrometry

MudPIT - Multidimensional protein identification analysis

MWCO – Molecular weight cut-off

NRPS – Non-ribosomal peptide synthetase

ONPG – Ortho-nitrophenyl-β-galactoside

ORF – Open reading frame

PCP – Peptidyl carrier protein

PDA – Photodiode array

phyH - Phytanoyl-CoA dioxygenase

PKS – Polyketide synthase

PPi – Pyrophosphate

PPTase – Phosphopantetheinylase

RACE – Rapid amplification of cDNA ends

RNAP – RNA polymerase

RPM – Revolutions per minute

RT-PCR – Reverse transcription polymerase chain reaction

SAM – S-adenosyl methionine

SARP – *Streptomyces* antibiotic resistance protein

SDS-PAGE - Sodium dodecyl sulfate polyacrylamide gel electrophoresis

SW – Saltwater

T – Thiolation domain

TSS – Transcription start site

μE – MicroEinstein

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My journey through graduate school has taken several years, but these years are some of the best of my life and I am eternally grateful for the people I've met, the constant support from my family and friends, and my growth as a person and as a scientist during this time. There are many, many people to thank and I appreciate all you have done for me.

My wonderful family, which continues to grow: My Mom and Dad, Sandy and Jerry, for always believing in me and for being wonderful parents. My brother Tyler and sister Calyn, who are always there for me and share my sense of humor. All of my stepfamily in Tucson and relatives throughout the country for your love, thoughts and constant prayers. And of course, all of my new in-laws, who are among the warmest, most sincere people I've ever met and whom I'm definitely blessed to have in my life.

My wife Carla: You are the best thing that has ever happened to me and I am so happy to be able to spend the rest of my life with you. Meeting you has easily been the best part of my time at Scripps and in San Diego and I am thrilled to find out what the future holds for us. I couldn't have made it this far without your love, support, and all of your advice. I love you.

All of my friends, especially from Providence, who remain loyal and steadfast and never lose contact despite me moving farther and farther away.

Dr. William Gerwick: Thank you for the tremendous privilege of working in your laboratory, and for giving me the opportunity to join the lab in the unpredictable period of transition to Scripps. I hope I have been able to make some lasting

contributions while I've been here. I can't imagine having worked anywhere else for my doctorate. It has been a pleasure to come to work each and every day, and this is mostly because I always knew you were supportive of my research, regardless of how much success some of my ideas held. I feel that my future career prospects are bright and I have you to thank for this. Thank you also for the opportunity to travel to some amazing places and visit parts of the world I may never have otherwise seen.

Dr. Lena Gerwick: Thank you for all of your help and support during my time here. You've always been available to troubleshoot, lend advice and support, and to make sure things were going okay. I appreciate all of your help and enjoyed working with you on many different projects.

The Gerwick Lab: Thank you for all being such great colleagues, past and present. You all have been such a big part of my life during this time and I've enjoyed working with each and every one of you. Thanks in particular to those with whom I've collaborated, especially Ed Esquenazi and Emily Monroe, for making all of this so enjoyable.

My officemates Sarah Glaser and Cameron Coates for support, encouragement, and conversation during some very long hours in front of the computer screen.

The Moore Laboratory: A big thank you to Brad and the whole Moore group for being such great hosts for me over the last two years of my research. Your kindness, support, and advice were unwavering and allowed me to feel at home every time I came to visit. I learned an incredible amount in a very short period of time, for which I will always be grateful.

The Natural Products Community at SIO: Thank you to all the NP graduate students, postdocs, and visitors who I've met during my time here. It's hard to put into words what it means to be a part of such a vibrant and intelligent research community. Thank you in particular to Drs. William Fenical and Paul Jensen for giving me my first insights into Scripps natural product research many years ago during our trips to the Bahamas.

Above all, I am eternally grateful to my Lord and Savior Jesus Christ. Nothing I have done or will ever do would be possible without You in my life. I am happy to have found such a great home in San Diego at St. Mark's and will always be appreciative of the times I've had and ways I've been able to serve.

Chapter 2 of this dissertation includes a portion of a manuscript: Jones, Adam C.; Gerwick, Lena; Gonzalez, David; Dorrestein, Pieter C.; Gerwick, William H. "Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins." *BMC Microbiology* 2009, 9:247. The dissertation author is the primary investigator and author of this paper.

Chapter 3 of this dissertation contains portions of two manuscripts: 1) Jones, Adam C.; Gerwick, Lena; Gonzalez, David; Dorrestein, Pieter C.; Gerwick, William H. "Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins." *BMC Microbiology* 2009, 9:247; and 2) Esquenazi, Eduardo; Jones, Adam C.; Byrum, Tara; Dorrestein, Pieter C.; Gerwick, William H. Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proceedings of the National*

Academy of Sciences 2011, 108:5226-5231. As mentioned above, the dissertation author is the primary investigator and author of the first paper. For the second paper, the dissertation author was involved in several of the MALDI experiments described, as well as organic extractions, LC-MS profiling of abstracts, experimental design, and writing the manuscript.

Chapter 4 of this dissertation includes a portion that is being prepared for submission in 2011. Jones, Adam C; Ottilie, Sabine; Eustáquio, Alessandra S.; Edwards, Daniel J; Gerwick, Lena; Moore, Bradley S; Gerwick, William H. Evaluation of *Streptomyces coelicolor* as a heterologous expression host for natural products from the filamentous cyanobacterium *Lyngbya majuscula*. The dissertation author was the primary investigator and author of this paper.

Chapter 5, in full, is reprinted as it appears in Jones, Adam C.; Monroe, Emily A.; Podell, Sheila; Hess, Wolfgang R.; Klages, Sven; Esquenazi, Eduardo; Niessen, Sherry; Hoover, Heather; Rothmann, Michael; Kube, Michael; Reinhardt, Richard; Lasken, Roger S.; Yates III, John R., Burkart, Michael D.; Allen, Eric E.; Dorrestein, Pieter C.; Gerwick, William H.; Gerwick, Lena. Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proceedings of the National Academy of Sciences* 2011, doi:10.1073/pnas.1101137108. The dissertation author is one of two primary investigators and authors of this paper (along with EA Monroe). For this project, the author was involved in genome annotation, natural product gene cluster elucidation, organic extractions, proteomic analysis, experimental design, data analysis, and writing the manuscript.

VITA

EDUCATION AND FIELDS OF STUDY

University of California, San Diego 2006-2011
Doctor of Philosophy in Oceanography
Major advisor: William H Gerwick

University of North Carolina Wilmington, Wilmington North Carolina 2002-2004
Master of Science in Marine Biology
Major Advisor: Joseph R. Pawlik

Providence College, Providence Rhode Island 1998-2002
Bachelor of Science in Biology, *magna cum laude*
Minor: History

HONORS AND AWARDS

Millis-Colwell Exchange Program for Early Career Scientists Fellowship, American Society for Microbiology (US Delegate) 2011
California Sea Grant Graduate Student Trainee, 2008-2009; 2010-2011
Achievement Rewards for College Scientists (ARCS) Fellowship, Los Angeles Chapter, 2009-2010
Fleet Admiral Chester W Nimitz Fellowship, Scripps Institution of Oceanography, 2007-2009
American Society of Pharmacognosy Graduate Student Travel Grant, 2007
Associate Member, Sigma Xi Honor Society, 2003
Charles V Reichart Award in Biology, Providence College, 2002
Phi Alpha Theta (History) Honor Society, Providence College, 2001
National Society of Collegiate Scholars, Providence College, 2000
Alpha Epsilon Delta (Health Preprofessional) Honor Society, Providence College, 1999
Dean's Scholarship, Providence College, 1998-2002

PUBLICATIONS

Jones AC, Ottilie S, Eustáquio AS, Edwards DJ, Gerwick L, Moore BS, Gerwick WH. Evaluation of *Streptomyces coelicolor* as a heterologous expression host for natural products from the filamentous cyanobacterium *Lyngbya majuscula*. Forthcoming.

Zubillaga AL, Jones AC, Lutes D, Meadows M, Nugues MM, Szmant AM. Behavior of the embryos and larvae of the threatened Caribbean coral, *Acropora palmata* (Lamarck 1816): Buoyancy, swimming, competency and settlement preferences. Forthcoming.

Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S, Hoover H, Rothmann M, Kube M, Reinhardt R, Lasken RS, Yates III JR, Burkart MD, Allen EE, Dorrestein PC, Gerwick WH, Gerwick L (2011) Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proceedings of the National Academy of Sciences of the USA* doi:10.1073/pnas.1101137108.

Udwary DW, Gontang EA, Jones AC, Jones CS, Schultz AW, Winter JM, Yang JY, Beauchemin N, Capson TL, Clark BR, Esquenazi E, Eustáquio AS, Freel K, Gerwick L, Gerwick WH, Gonzalez DJ, Liu W-T, Malloy KL, Maloney KN, Nett M, Nunnery JK, Penn K, Prieto-Davo A, Simmons TL, Weitz S, Wilson MC, Tisa LS, Dorrestein PC, Moore BS (2011). Comparative genomic and proteomic analysis of the actinorhizal symbiont *Frankia* reveals significant natural product biosynthetic potential. *Applied and Environmental Microbiology* doi:10.1128/AEM.00038-11.

Jones AC, Monroe EA, Gerwick WH (2011) Elegant metabolite biosynthesis. *Chem Biol* 18:281-283.

Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226-5231.

Jones AC, Monroe EA, Eisman EB, Gerwick L, Sherman DH, Gerwick WH (2010) The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* 27:1048-1065.

Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.

Jones AC, Gu L, Sorrels CM, Sherman DH, Gerwick WH (2009) New tricks from ancient algae: Natural products biosynthesis in marine cyanobacteria. *Curr Op Chem Biol* 13:216-223.

Gerwick WH, Byrum T, Coates RC, Engene N, Esquenazi E, Gerwick L, Grindberg R, Johnson M, Jones A, Malloy K, Nunnery J, Pereira A, Soria I, Sorrels C, Suyama T, Tamaguchi M, Tidgewell K, Villa F, Vining O, Dorrestein P, Gu L, Sherman DH (2008) Integrating chemical and biochemical approaches to natural products drug discovery from marine cyanobacteria. *Proceedings of the International Conference on Drug Discovery from Natural Products and Traditional Medicines*, 33-43 (Punjab, India).

Gerwick WH, Coates RC, Engene N, Gerwick L, Grindberg RV, Jones AC, Sorrels CM (2008) Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3:277-284.

Cooper WJ, Jones AC, Whitehead RF, Zika RG (2007) Sunlight induced photochemistry of oxidants in natural waters: Implications in ballast water treatment. *Environ Sci Technol* 41:3728-3733.

Jones AC, Gensemer RW, Stubblefield WA, Van Genderen E, Dethloff GA, Cooper WJ (2006) Toxicity of ozonated seawater to marine organisms. *Environ Toxicol Chem* 25:2683-2691.

Harris LG, Jones AC (2005) Temperature, herbivory and epibiont aquisition as factors controlling the distribution and ecological role of an invasive seaweed. *Biological Invasions* 7:913-924.

Jones AC, Blum JE, Pawlik JR (2005) Testing for defensive synergy in Caribbean sponges: Bad taste or glass spicules? *J Exp Mar Biol Ecol* 322:67-81.

SELECTED RECENT PRESENTATIONS

Jones AC. Regulation and heterologous expression of bioactive natural products from the marine cyanobacterium *Lyngbya majuscula* (invited oral presentation) *California and the World Ocean 2010*, San Francisco, CA, September 2010

Jones AC, Ottilie S, Esquenazi E, Gonzalez D, Eustáquio AS, Saha S, Gerwick L, Dorrestein PC, Golden JW, Moore BS, Gerwick WH. Regulation and heterologous expression of bioactive natural products from the marine cyanobacterium *Lyngbya majuscula* (poster) *110th General Meeting of the American Society for Microbiology*, San Diego, CA May 2010

Jones AC, Gerwick L, Gonzalez D, Esquenazi E, Dorrestein PC, Gerwick WH. Regulatory elements associated with the jamaicamide gene cluster in *Lyngbya majuscula* (poster) *50th Meeting of the American Society of Pharmacognosy*, Honolulu, HI, June 2009

Jones AC, Gerwick L, Gerwick WH. Regulation of jamaicamide expression in the marine cyanobacterium *Lyngbya majuscula*. (poster) *L.S. Skaggs Biomedical Research Symposium,* Salt Lake City, UT, November 2008

Jones AC, Gerwick L, Gerwick WH. Regulation of jamaicamide expression in the marine cyanobacterium *Lyngbya majuscula*. (poster) *Society for Industrial Microbiology Annual Meeting and Exhibition,* San Diego, CA, August 2008

Jones AC, Gerwick L, Gerwick WH. Regulation of jamaicamide expression in *Lyngbya majuscula.* (poster) *17th Gordon Research Conference in Marine Natural Products*, Ventura, CA, February 2008

Jones AC, Gerwick L, Gerwick WH. Transcriptional regulation of cyanobacterial toxins. (poster) *48th Annual American Society of Pharmacognosy Meeting*, Portland, ME, July 2007

Jones AC, Pawlik JR. Sponge Synergy: Evidence for interactions between chemical and physical defenses in Caribbean sponges. (oral presentation) *33rd Annual Benthic Ecology Meeting*, Mobile, AL, March 2004

OUTREACH

Center for Ocean Sciences Education Excellence (COSEE) – California: Teaching Ocean Sciences in the 21st Century Classroom (2008 - 2010). Present drug discovery research to middle school classrooms (approximately 300 students) via internet broadcast from the Gerwick laboratory at Scripps Institution of Oceanography, UCSD.

PUBLIC SERVICE

St. Mark's United Methodist Church Lay Leader (San Diego, CA), 2010-present
St. Mark's United Methodist Church Council Member, 2008-present
St. Mark's United Methodist Church Youth Group Leader, 2007-2008
Providence College Board of Programmers, 2000-2002
Providence College Student Congress Committee Chair, 2000-2001

MEMBERSHIPS

American Academy of Underwater Sciences (AAUS) Scientific Diver, ~300 dives
American Society for Microbiology
American Chemistry Society

ABSTRACT OF THE DISSERTATION

Regulatory, genetic, and genomic investigations of
natural products biosynthesis in marine cyanobacteria

by

Adam Christopher Jones

Doctor of Philosophy in Oceanography

University of California, San Diego, 2011

William H. Gerwick, Chair

Natural products from the marine environment constitute one of the richest and most reliable sources of bioactive molecules for drug discovery and biotechnology. Filamentous marine cyanobacteria are among the most prolific natural product producers, and with significant advances in genetics and genomics over the past decade are gaining increasing recognition for unusual or unprecedented biochemistry involved in natural product biosynthesis. However, the further development of several

of these cyanobacterial secondary metabolites has not been accomplished because of low yields from field collections, slow growth of these organisms in laboratory cultures, and an absence of any genetic techniques for DNA manipulation.

The aim of the research presented in this dissertation is to study the biosynthesis of natural products from the marine filamentous cyanobacterium *Lyngbya majuscula* to investigate how access to these compounds can be improved and the full potential of cyanobacterial strains can be realized. A variety of experiments are described herein, including: 1) A transcriptional analysis of the neurotoxin jamaicamide gene cluster from *Lyngbya majuscula*, which employed RT-PCR and reporter gene assays to better understand how cyanobacterial biosynthetic pathways are transcribed; 2) Investigations into jamaicamide regulation, which included a) protein pulldown assays to isolate possible light related regulatory proteins for this pathway and b) the development of a novel mass spectrometry approach to monitor natural product biosynthetic turnover, the effect of environmental parameters on this turnover, and confirm the impact of light on jamaicamide production; 3) Evaluating different methods for transferring DNA into *Lyngbya* filaments and taking the first steps in heterologously expressing portions of a *Lyngbya* gene cluster in the actinomycete *Streptomyces coelicolor*, and 4) Sequencing the genome of *Lyngbya majuscula* 3L, thereby revealing its natural product potential, a complex gene regulatory network and a surprising inability to fix atmospheric nitrogen. Collectively, these experiments provide new perspective on natural product biosynthesis in *Lyngbya* strains and direction on how these organisms and their natural products can be harnessed for biomedical and biotechnological applications.

CHAPTER 1

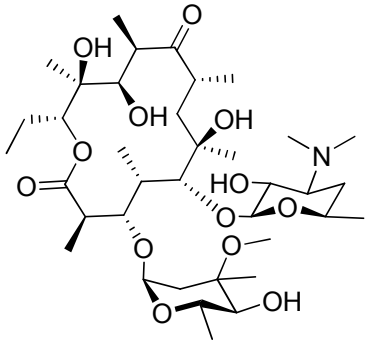
Introduction

1.1 Importance of natural products in drug discovery

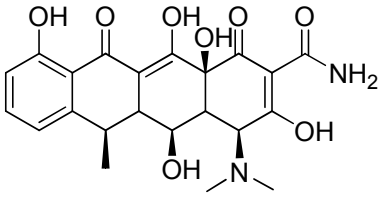
For thousands of years, humans have relied on the medicinal qualities of compounds produced in Nature to treat a variety of ailments. Remains of marine algae recovered from Monte Verde in Southern Chile from over 13,000 years ago indicated early settlers in the area likely relied on seaweeds for food and medicine (Dillehay et al. 2008). Likewise, recent analysis of pottery jars from ancient Egypt (ca. 3150 BC) have found that herbs and tree resins were dispensed into grape wine for their supposed health benefits (McGovern et al. 2009). In modern medicine, naturally produced compounds, or "natural products," have had an enormous impact on drug discovery and development, and have become standard treatment for many human diseases. The term "natural product" is often interchanged with the designation "secondary metabolite", as many of these compounds are thought to serve a specific ecological role in the producing organism and are not vital for primary metabolic functions (Fraenkel 1959). Currently, the majority of drugs in use today are either natural products themselves, derivatives of natural products, or are synthetic molecules that have some feature inspired by a natural product (Newmann and Cragg 2007).

The scientific field of natural products chemistry has long been associated with the pursuit of new compounds from terrestrial sources, and several of the most significant drugs in use today are hallmarks of these efforts over the last century

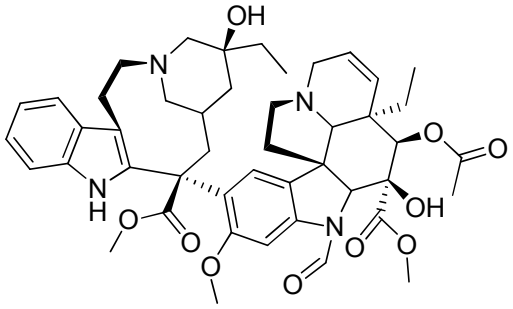
(Figure 1.1). Penicillin, a β-lactam antibiotic discovered by Alexander Fleming in 1928 from *Penicillium* fungi, is one of the most frequently used antibiotics and brought fundamental change to the treatment of bacterial infections (Drews 2000). Other antibiotics produced by actinobacteria, including erythromycin (Oliynyk et al. 2008), novobiocin (Heide 2009), vancomycin (Levine 2006), and tetracycline (Chopra and Roberts 2001) have become benchmarks for modes of action against specific diseases. Doxycycline, one antibiotic of the tetracycline class of compounds, is also used in prophylaxis against malaria (Anderson et al. 1998), a disease first treated with the natural product quinine from the bark of the *Chincona* tree (White 1996). Several anticancer agents from terrestrial organisms in use today include the antimitotic paclitaxel (taxol) originally isolated from the Pacific Yew Tree *Taxus brevifola* (Wani et al. 1971), and the vinca alkaloids from the Madagascar Periwinkle *Catharanthus* (Johnson et al. 1963).



Erythromycin

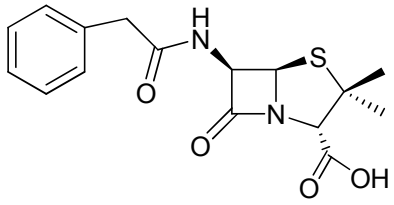


Doxycycline

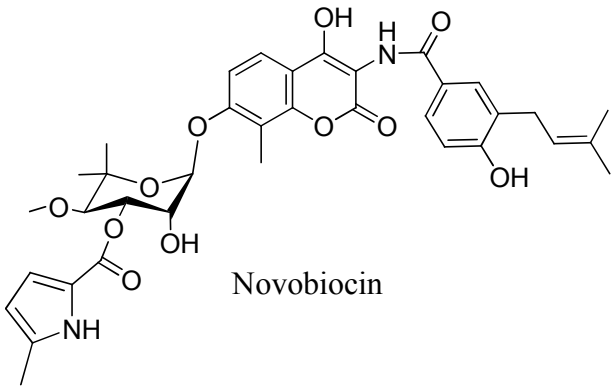


Vincristine (Vinca alkaloid)

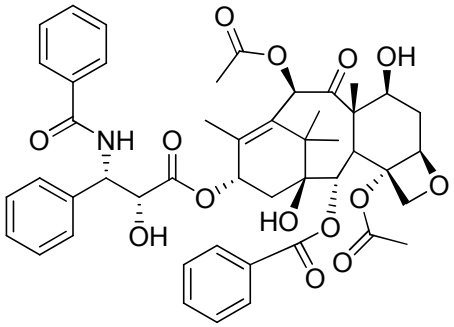
Figure 1.1: Natural products approved for human use isolated from terrestrial sources.



Penicillin G

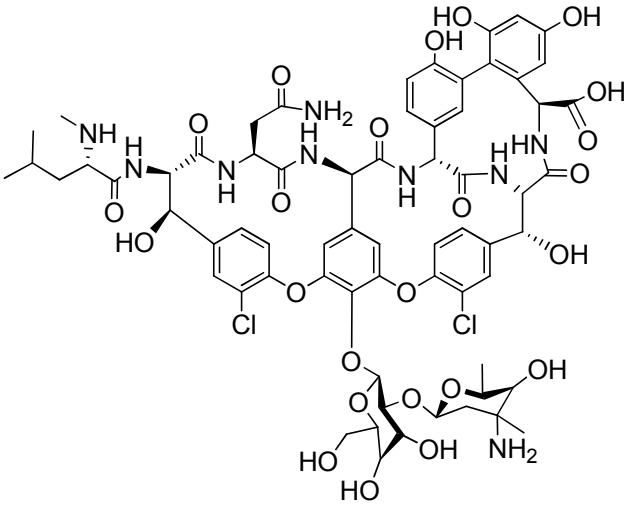


Novobiocin

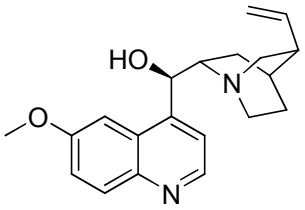


Paclitaxel

Figure 1.1 (continued): Natural products approved for human use isolated from terrestrial sources.



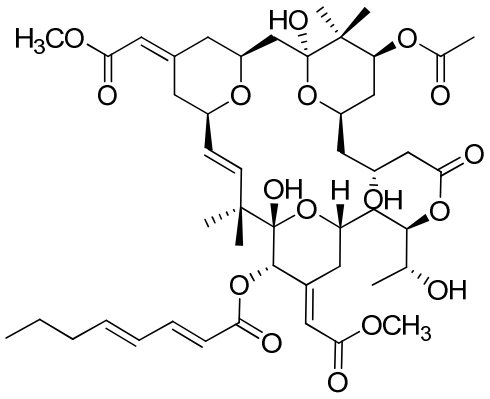
Vancomycin



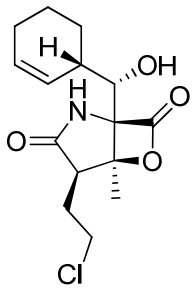
Quinine

Figure 1.1 (continued): Natural products approved for human use isolated from terrestrial sources.

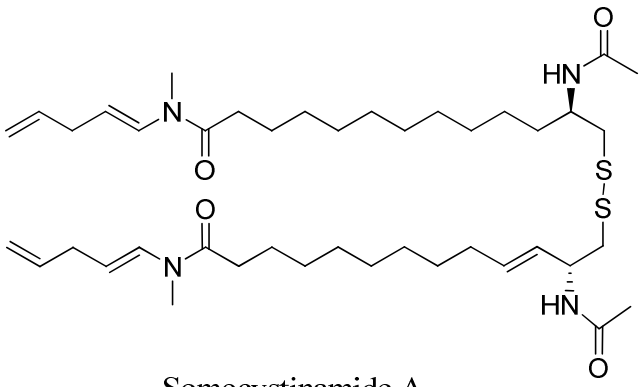
Over the last 50 years, the search for natural products has been extended in earnest to include organisms from the marine realm. Initial investigations into the chemistry of marine algae, cyanobacteria, and sponges from research groups at the University of Hawaii and Scripps Institution of Oceanography at the University of California San Diego have led to the formation of a large number of laboratories dedicated to studying the chemistry and biochemistry of marine natural products. To date, approximately 17,000 natural products have been isolated from marine organisms (Blunt et al. 2009), and the bioactivity and clinical potential of many of these compounds rival the most promising terrestrial discoveries (Figure 1.2). For example, bryostatin, a protein kinase C modulator from a microbial symbiont of the bryozoan *Bugula neritina* (Sudek et al. 2007), has advanced to clinical trials to be evaluated as an anticancer agent and a treatment for Alzheimer's disease. Salinosporamide A, a γ-lactam-β-lactone compound from the marine actinomycete *Salinispora tropica*, is an especially potent inhibitor of the 20S proteasome and is currently in clinical trials for treating multiple myeloma (Feling et al. 2003; Molinski et al. 2009). Somocystinamide is a lipopeptide dimer from a *Lyngbya*/*Schizothrix* Fijian cyanobacterial assemblage (Nogle and Gerwick 2002) that was recently found to trigger tumor cell apoptosis via Caspase 8 (Wrasidlo et al. 2008) and is being considered for Phase 0 trials in treating solid tumors, with possible delivery using nanoparticles.



Bryostatin 1



Salinosporamide A



Somocystinamide A

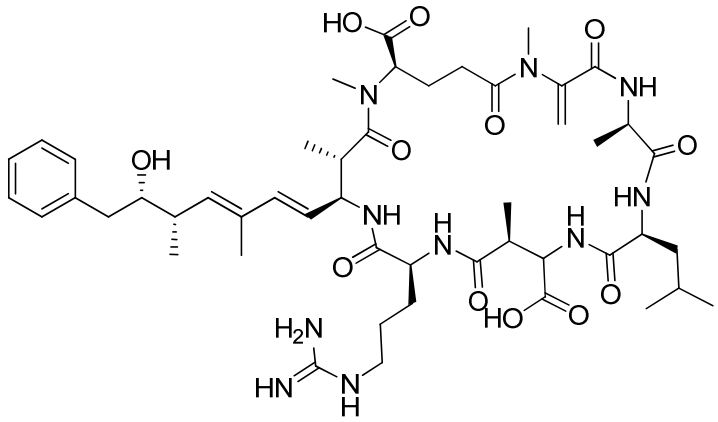
Figure 1.2: Bioactive natural products from marine organisms.

1.2 Cyanobacteria

Unarguably, members of the phylum cyanobacteria are among the richest and most important sources of natural products yet discovered. Cyanobacteria are some of the oldest life forms on earth, originating at least 2 billion years ago (Rasmussen et al. 2008), and can be found almost everywhere on the planet, including freshwater lakes and ponds, oceans, deserts, polar regions, and hot springs (Seckbach 2007). Cyanobacteria are photoautotrophic, oxygen producing bacteria that are thought to

have triggered the conversion of Earth's early atmosphere to an oxidizing state, and are likely the ancestors of chloroplasts found in algae and higher plants (Falcón et al. 2010). Cyanobacteria are so named because of the presence of the pigment phycocyanin, which results in many cyanobacteria having a blue-green coloration (leading to the common name "blue-green" algae). Collectively, cyanobacteria perform a large percentage of global carbon and nitrogen fixation, with much of this contribution coming from unicellular oceanic species such as *Prochlorococcus*, *Synechococcus*, and *Trichodesmium* (Zehr et al. 2001, Ting et al. 2002).

Increasingly, bioactive compounds from filamentous cyanobacteria, especially from the orders Oscillatoriales and Nostocales, are receiving both acclaim for their pharmacological potential despite their notoriety for environmental toxicity (Burja et al. 2001). Approximately 50 species of cyanobacteria are responsible for the production of toxins found in both freshwater and marine environments (Antoniou et al. 2005). Toxic freshwater blooms can occur in lakes, rivers, and ponds worldwide and are blamed for sporadic but recurring incidents of animal poisonings and death. Human poisonings have also occurred from exposure to municipal and recreational water supplies (Carmichael 2001). The hepatotoxic heptapeptide microcystin is of particular concern due to its wide geographic distribution and production by multiple organisms. The original source of microcystin was *Microcystis aeruginosa*, but over 65 structural isoforms have been isolated from cyanobacterial species of the genera *Anabaena*, *Aphanocapsa*, *Hapalosiphon*, *Nostoc*, and *Plankothrix*, in addition to *Microcystis* (e.g., Rinehart et al. 1994, Carmichael 2001).



Microcystin L-R

Figure 1.3: Structure of the cyanobacterial toxin microcystin L-R.

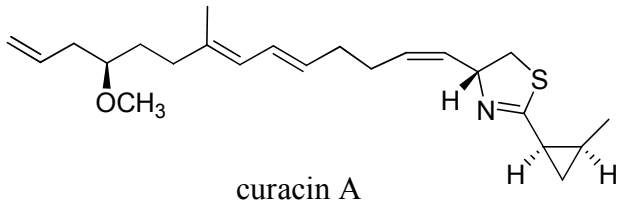
In the marine environment, toxins produced by cyanobacteria such as *Lyngbya majuscula* (lyngbyatoxin) cause blistering dermatitis known as "swimmer's itch" (e.g., Cardellina II et al. 1979). Freshwater *Lyngbya* species can produce neurotoxins such as saxitoxin (Carmichael et al. 1997), a compound also produced by marine dinoflagellates that causes paralytic shellfish poisoning. Lyngbyatoxin and another *L. majuscula* compound, debromoaplysiatoxin (a tumor promoter) have been found in high levels in Australian sea slugs (Capper et al. 2005), which may indicate accumulation in higher trophic levels of the food web.

Investigations of marine natural products from cyanobacteria were first initiated by Professor Richard Moore from the University of Hawaii in the mid 1970s (Cardellina II and Moore 2010). Early field collections in the Hawaiian and Marshall Islands led to the discovery of the malyngamides (Cardellina II et al. 1979), debromoaplysiatoxin (Mynderse et al. 1977) and lyngbyatoxin as described above.

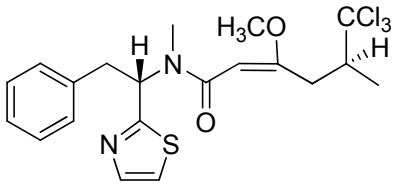
Since that time, over 800 bioactive natural products have been isolated from marine cyanobacteria around the world. The majority of these compounds (over 250) have been ascribed to the genus *Lyngbya*, with 76% from this genus assigned to the species *Lyngbya majuscula* (Tidgewell et al. 2010). The *Lyngbya* genus is characterized morphologically on the basis of large, macroscopic filaments that grow in shallow water environments. These filaments consist of a thick polysaccharide sheath of approximately 50-60 μm in width that encases stacks of *Lyngbya* cells (Engene et al. 2010). In contrast to other filamentous cyanobacteria from freshwater environments such as *Anabaena (Nostoc)* sp. PCC 7120 (Kaneko et al. 2001), *Lyngbya* strains do not form heterocysts, specialized cells designated for nitrogen fixation. Marine *Lyngbya* strains are found pantropically in shallow coral reef environments, where they commonly grow in close association with other cyanobacteria and algae and provide an ideal substrate for a variety of heterotrophic bacteria. When grown in laboratory culture, most *Lyngbya* strains retain these heterotrophic bacteria on their filament sheaths and cannot be grown axenically (Gerwick et al. 2008).

A number of compounds originally isolated from *Lyngbya* field collections have been the subject of further investigations concerning their bioactivity, biosynthesis, and use in downstream applications, including consideration for preclinical and clinical trials as potential pharmaceuticals. Several natural products from Caribbean and Pacific collections have received substantial attention for their cytotoxicity (Figure 1.4). Curacin A is a compound produced by the strain *Lyngbya majuscula* 3L from Curaçao, Netherlands Antilles. Curacin A exhibits antimitotic and antiproliferative activity, serving as an inhibitor of colchicine binding to tubulin, and

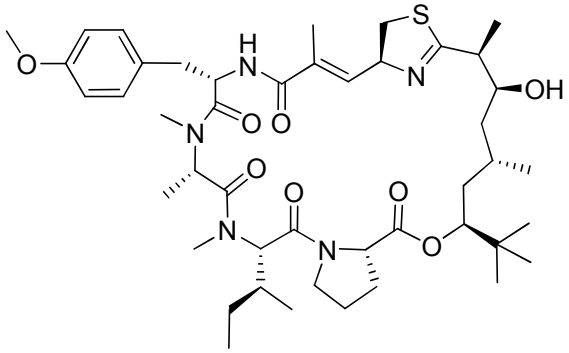
shows toxicity in brine shrimp assays (Gerwick et al. 1994, Verdier-Pinard et al. 1998). Initial work with this compound found that it composes approximately 10% of *L. majuscula* 3L extracts, and further efforts were expended to adapt this strain to laboratory culture to provide consistently high curacin A yields (Rossi et al. 1997). Curacin A is also available by total synthesis via several different routes (e.g., White et al. 1997; Muir et al. 1998). Two other compounds produced by *L. majuscula* 3L are barbamide (Orjala and Gerwick 1996), a potent molluscicide featuring an unusual trichlorinated methyl group, and the lipopeptide carmabin (Hooper et al. 1998). Chapter 5 of this dissertation will discuss these compounds in more detail as the subject of a project to sequence the genome of *L. majuscula* 3L. A particularly intriguing series of cytotoxic molecules from *Lyngbya* strains are the apratoxins (Luesch et al. 2001), which are produced by the Pacific strain *Lyngbya bouillonii*. *Lyngbya bouillonii* is commonly found in somewhat deeper reef environments (10-20 m depth), where it grows in thickly entangled mats or tubes. These assemblages explain the moniker "cobweb *Lyngbya* (Tan et al. 2002)" and are likely created by a species of snapping shrimp that lives within the *L. bouillonii* filaments. Apratoxin is a nanomolar level cancer cell toxin that appears to act via G1 cell cycle arrest and resulting apoptosis (Luesch et al. 2005) and has also been characterized as a kinase inhibitor and a preventative agent of secretory protein translocation (Liu et al. 2009).



curacin A



barbamide



apratoxin A

Figure 1.4: Cytotoxic compounds from marine *Lyngbya* strains.

Neurotoxic natural products from *Lyngbya* strains are well recognized for their potency and strong potential for development as therapeutics (Figure 1.5). Antillatoxin is a neurotoxic lipopeptide from *L. majuscula* collections in Curaçao (Orjala et al. 1995) and Puerto Rico (Nogle et al. 2001) that has been found to be a potent ichthyotoxin and molluscicidial agent (Orjala et al. 1995) as well as neurotoxic (nM scale) against rat cerebellar granule cells (Li et al. 2001). The voltage gated

sodium channel activity of antillatoxin was recently found to enhance neurite outgrowth in immature cerebrocortical neurons (Jabba et al. 2010), which may offer a new way of regulating neuronal plasticity. Another neurotoxic lipopeptide from Caribbean *L. majuscula* strains is kalkitoxin, which also has ichthyotoxic activity but acts as a sodium channel blocking agent (Wu et al. 2000). A Jamaican strain of *L. majuscula* collected in 2004 was found to produce the jamaicamides, which are neurotoxins acting as sodium channel blockers (Edwards et al. 2004). The jamaicamides feature unusual functional groups including a vinyl chloride moiety and an alkynyl bromide in jamaicamide A, which has only been observed in one other marine natural product (Mevers et al. 2011). The jamaicamides form the basis of several experiments described in this dissertation in Chapters 2 and 3.

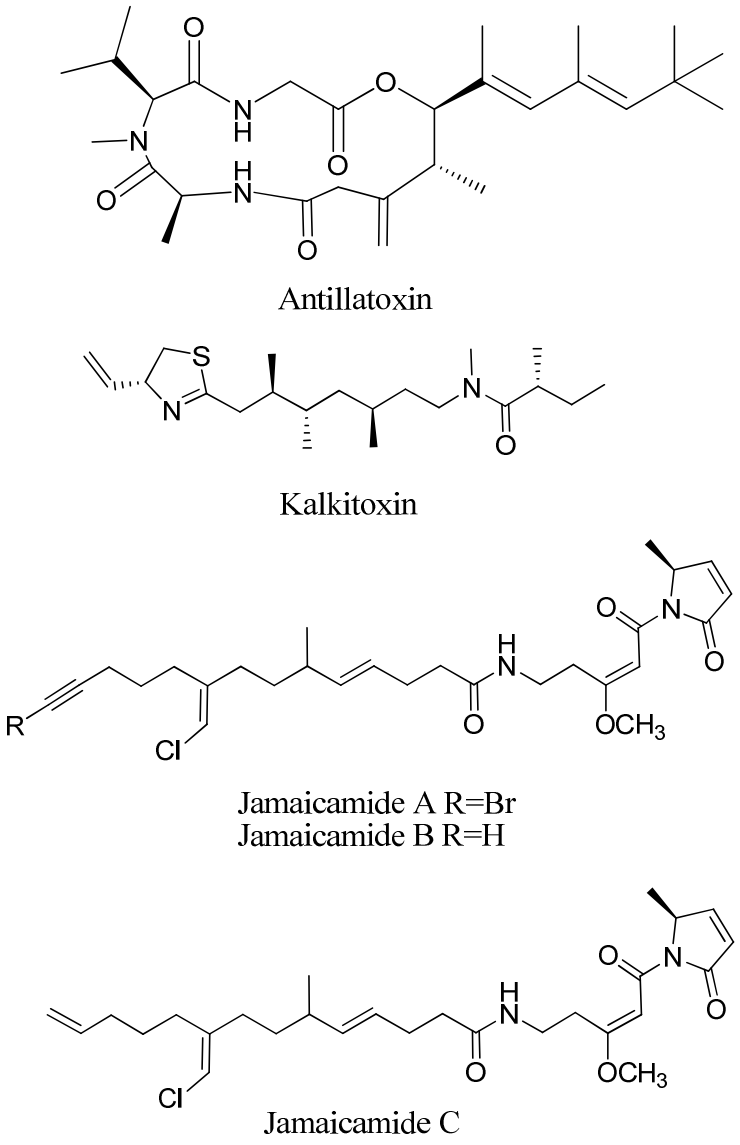


Figure 1.5: Neurotoxic natural products from marine *Lyngbya* strains.

1.3 Biosynthesis of natural products in marine cyanobacteria

Important advances in discovering more about the biological origins of cyanobacterial natural products have been achieved in the past decade, coinciding with new developments in the areas of genetics and genomics. Structural elucidation of

bioactive compounds from cyanobacteria has been complemented with the isolation and characterization of the biosynthetic gene clusters that code for their assemblies. Gene clusters for the hepatotoxins microcystin (Tillett et al. 2000, Kaebernick et al. 2002) and nodularin (Moffitt & Neilan 2004) have been identified, as well as several from *Lyngbya* strains, including those for the molluscicide barbamide (Chang et al. 2002), cancer cell toxins curacin A (Chang et al. 2004) and apratoxin A (Grindberg et al. 2011), antifungal agent hectochlorin (Ramaswamy et al. 2007) dermatotoxin lyngbyatoxin (Edwards & Gerwick 2004) and the neurotoxic jamaicamides (Edwards et al. 2004). Cyanobacterial natural product gene clusters are often composed of and have become typified by mixed polyketide synthases (PKS) and non-ribosomal peptide synthetases (NRPS). PKSs and NRPSs are multifunctional protein complexes that are responsible for catalyzing the stepwise condensation of metabolic building blocks. Both are organized in modules, where each module is capable of recognizing and adding either coenzyme A thioester derivatives of carboxylic acids (PKS) or amino acids (NRPS) into the growing chain. Modules are comprised of domains that each correspond to a particular biochemical reaction (Figure 1.6; Du et al. 2001, Chang et al. 2002, Fischbach & Walsh 2006).

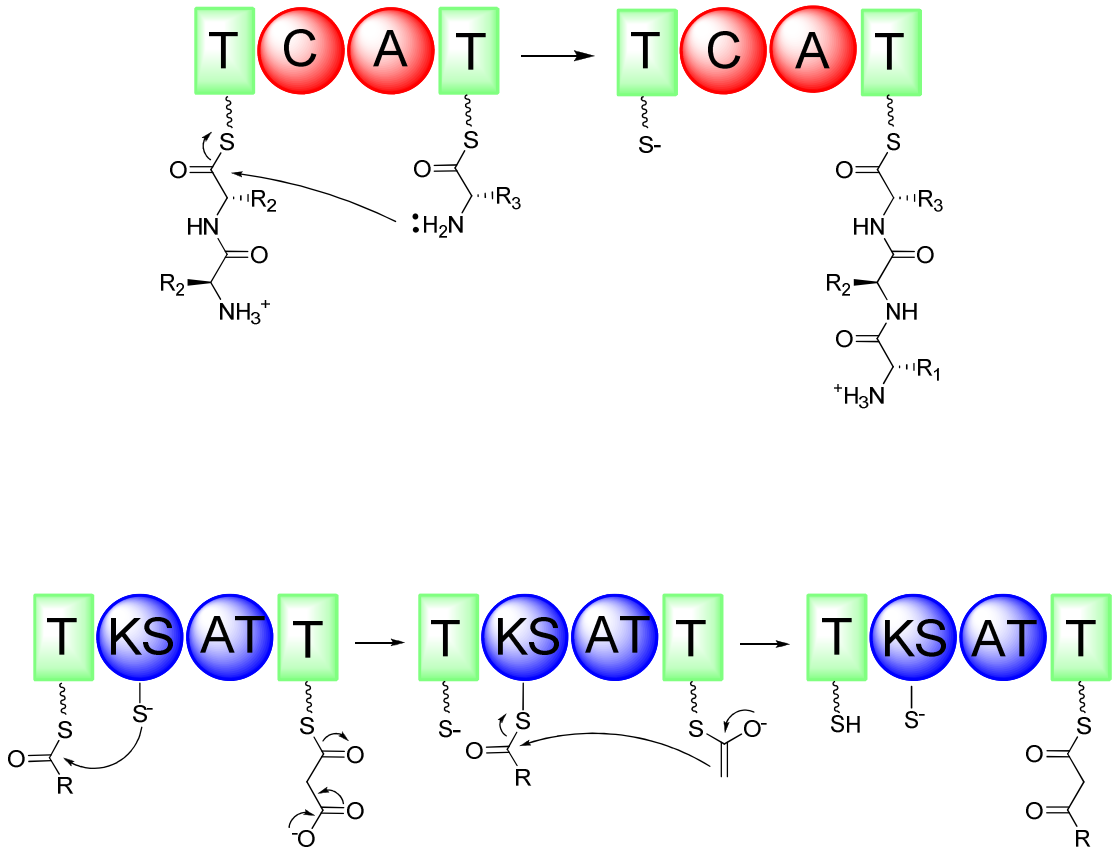


Figure 1.6: Condensation reactions occurring during NRPS (top) and PKS (bottom) biosynthesis. Domain abbreviations: T = thiolation, C = condensation, A = Adenylation, KS = ketosynthase, AT = acyltransferase. Figure adapted from Fischbach and Walsh (2006).

Identification and characterization of the barbamide biosynthetic gene cluster represented the first such analysis of a natural product pathway from a marine microorganism (Chang et al. 2002). Precursor feeding experiments indicated that barbamide is derived from acetate, phenylalanine, leucine, and cysteine, and identification of the cluster genes from a cosmid library revealed a 26 kb pathway with PKS and NRPS elements consistent with the barbamide structure. ATP-

pyrophosphate experiments testing the adenylation domains of the barbamide NRPS modules confirmed their specificity for the amino acids tested in the feeding experiments (Chang et al. 2002). A one carbon truncation in barbamide biosynthesis appears to be necessary to explain the final molecule, and the ATP-PPi experiments support the idea that this truncation likely occurs in the leucine portion as substrate specificity for BarE was identified for α-ketoisocaproic and α-trichloroketoisocaproic acids, but not isovaleric or trichloroisovaleric acids (Chang et al. 2002). The trichloromethyl group on the leucine residue of barbamide prompted additional experiments to determine how this particular functional group is biosynthesized. Leucine was first demonstrated as the substrate for halogenation through stable isotope feeding experiments (Sitachitta et al. 2000), which showed that chlorination occurs after leucine is biosynthesized. Recombinant expression of the non-heme Fe (II) α-ketoglutarate dependent halogenases BarB1 and BarB2 and incubation of these proteins with ^{14}C leucine, the BarD adenylation domain, ATP, and the holo version of the PCP BarA allowed for the predicted activity of BarB1 and BarB2 to be confirmed (Galonić et al. 2006). BarB2 adds the first two chlorine atoms, while the third is added by BarB1 (Flatt et al. 2006, Galonić et al. 2006).

Characterization of curacin A biosynthesis (Chang et al. 2004) represented another important achievement in understanding biosynthetic mechanisms used in *Lyngbya majuscula*, and the initial identification of this gene cluster has led to additional studies resulting in the curacin A pathway being the best understood among the mixed PKS/NRPS gene clusters from marine cyanobacteria identified to date. Following a similar strategy to that used in the barbamide work discussed above,

stable isotope precursor feeding experiments revealed that curacin A is composed of a cysteine residue, 10 acetate units, and two methyl groups derived from S-adenosyl methionine (SAM; Chang et al. 2004). Screening of a genomic library using PKS and NRPS probes led to the sequencing of the approximately 64 kb gene cluster. Subsequent to this work, detailed investigations into the enzymatic mechanisms used in *L. majuscula* 3L to biosynthesize curacin A have been completed, which collectively showcase the wide variety of unusual and unprecedented strategies marine *Lyngbya* strains possess in creating natural products. Chain initiation in curacin A is conducted using the CurA enzyme, which is composed of an adaptor domain, a GCN5-related *N*-acetyltransferase (GNAT) domain, and an acyl carrier protein (ACP). This GNAT domain functions as both as a decarboxylase and S-acetyltransferase, and was crystallized in both free form and in complex with acyl-CoA to confirm this activity (Gu et al. 2007). Recent work has shown that chain termination to yield the terminal olefin in curacin A is preceded by sulfonation that is catalyzed by a sulfotransferase located in the CurM enzyme. This sulfate moiety acts as an excellent leaving group during the final decarboxylation step (Gu et al. 2009).

Perhaps the most intriguing component of curacin A biosynthesis is the β-branching conducted by an HMG-CoA cassette located in the middle of the gene cluster – a cassette that is shared in high levels of homology with the jamaicamide biosynthetic pathway (Edwards et al. 2004, Gu et al. 2009). Each cassette contains three tandem ACPs from a single open reading frame (ORF; ACPs $_{I\text{-}III}$), a standalone ACP_{IV}, an HMG CoA synthase, and a dehydratase (ECH_1) and decarboxylase (ECH_2). These cassettes are flanked by α-ketoglutarate dependent non-haem halogenases and

enoyl reductase (ER) domains. Biochemical characterization of several components of these pathways (Gu et al. 2006, Gu et al. 2009) have found that in both pathways modification of a polyketide intermediate proceeds in parallel with reaction with the HCS, chlorination, and reduction via ECH1, but that differing activity between the two ECH2 domains yields a α,ß-enoyl thioester in curacin A and a ß, γ-enoyl thioester in jamaicamide. The former substrate reacts with the ER in the curacin pathway resulting in cyclopropyl ring formation, while the latter substrate in jamaicamide is unreactive with the curacin ER in *in vitro* experiments and instead corresponds with the vinyl chloride moiety seen in the jamaicamide molecule (Figure 1.7; Gu et al. 2009), suggesting that the jamaicamide ER is likely superfluous and an evolutionary remnant. This highly similar chemistry is a significant example of versatility in natural products biosynthesis due to horizontal gene transfer.

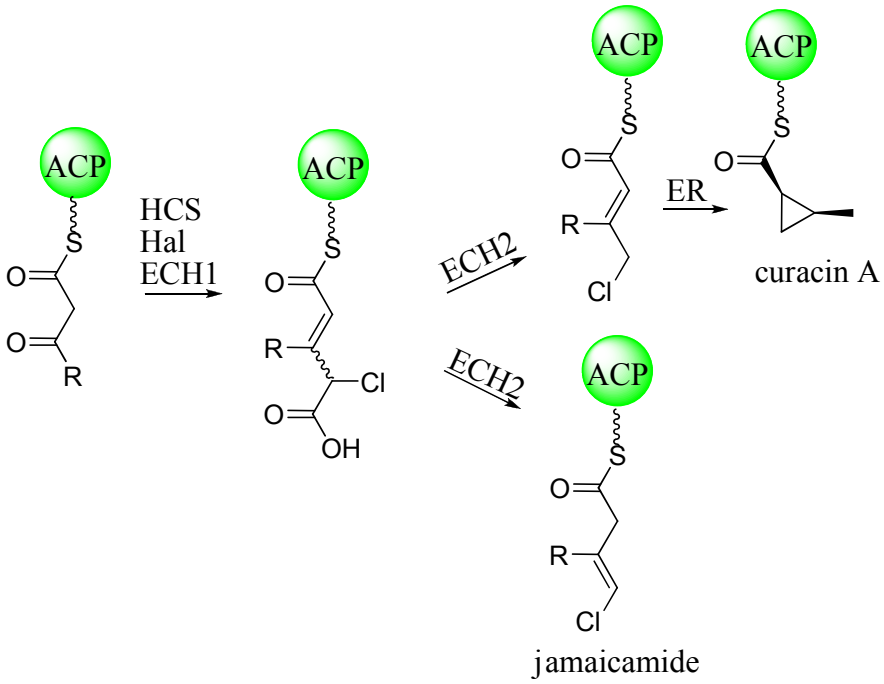


Figure 1.7: β-branching of curacin A and jamaicamide proceeds using a similar route until deviating at the ECH2 domain expressed from each cluster. Curacin A proceeds to form the terminal cyclopropyl ring while jamaicamide biosynthesis yields a vinyl chloride. Figure scheme adapted from Lane and Moore (2010).

The variety of mechanisms in other modular marine cyanobacterial natural product pathways (Jones et al. 2010) are further illustrated by dichlorination in hectochlorin (Ramaswamy et al. 2007), incorporation of a tert-butyl starter unit in apratoxin A (Grindberg et al. 2011), and reductive offloading from an NRPS enzyme in lyngbyatoxin A (Edwards and Gerwick 2004, Read and Walsh 2007). Marine cyanobacteria are also capable of biosynthesizing cyclic peptide natural products that are ribosomally encoded. The symbiotic cyanobacterium *Prochloron* lives in association with didemnin ascidians such as *Lissoclium patella* in the tropical Pacific and produces molecule classes such as the patellamides, ulithiacyclamides, and

lissoclinamides (Jones et al. 2009). In 2005, the patellamide gene cluster was isolated (Schmidt et al. 2005), which contains a precursor peptide (*patE*) that encodes the exact sequence of amino acids in patellamides A and C. The remaining genes function as tailoring and cyclizing enzymes to yield the final molecules. The patellamide pathway has been heterologously expressed in *E. coli* (Schmidt et al. 2005), and genetic engineering has allowed for new analogs of these cyclic peptides to be produced (Donia et al. 2006, 2008). Ribosomally encoded peptides similar to the patellamides, now referred to as "cyanobactins", appear to be widespread as over 100 compounds derived or anticipated from this proccess have been identified from free living and obligate symbiont cyanobacteria (Donia et al. 2008, Jones et al. 2009, Li et al. 2010).

1.4 Current limitations in accessing natural products from marine cyanobacteria for biotechnology and biomedicine

Despite the enormous potential of natural products from marine filamentous cyanobacteria in drug discovery and as biotechnological tools, there are a number of obstacles inherent in cyanobacterial physiology that limit access to large amounts of these compounds for further development to impractically large scale field collections or elaborate and difficult synthetic methods. Probably the biggest challenge is the slow growth rate of these organisms in laboratory culture (Rossi et al. 1997; Esquenazi et al. 2011). Typical natural product isolation efforts from *Lyngbya* and other strains begin with field collections of biomass, which is returned to the laboratory after being preserved in alcohol and frozen for future extraction. If live filaments are obtained and grown in the laboratory, they require several months and gradual transfer between

culture vessels to grow to the point where traditional natural product isolation projects can begin (Rossi et al. 1997). *Lyngbya* filaments double in size only about every 6-10 days (Esquenazi et al. 2011), in contrast to other bacterial strains such as *E. coli* that are capable of doubling every 20 minutes (Cooper and Helmstetter 1968). Many cyanobacterial strains brought back from field collections do not adapt well to laboratory conditions and cannot be reliably archived using glycerol stocks or other common storage methods, thus limiting their preservation only for chemical, genetic, or microscopic analyses.

Another substantial obstacle to studying natural product chemistry and biochemistry in marine cyanobacteria is the lack of genetic techniques for most filamentous strains. Unlike other model freshwater cyanobacteria such as *Synechococcus* or *Anabaena* (*Nostoc*) sp. PCC7120 (Koksharova and Wolk 2002), stable transfer of DNA for gene knockouts, replacement, or expression is not yet possible for *Lyngbya* or other tropical, benthic cyanobacteria often targeted for their secondary metabolites. In addition to their slow growth, *Lyngbya* strains are almost always found growing in close association with other cyanobacteria (Nogle and Gerwick 2002), algae, or heterotrophic bacteria that live on the surface of the *Lyngbya* polysaccharide sheath (Gerwick et al. 2008, Simmons et al. 2008). These associated organisms, especially heterotrophic bacteria, persist in laboratory culture and often prohibit maintenance of axenic *Lyngbya* strains (Gerwick et al. 2008). It is therefore probable that resistance to stable introduction of DNA into cyanobacterial strains may be increased because of the presence of additional restriction endonucleases (see Chapter 5). The biosynthetic origin of natural products attributed to non-axenic

cyanobacterial strains has also been questioned, as heterotrophic bacteria could possess the necessary genes to encode their own secondary metabolites. Recent evidence of malyngic acid isolated from the marine bacterium *Rhodopirellula baltica* has shown this could occasionally be a valid conclusion (Lee et al. 2010), as this molecule was earlier found from *Lyngbya majuscula* collections (Cardellina II and Moore 1980). However, successful imaging techniques of *Lyngbya* filaments using Matrix Assisted Laser Desorption Ionization (MALDI) mass spectrometry (Simmons et al. 2008, Esquenazi et al. 2008, Esquenazi et al. 2011) have localized the presence of many of these compounds to cyanobacterial filaments and single cells isolated from filaments using micromanipulation (Esquenazi 2011), confirming that in most cases these compounds are likely produced by cyanobacteria. Despite significant progress in heterologous expression of natural product pathways in a variety of bacteria, including actinobacterial antibiotics (Eustáquio et al. 2005), myxobacterial cytotoxins (Tang et al. 2000) and *Streptomyces* antitumor compounds (Challis 2006), recombinant expression of non-ribosomal cyanobacterial natural products has been limited to individual proteins or domains (Edwards and Gerwick 2004, Gu et al. 2009). No stable heterologous platform for compounds from *Lyngbya* has been established.

Along with physiological characteristics of marine filamentous cyanobacteria that impede natural product investigations, our knowledge of how expression of the biosynthetic clusters of genes in marine cyanobacterial genomes is controlled is extremely limited. If the regulatory triggers responsible for turning the transcription of these pathways “on” and “off” can be identified and manipulated, yields of natural

products can be significantly increased in laboratory culture and reduce the need for repeated field collections or organic synthesis. In general, gene transcription in bacteria is initiated by the binding of the enzyme RNA polymerase to gene regions known as ‘promoters’. The promoter binding sites are typically located about 35 and 10 base pairs (bp) upstream (5’ end) of a transcription start site (TSS). These regions are detected by bacterial polymerase protein subunits called ‘sigma factors’ which cause RNA polymerase to tightly bind to DNA and begin transcription (Figure 1.8; Alberts et al. 1998).

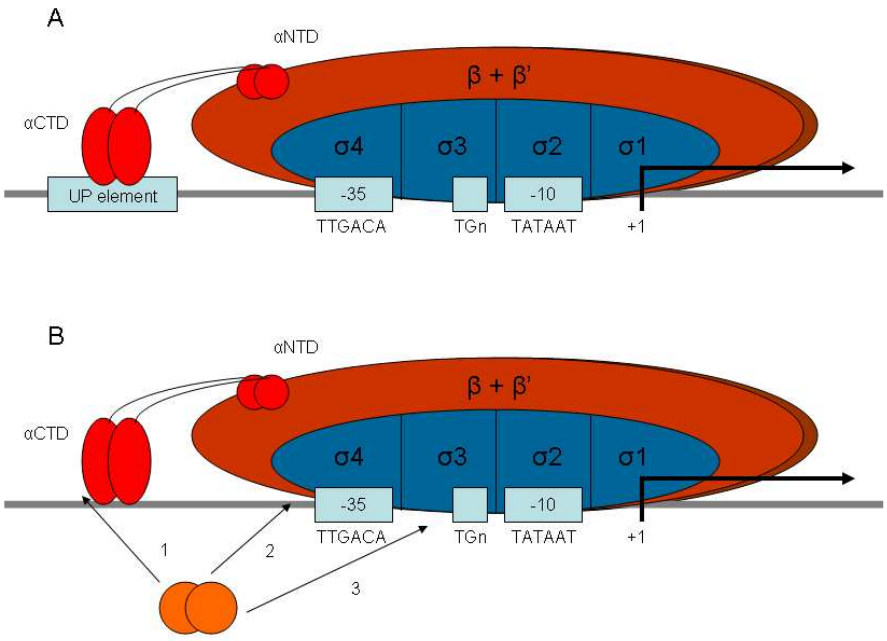


Figure 1.8: Example of bacterial RNA polymerase binding to a gene promoter region (top) and being activated by interacting with a dimerized transcription factor (bottom; figure adapted from Browning and Busby 2004). A) The sigma factor subunit of the RNA polymerase enzyme is shown in dark blue, and binding to DNA occurs at -35 and -10 hexamers, as well as TGn and extended UP elements in some instances. Bottom: Transcription factors (commonly present as dimers) can interact with 1) the RNA polymerase alpha subunit (class I activation), 2) sigma factor subunit 4 (class II activation) or 3) at/near the -35 and -10 promoter elements (class III activation; Browning and Busby 2004).

The binding region in the promoter of each gene can be identified by a specific sigma factor. Sigma factors are categorized in two broad families based on their similarity to sigma factors in *E. coli*: sigma 70 and sigma 54 (referring to their respective masses in kilodaltons (kD) (Mathews et al. 2000). Sigma 70 factors are the predominant type of sigma factor and can be further classified into three groups based on the genes with which they associate: group 1 (primary sigma factors – most genes, such as housekeeping genes), group 2 (non-essential genes, such as those used by

bacteria in stationary phase) and group 3 (alternative/species specific genes, such as sporulation or heat shock) (Wösten 1998). Sigma factors are also used for the transcription of the chloroplast genome in higher plants (Hakimi et al. 2000). Cyanobacteria use specific sigma factors (composing a subgroup of group 2) that can, for example, modify gene expression in circadian rhythms (Tsinoremas et al., 1996) or in response to carbon and nitrogen availability (Caslake et al. 1997).

Another necessary component of bacterial gene expression is the regulatory network of transcription factors that govern transcription in response to local environmental changes. Transcription factors are proteins that bind in proximity to nearby genes and are responsible for gene activation, gene repression, or both (Figure 1.8). Several hundred transcription factors have been identified in *E. coli*, many of which, known as global regulators, act as part of a cascade system where they initiate expression of other transcription factors (Babu and Teichmann 2003). Transcription factors associated with bacterial secondary metabolites are perhaps best known in the genus *Streptomyces* (Cundliffe 2006), which have been described for a number of antibiotics (Bibb 1996) and siderophores (Tunca et al. 2007). Other secondary metabolites for which transcription factors or other regulatory genes have been characterized include the toxin hemolysin II from *Bacillus cereus* (Kovalevskiy et al., 2007) and stigmatellin from the myxobacterium *Cystobacter fuscus* (Rachid et al., 2006). In the case of *C. fuscus*, the stigmatellin regulator (*stiR*) is the first secondary metabolite regulatory element described for a myxobacterium. In the fungus *Aspergillus nidulans*, Bergmann et al. (2007) activated a silent PKS-NRPS pathway (= orphan gene cluster) via transformation of a construct containing a copy of a normally

inactive regulatory element from the gene cluster. With the regulatory gene in this construct under the control of an inducible promoter, the authors were successful in driving transcription and translation of the pathway. The PKS-NRPS derived aspyridones were subsequently isolated and identified.

A few gene expression studies have been conducted in cyanobacteria to determine the identity and activity of inducible promoters. Qi et al. (2005) used a *Synechococcus* inducible nitrate reductase (*nirA*) promoter, enhanced by the presence of nitrate and repressed by ammonia, to drive transgenic expression in *Synechocystis*. This promoter was used to drive expression of the *Arabidopsis thaliana* gene for *p*-hydroxyphenylpyruvate dioxygenase (hpd_{At}), a gene that catalyzes the formation of homogentisic acid and is important in tocopherol (vitamin E) biosynthesis. Production of tocopherols in *Synechocystis* was increased five-fold by hpd_{At} over-expression. In *Lyngbya majuscula*, Leitão et al. (2005) isolated the genes and putative promoter regions for a hydrogenase (*hupL*) thought to be responsible for scavenging H_2 during nitrogen fixation. The putative promoter region appeared to be regulated by changes in light levels.

The best studied cyanobacterial natural product gene cluster in terms of its regulation is that encoding the microcystins. Expression of the microcystin gene cluster is positively correlated with increased light intensity and red light (Kaebernick et al. 2000). The gene cluster also has different transcription start sites depending on light levels (Kaebernick et al. 2002). Other environmental factors have been evaluated for their effects on microcystin production, and increasing evidence suggests that iron may be important. Transcription of genes from the microcystin gene cluster increases

with iron starvation (Sevilla et al. 2008), and in the presence of iron, a ferric uptake regulator (Fur) protein appears to bind to the microcystin bidirectional promoter and decrease microcystin production (Martin-Luna et al. 2006). Because it complexes with iron and other metals (Saito et al. 2008), microcystin may function as a siderophore-like molecule. Alternatively, microcystin has been proposed to serve in intraspecies communication, where release of the compound is interpreted as cell death by other *Microcystis* sp. and causes increased microcystin production to enhance localized toxicity (Schatz et al. 2007). Contrary to gene clusters found in certain bacteria such as actinomycetes (Bibb 1996), none of the biosynthetic pathways identified to date in *Lyngbya* contain any putative transcription factors within the cluster boundaries, which has led to the suggestion that these pathways could be constitutively expressed (Edwards et al. 2004) or alternatively, pathway transcription factors could be encoded elsewhere in the genomes of the producing strains.

A final, emerging issue that complicates natural product discovery efforts from marine filamentous cyanobacteria is the large degree of phylogenetic ambiguity that has resulted from the use of morphological techniques and comparisons with freshwater cyanobacteria to define taxonomic lineages in marine systems, and the lack of genomic information for the *Lyngbya* genus (as well as other natural product rich species). For most of the modern marine natural products era, compounds isolated from marine cyanobacteria have often been linked to field collections of *Lyngbya majuscula* or *Lyngbya* sp. based on morphological identification (Nogle and Gerwick 2002, Balunas et al. 2010). Collectively, 35% of all marine natural products isolated from cyanobacteria are associated with the *Lyngbya* genus, with over 75% of these

associated with *L. majuscula* as mentioned above (Tidgewell et al. 2010). Recent efforts to better understand the phylogenetics of *Lyngbya* have revealed that based on 16S rRNA gene comparisons, the *Lyngbya* genus forms three distinct lineages, with a marine lineage entirely separate from that containing *Lyngbya* strains from freshwater and brackish environments (Sharp et al. 2009; Engene et al. 2010). This distinction between freshwater and marine *Lyngbya* clades has prompted taxonomic revision of this genus, which will likely result in a new genus designation for the marine *Lyngbya* strains (Engene et al., unpublished data). Moreover, the polyphyly found for *Lyngbya* strongly implies that the percentage of natural products attributed to "*Lyngbya*" strains is higher than in reality, and it is therefore unclear as to how many natural products a given marine *Lyngbya* strain can produce. Although over 30 cyanobacterial genomes have been sequenced, marine sequencing projects have been limited to pelagic genera such as *Prochlorococcus*, *Synechococcus*, and *Trichodesmium* (Jones et al. 2010). To date, genomic information for marine *Lyngbya* and other natural product producing marine cyanobacteria has been unavailable.

1.5 Contents of the dissertation

The focus of this doctoral dissertation is to investigate the biosynthesis of natural products from marine filamentous cyanobacteria, with the overall aim of improving access to these compounds for drug discovery and biotechnology. To this end, the results of several projects studying gene cluster transcription and regulation, protein and gene cluster heterologous expression, and genome sequencing are presented.

Chapter 2 of this dissertation discusses experiments used in the transcriptional analysis of the jamaicamides, mixed PKS/NRPS neurotoxins from the Jamaican strain *Lyngbya majuscula* JHB. Results presented include identification of the primary pathway promoter and putative TSS, as well as analysis of the pathway intergenic regions, using Reverse Transcription (RT-PCR). Most of the jamaicamide gene cluster intergenic regions were also evaluated for their potential promoter activity using a β-galactosidase reporter assay in *E. coli.* This chapter is taken in part from Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.

Chapter 3 of this dissertation is focused on experiments aimed at identifying how expression of the jamaicamide gene cluster may be regulated by genetic and environmental factors. This chapter includes protein isolation experiments wherein possible regulatory proteins associated with the jamaicamide promoter region were obtained, identified using mass spectrometry, heterologously expressed in *E. coli*, and used in *in vitro* ElectroMobility Shift Assays (EMSAs) to assess their binding to the intergenic region encompassing the primary jamaicamide promoter and upstream noncoding sequence of *jamA*. In addition, production and turnover of jamaicamides A and B was measured using a novel mass spectrometry approach in combination with stable isotope feeding, and various environmental factors were assessed for their effects on jamaicamide biosynthesis. This chapter is taken in part from Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis

of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247, and Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226-5231.

Chapter 4 includes efforts aimed at the heterologous production of the lyngbyatoxin biosynthetic gene cluster from a Hawaiian strain of *Lyngbya majuscula*, as well as individual lyngbyatoxin proteins, in *Streptomyces coelicolor*, *Anabaena* (*Nostoc*) sp. PCC7120, and *Escherichia coli* BL21. The text of this chapter is taken in part from Jones AC, Ottilie S, Eustaquio AS, Edwards DJ, Gerwick L, Moore BS, Gerwick WH. Evaluation of *Streptomyces coelicolor* as a heterologous expression host for natural products from the filamentous cyanobacterium *Lyngbya majuscula*. Manuscript in preparation.

Chapter 5 describes the genome sequencing of *Lyngbya majuscula* 3L, a *Lyngbya* strain from Curaçao, Netherlands Antilles that produces several natural products, including the cytotoxin curacin A, the molluscicide barbamide, and the lipopeptide carmabin. This genome project included isolation of genomic DNA, sequencing using a combination of Sanger and 454 technologies, and stringent binning and assembly to yield a *L. majuscula* 3L draft genome. Annotation and analysis of secondary metabolite gene clusters, organic extraction and profiling, proteomic analyses, and wet lab experiments were all conducted to support initial findings from surveys of *L. majuscula* 3L genome information. This chapter is taken in full from Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S,

Hoover H, Rothmann M, Lasken RS, Yates III JR, Reinhardt R, Kube M, Burkart MD, Allen EE, Dorrestein PC, Gerwick WH, Gerwick L (2011) Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci USA* doi:10.1073/pnas.1101137108.

Chapter 6 provides a general summary of this dissertation and perspectives on future research directions predicated on the research presented here. Chapter 6 concludes the dissertation.

1.6 References

Alberts B, Bray D, Johnson A, Lewis J, Raff M, Roberts K, Walter P (1998) Essential cell biology: An introduction to the molecular biology of the cell. Garland Publishing, New York.

Anderson SL, Oloo AJ, Gordon DM, Ragama OB, Aleman GM, Berman JD, Tang DB, Dunne MW, Shanks GD (1998) Successful double-blinded, randomized, placebo-controlled field trial of azithromycin and doxycycline as prophylaxis for malaria in Western Kenya. *Clin Infect Dis* 26:146-150.

Antoniou MG, de la Cruz AA, Dionysiou DD (2005) Cyanotoxins: New generation of water contaminants. *J Environ Eng* 131: 1239-1243.

Babu MM, Teichmann SA (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nuc. Acid. Res.* 31:1234-1244.

Balunas MJ, Linington RG, Tidgewell K, Fenner AM, Ureña L-D, Togna GD, Kyle DE, Gerwick WH (2010) Dragonamide E, a modified linear lipopeptide from *Lyngbya majuscula* with antileishmanial activity. *J Nat Prod* 73:60-66.

Bergmann S, Schümann J, Scherlach K, Lange C, Brakhage AA, Hertweck C (2007) Genomics-driven discovery of PKS-NRPS hybrid metabolites from *Aspergilus nidulans. Nat Chem Biol* 3:213-217.

Bibb M (1996) The regulation of antibiotic production in *Streptomyces coelicolor* A3(2). *Microbiol* 142:1335-1344.

Blunt JW, Copp BR, Hu W-P, Munro MHG, Northcote PT, Prinsep MR (2009) Marine natural products. *Nat Prod Rep* 26:170-244.

Browning DF, Busby SJW (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2:57-65.

Burja AM, Banaigs B, Abou-Mansour E, Burgess JG, Wright PC (2001) Marine cyanobacteria – a prolific source of natural products. *Tetrahedron* 57:9347-9377.

Capper A, Tibbetts IR, O'Neil JM, Shaw GR (2005) The fate of *Lyngbya majuscula* toxins in three potential consumers. *J Chem Ecol* 31:1595-1606.

Cardellina II JH, Marner FJ, Moore RE (1979) Malyngamide A, a novel chlorinated metabolite of the marine cyanophyte *Lyngbya majuscula*. *J Am Chem Soc* 101:240-242.

Cardellina II JH, Marner F-J, Moore RE (1979) Seaweed dermatitis: Structure of lyngbyatoxin A. *Science* 204:193-195.

Cardellina II JH, Moore BS (2010) Richard E. Moore (1933-2007). *J Nat Prod* 73: 301-302.

Cardellina II JH, Moore RE (1980) Malyngic acid, a new fatty acid from *Lyngbya majuscula*. *Tetrahedron* 36:993-996.

Carmichael WW (2001) Health effects of toxin-producing cyanobacteria: "The cyanoHABs". *Human Ecol Risk Assess* 7:1393-1407

Carmichael WW, Evans WR, Yin QQ, Bell P, Moczydlowski E (1997) Evidence for paralytic shellfish poisons in the freshwater cyanobacterium *Lyngbya wollei* (Farlow ex Gomont) comb. nov. *Appl Environ Microbiol* 63: 3104-3110.

Caslake LF, Gruber TM, Bryant DA (1997) Expression of two alternative sigma factors of *Synechococcus* sp. strain PCC 7002 is modulated by carbon and nitrogen stress. *Microbiol* 143:3807-3818.

Challis G (2006) Engineering *Escherichia coli* to produce nonribosomal peptide antibiotics. *Nat Chem Biol* 2:398-400.

Chang Z, Flatt P, Gerwick WH, Nguyen VA, Willis CL, Sherman DH (2002) The barbamide biosynthetic gene cluster: a novel cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296: 235-247.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 67:1356-1367.

Chopra I, Roberts M (2001) Tetracycline antibiotics: Mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol Mol Bio Rev* 65:232-260.

Cooper S, Helmstetter CE (1968) Chromosome replication and the division cycle of *Escherichia coli* B/r. *J Mol Biol* 31:519-540.

Cundliffe E (2006) Antibiotic production by actinomycetes: The Janus faces of regulation. *J Ind Microbiol Biotechnol* 33:500-506.

Dillehay TD, Ramírez C, Pino M, Collins MB, Rossen J, Pino-Navarro JD (2008) Monte Verde: Seaweed, food, medicine, and the peopling of South America. *Science* 320:784-786.

Donia MS, Hathaway BJ, Sudek S, Haygood MG, Rosovitz MJ, Ravel J, Schmidt EW (2006) Natural combinatorial peptide libraries in cyanobacterial symbionts of marine ascidians. *Nat Chem Biol* 2:729-735.

Donia MS, Ravel J, Schmidt EW (2008) A global assembly line for cyanobactins. *Nat Chem Biol* 4:341-343.

Drews J (2000) Drug discovery: A historical perspective. *Science* 287:1960-1964.
Du L, Sanchez C, Shen B (2001) Hybrid peptide-polyketide natural products: biosynthesis and prospects toward engineering novel molecules. *Met Engineer* 3:78-95.

Edwards DJ, Gerwick WH (2004) Lyngbyatoxin biosynthesis: Sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J Am Chem Soc* 126: 11432-11433

Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* 11: 817-833

Engene N, Coates RC, Gerwick WH (2010) 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46:591-601.

Esquenazi E (2010) Spatial and Temporal Dynamics of Marine Natural Products Biosynthesis. PhD dissertation, University of California San Diego.

Esquenazi E, Coates C, Simmons L, Gonzalez D, Gerwick WH, Dorrestein PC (2008) Visualizing the spatial distribution of secondary metabolites produced by marine cyanobacteria and sponges *via* MALDI-TOF imaging. *Mol BioSyst* 4:562-570.

Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural product biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226-5231.

Eustáquio AS, Gust B, Galm U, Li S-M, Chater KF, Heide L (2005) Heterologous expression of novobiocin and chlorobiocin biosynthetic gene clusters. *Appl Environ Microbiol* 71:2452-2459.

Falcón LI, Magallón S, Castillo A (2010) Dating the cyanobacterial ancestor of the chloroplast. *ISME J* 4:777-783.

Feling RH, Buchanan GO, Mincer TJ, Kauffman CA, Jensen PR, Fenical W (2003) Salinosporamide A: A highly cytotoxic proteasome inhibitor from a novel microbial source, a marine bacterium of the new genus *Salinospora*. *Angew Chem Int Ed Engl* 42:355-357.

Fischbach MA, Walsh CT (2006) Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem Rev* 106: 3468-3496.

Flatt PM, O'Connell SJ, McPhail KL, Zeller G, Willis CL, Sherman DH, Gerwick WH (2006) Characterization of the initial enzymatic steps of barbamide biosynthesis. *J Nat Prod* 69:938-944.

Fraenkel GS (1959) The raison d'Etre of secondary plant substances. *Science* 129:1466-1470.

Galonić DP, Vaillancourt FH, Walsh CT (2006) Halogenation of unactivated carbon centers in natural product biosynthesis: Trichlorination of leucine during barbamide biosynthesis. *J Am Chem Soc* 128:3900-3901.

Gerwick WH, Coates RC, Engene N, Gerwick L, Grindberg RV, Jones AC, Sorrels CM (2008) Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3:277-284.

Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blokhin A, Slate DL (1994) Structure of curacin A, a novel antimitoic, antiproliferative and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*. *J Org Chem* 59:1243-1245.

Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu W-T, Gerwick L, Dorrestein PC, Pevzner P, Lasken R, Gerwick WH (2011) Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS ONE* (in press).

Gu L, Geders TW, Wang B, Gerwick WH, Håkansson K, Smith JL, Sherman DH (2007) GNAT-like strategy for polyketide chain initiation. *Science* 318:970-974.

Gu L, Jia J, Liu H, Håkansson K, Gerwick WH, Sherman DH (2006) Metabolic coupling of dehydration and decarboxylation in the curacin A pathway: Functional identification of a mechanistically diverse enzyme pair. *J Am Chem Soc* 128:9014-9015.

Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Håkansson K, Wipf P, Smith JL, Gerwick WH, Sherman DH (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature* 459:731-735.

Gu L, Wang B, Kulkarni A, Gehret JJ, Lloyd KR, Gerwick L, Gerwick WH, Wipf P, Håkansson K, Smith JL, Sherman DH (2009) Polyketide decarboxylative chain termination preceded by *O*-sulfonation in curacin A biosynthesis. *J Am Chem Soc* 131:16033-16035.

Hakimi MA, Privat I, Valay JG, Lerbs-Mache S (2000) Evolutionary conservation of C-terminal domains of primary sigma 70–type transcription factors between plants and bacteria. *J Biol Chem* 275:9215-9221.

Heide L (2009) Aminocoumarins: Mutasynthesis, chemoenzymatic synthesis, and metabolic engineering. *Meth Enzymol* 459:437-455.

Hooper GJ, Orjala J, Schatzman RC, Gerwick WH (1998) Carmabins A and B, new lipopeptides from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 61:529-533.

Jabba SV, Prakash A, Dravid SM, Gerwick WH, Murray TF (2010) Antillatoxin, a novel lipopeptide, enhances neurite outgrowth in immature cerebrocortical neurons through activation of voltage-gated sodium channels. *J Pharmacol Exper Therap* 332:698-709.

Johnson IS, Armstrong JG, Gorman M, Burnett Jr., JP (1963) The vinca alkaloids: A new class of oncolytic agents. *Cancer Res* 23:1390-1427.

Jones AC, Gu L, Sorrels CM, Sherman DH, Gerwick WH (2009) New tricks from ancient algae: Natural products biosynthesis in marine cyanobacteria. *Curr Op Chem Biol* 13:216-223.

Jones AC, Monroe EA, Eisman EB, Gerwick L, Sherman DH, Gerwick WH (2010) The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* 27:1048-1065.

Kaebernick M, Dittmann E, Börner T, Neilan BA (2002) Multiple alternate transcripts direct the biosynthesis of microcystin, a cyanobacterial nonribosomal peptide. *Appl Environ Microbiol* 68:449-455.

Kaebernick M, Neilan BA, Borner T, Dittmann E (2000) Light and the transcriptional response of the microcystin biosynthesis gene cluster. *Appl Environ Microbiol* 66:3387-3392.

Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M, Tabata S (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res* 8:205-213.

Koksharova OA, Wolk CP (2002) Genetic tools for cyanobacteria. *Appl Microbiol Biotechnol* 58:123-137.

Kovalevskiy OV, Lebedev AA, Surin AK, Solonin AS, Antson AA (2007) Crystal structure of *Bacillus cereus* HIyIIR, a transcriptional regulator of the gene for pore-forming toxin hemolysin II. *J Mol Biol* 365:825-834.

Lane AL, Moore BS (2011) A sea of biosynthesis: Marine natural products meet the molecular age. *Nat Prod Rep* 28:411-428.

Lee Y-J, Lee JW, Oh C, Heo S-J, Kang D-H, Shin HJ, Lee H-S (2010) Identification of a novel chlorinated fatty acid from the marine *Rhodopirellula baltica*. *Bull Korean Chem Soc* 31:3421-3222.

Leitão E, Oxelfelt F, Oliveira P, Moradas-Ferreira P, Tamagnini P (2005) Analysis of the *hupSL* operon of the nonheterocystous cyanobacterium *Lyngbya majuscula* CCAP 1446/4: Regulation of transcription and expression under a light-dark regimen. *Appl Env Microbiol* 71:4567-4576.

Levine DP (2006) Vancomycin: A History. *Clin Infect Dis* 42 (Supplement 1) S5:S12.
Liu Y, Law BK, Luesch H (2009) Apratoxin A reversibly inhibits the secretory pathway by preventing cotranslational translocation. *Mol Pharmacol* 76:91-104.

Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, Joewono I, Rusch D, Chisholm SW, van der Donk WA (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. *Proc Natl Acad Sci USA* 107:10430-10435.

Luesch H, Chanda SK, Raya RM, DeJesus PD, Orth AP, Walker JR, Belmonte JCI, Schultz PG (2005) A functional genomics approach to the mode of action of apratoxin A. *Nat Chem Biol* 2:158-167.

Luesch H, Yoshida WY, Moore RE, Paul VJ, Corbett TH (2001) Total structure determination of apratoxin A, a potent novel cytotoxin from the marine cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 123:5418-5423.

Martin-Luna B, Sevilla E, Hernandez JA, Bes MT, Fillat MF, Peleato ML (2006) Fur from *Microcystis aeruginosa* binds in vitro promoter regions of the microcystin biosynthesis gene cluster. *Phytochemistry* 67: 876-881

Mathews CK, van Holde KE, Ahern KG (2000) Biochemistry, 3rd Ed. Benjamin Cummings, San Francisco.

McGovern PE, Mirzoian A, Hall GR (2009) Ancient Egyptian herbal wines. *Proc Natl Acad Sci USA* 106:7361-7366.

Mevers E, Liu WT, Engene N, Mohimani H, Byrum T, Pevzner PA, Dorrestein PC, Gerwick WH. Cytotoxic veraguamides, alkynyl bromine containing cyclic depsipeptides from the marine cyanobacterium cf. *Oscillatoria margaritifera*. *J Nat Prod*, in review.

Moffitt MC, Neilan BA (2004) Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Appl Environ Microbiol* 70: 6353-6362

Molinski TF, Dalisay DS, Lievens SL, Saludes JP (2009) Drug development from marine natural products. *Nat Rev Drug Disc* 8:69-85.

Muir JC, Pattenden G, Ye T (1998) A concise total synthesis of (+)-curacin A, a novel cyclopropyl-substituted thiazoline from the cyanobacterium *Lyngbya majuscula*. *Tett Lett* 39:2861-2864.

Mynderse JS, Moore RE, Kashiwagi M, Norton TR (1977) Antileukemia activity in the Osillatoriaceae: Isolation of debromoaplysiatoxin from *Lyngbya*. *Science* 196:538-540.

Newman DJ, Cragg GM (2007) Natural products as sources of new drugs over the last 25 years. *J Nat Prod* 70:461-477.

Nogle LM, Gerwick WH (2002) Somocystinamide A, a novel cytotoxic disulfide dimer from a Fijian marine cyanobacterial mixed assemblage. *Org Lett* 4:1095-1098.

Nogle LM, Okino T, Gerwick WH (2001) Antillatoxin B, a neurotoxic lipopeptide from the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 64:983-985.

Oliynyk M, Samborskyy M, Lester JB, Mironenko T, Scott N, Dickens S, Haydock SF, Leadlay PF (2007) Complete genome sequence of the erythromycin-producing bacterium *Saccharopolyspora erythraea* NRRL23338. *Nat Biotechnol* 25:447-453.

Orjala J, Gerwick WH (1996) Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 59:427-430.

Orjala J, Nagle DG, Hsu V, Gerwick WH (1995) Antillatoxin: An exceptionally ichthyotoxic cyclic lipopeptide from the tropical cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 117:8281-8282.

Qi Q, Hao M, Ng W, Slater SC, Baszis SR, Weiss JD, Valentin HE (2005)Application of the *Synechococcus nirA* promoter to establish an inducible expression system for engineering the *Synechocystis* tocopherol pathway. *Appl Env Microbiol* 71:5678-5684.

Rachid S, Sasse F, Beyer S, Müller R (2006) Identification of StiR, the first regulaor of secondary metabolite formation in the myxobacterium *Cystobacter fuscus* Cb f17.1. *J Biotech* 121:429-441.

Ramaswamy AV, Sorrels CM, Gerwick WH (1997) Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 70:1977-1986.

Rasmussen B, Fletcher IR, Brocks JJ, Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* 455:1101-1104.

Read JA, Walsh CT (2007) The lyngbyatoxin biosynthetic assembly line: Chain release by four-electron reduction of a dipeptidyl thioester to the corresponding alcohol. *J Am Chem Soc* 129:15762-15763.

Rinehart KL, Namikoshi M, Choi BW (1994) Structure and biosynthesis of toxins from blue-green algae (cyanobacteria). *J Appl Phycol* 6: 159-176.

Rossi JV, Roberts MA, Yoo H-D, Gerwick WH (1997) Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically-useful natural metabolite curacin A. *J Appl Phycol* 9:195-204.

Saito K, Sei Y, Miki S, Yamaguchi K (2008) Detection of microcystin-metal complexes by using cryospray ionization-Fourier transform ion cyclotron resonance mass spectrometry. *Toxicon* 51:1496-1498

Schatz D, Keren Y, Vardi A, Sukenik A, Carmeli S, Börner T, Dittmann E, Kaplan A (2007) Towards clarification of the biological role of microcystins, a family of cyanobacterial toxins. *Environ Microbiol* 9:965-970.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J (2005) Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci USA* 102:7315-7320.

Seckbach J, Ed. (2007) Algae and Cyanobacteria in Extreme Environments. Cellular Origin, Life in Extreme Habitats, and Astrobiology, vol 11. Springer: the Netherlands.

Sevilla E, Martin-Luna B, Vela L, Bes MT, Fillat MF, Peleato ML (2008) Iron availability affects mcyD expression and microcystin-LR synthesis in *Microcystis aeruginosa* PCC7806. *Environ Microbiol* 10: 2476-2483.

Sharp K, Arthur KE, Gu L, Ross C, Harrison G, Gunasekera SP, Meickle T, Matthew S, Luesch H, Thacker RW, Sherman DH, Paul VJ (2009) Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* species (*Cyanobacteria: Oscillatoriales*) from reefs of Southeastern Florida. *Appl Environ Microbiol* 75:2879-2888.

Simmons TL, Coates RC, Clark BR, Engene N, Gonzalez D, Esquenazi E, Dorrestein PC, Gerwick WH (2008) Biosynthetic origin of natural products isolated from marine microorganism-invertebrate assemblages. *Proc Natl Acad Sci USA* 105:4587-4594.

Sitachitta N, Márquez BL, Williamson RT, Rossi J, Roberts MA, Gerwick WH, Nguyen V-A, Willis CL (2000) Biosynthetic pathway and origin of the chlorinated methyl group in barbamide and dechlorobarbamide, metabolites from the marine cyanobacterium *Lyngbya majuscula*. *Tetrahedron* 56:9103-9113.

Sudek S, Lopanik NB, Waggoner LE, Hildebrand M, Anderson C, Liu H, Patel A, Sherman DH, Haygood MG (2007) Identification of the putative bryostatin polyketide synthase gene cluster from "*Candidatus* Endobugula sertula", the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *J Nat Prod* 70:67-74.

Tan LT, Márquez BL, Gerwick WH (2002) Lyngbouilloside, a novel glycosidic macrolide from the marine cyanobacterium *Lyngbya bouillonii*. *J Nat Prod* 65:925-928.

Tang L, Shah S, Chung L, Carney J, Katz L, Khosla C, Julien B (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science* 287:640-642.

Tidgewell K, Clark BR, Gerwick WH (2010) The natural products chemistry of cyanobacteria. In Mander L, Lui H-W, Eds. Comprehensive Natural Products II Chemistry and Biology, Oxford, Elsevier 2:141-188.

Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA (2000) Structural organization of microcystin biosynthesis in *M. aeruginosa* PCC7806: an integrated peptide-polyketide synthetase system. *Chem Biol* 7:753-764.

Ting CS, Rocap G, King J, Chisholm SW (2002) Cyanobacterial photosynthesis in the oceans: The origins and significance of divergent light-harvesting strategies. *Trends Microbiol* 10:134-142.

Tsinoremas NF, Ishiura M, Kondo T, Anderson CR, Tanaka K, Takahashi H, Johnson CH, Golden SS (1996) A sigma factor that modifies the circadian expression of a subset of genes in cyanobacteria. *EMBO J* 15*:*2488-2495.
Tunca S, Barreiro C, S-L A, Coque JJR, Martín JF (2007) Transcriptional regulation of the desferrioxamine gene cluster of *Streptomyces coelicolor* is mediated by binding of DmdR1 to an iron box in the promoter of the *desA* gene. *FEBS J* 274: 1110-1122.

Verdier-Pinard P, Lai J-Y, Yoo H-D, Yu J, Marquez B, Nagle DG, Nambu M, White JD, Falck JR, Gerwick WH, Day BW, Hamel E (1998) Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. *Mol Pharmacol* 53:62-76.

Wani MC, Taylor HL, Wall ME, Coggon P, McPhail AT (1971) Plant antitumor agents. VI. The isolation and structure of taxol, a novel antileukemic and antitumor agent from *Taxus brevifola*. *J Am Chem Soc* 93:2325-2327.

White JD, Kim T-S, Nambu M (1997) Absolute configuration and total synthesis of (+)-curacin A, an antiproliferative agent from the cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 119:103-111.

White NJ (1996) The treatment of malaria. *N Engl J Med* 335:800-806.

Wösten MMSM (1998) Eubacterial sigma-factors. *FEMS Microb Rev* 22:127-150.

Wrasidlo W, Mielgo A, Torres VA, Barbero S, Stoletov K, Suyama TL, Klemke RL, Gerwick WH, Carson DA, Stupack DG (2008) The marine lipopeptide somocystinamide A triggers apoptosis via caspase 8. *Proc Natl Acad Sci USA* 105:2313-2318.

Wu M, Okino T, Nogle LM, Marquez BL, Williamson RT, Sitachitta N, Berman FW, Murray TF, McGough K, Jacobs R, Colsen K, Asano T, Yokokawa F, Shioiri T, Gerwick WH (2000) Structure, synthesis, and biological properties of kalkitoxin, a novel neurotoxin from the marine cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 122:12041-12042.

Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM (2001) Unicelluar cyanobacteria fix N2 in the subtropical North Pacific Ocean. *Nature* 412:635-638.

CHAPTER 2

Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* JHB

2.1 Abstract

The marine cyanobacterium *Lyngbya majuscula* is a prolific producer of bioactive secondary metabolites. Although biosynthetic gene clusters encoding several of these compounds have been identified, little is known about how these clusters of genes are transcribed or regulated, and techniques targeting genetic manipulation in *Lyngbya* strains have not yet been developed. This chapter contains experiments conducting transcriptional analyses of the jamaicamide gene cluster from a Jamaican strain of *Lyngbya majuscula*. An unusually long untranslated leader region of approximately 840 bp is located between the jamaicamide transcription start site (TSS) and gene cluster start codon. All of the intergenic regions between the pathway ORFs were transcribed into RNA in RT-PCR experiments; however, a promoter prediction program indicated the possible presence of promoters in multiple intergenic regions. Because the functionality of these promoters could not be verified *in vivo*, a reporter gene assay in *E. coli* was used to show that several of these intergenic regions, as well as the primary promoter preceding the TSS, are capable of driving β-galactosidase production. This is the first attempt to characterize the transcription of secondary metabolism in a marine cyanobacterium.

2.2. Introduction

Over the past 30 years, the search for bioactive secondary metabolites (natural products) from marine organisms has yielded a wealth of new molecules (estimated at ~17,000) with many fundamentally new chemotypes and extraordinary potential for biomedical research and applications (Blunt et al. 2009 and previous references therein). Marine cyanobacteria continue to be among the most fruitful sources of marine natural products, with nearly 700 compounds described (Tan 2007; Tidgewell et al. 2010). The filamentous marine cyanobacterium *Lyngbya majuscula* (Gomont) is of particular importance, as approximately 35% of all cyanobacterial bioactive compounds have been reported from the genus *Lyngbya*, with 76% of these coming from *L. majuscula* (Tidgewell et al. 2010). More recently, compound isolation and structure elucidation from *L. majuscula* has been complemented with the characterization of biosynthetic gene clusters that encode a number of these compounds. The gene clusters for several potent anticancer and neurotoxic agents such as curacin A, barbamide, and the jamaicamides have provided new insight into the biosynthetic strategies and logic used by this organism for compound production, as well as unique enzymes involved in unprecedented molecular tailoring reactions (Chang et al. 2002; Chang et al. 2004; Edwards et al. 2004; Gu et al. 2007).

Despite considerable interest in pursuing cyanobacterial lead compounds as potential drug candidates, an adequate supply of these compounds for clinical research is often impossible to obtain without impractically large scale field collections or sophisticated and expensive synthetic methods (Cragg et al. 1999, Suyama and Gerwick 2008). With some notable examples (Pfeifer et al. 2003; Schmidt et al. 2005;

Watanabe et al. 2006; Wilkinson and Micklefield 2007) it has been difficult to induce microbial gene clusters to produce their natural products in heterologous hosts, and thus this technology is not always predictable (Galm and Shen 2006). Equally problematic, filamentous marine cyanobacteria such as *Lyngbya* grow slowly in laboratory culture, with doubling times in some cases as long as 18 days (Rossi et al. 1997).

One avenue for increasing compound production from marine cyanobacteria could be to take advantage of regulatory elements associated with a biosynthetic gene cluster of interest. Although genetic controls of several primary metabolic functions in cyanobacteria including circadian rhythms (Ishiura et al. 1998), heterocyst development (Yoon and Golden 2001), and nutrient uptake (Lindell et al. 1998) have been described, information regarding transcriptional regulation of cyanobacterial secondary metabolites is currently limited to freshwater toxins such as the microcystins. The microcystins are potent hepatotoxins synthesized by several freshwater cyanobacteria of worldwide occurrence (de Figueiredo et al. 2004) and are generated via a mixed polyketide synthase/non-ribosomal peptide synthetase (PKS/NRPS) gene cluster (Tillett et al. 2000). Expression of the microcystin gene cluster is positively correlated with increased light intensity and red light in particular (Kaebernick et al. 2000). Moreover, the gene cluster has different transcriptional start sites depending on light levels (Kaebernick et al. 2002). Other environmental factors have been evaluated for their effects on microcystin production, and increasing evidence suggests that iron may be important. Transcription of genes from the microcystin gene cluster increases with iron starvation (Sevilla et al. 2008),

and in the presence of iron, a ferric uptake regulator (Fur) protein appears to bind to the microcystin bidirectional promoter and may decrease microcystin production (Martin-Luna et al. 2006). Because it complexes with iron and other metals (Saito et al. 2008) microcystin may therefore function as a siderophore. Alternatively, microcystin has been proposed to serve in intraspecies communication, where release of the compound is interpreted as cell death by other *Microcystis* sp. and causes increased microcystin production to enhance localized toxicity (Schatz et al. 2007).

As with microcystin, many of the toxins found in *L. majuscula* are also produced by gene clusters comprised of PKS/NRPS architecture. PKS/NRPS gene clusters in other bacteria have been found to include imbedded regulatory proteins, such as the ***S****treptomyces* **A**ntibiotic **R**egulatory **P**roteins (SARPs) found within the confines of several antibiotic pathways in *Streptomyces* (Bibb 2005). However, cyanobacterial natural product gene clusters identified to date do not contain any putative regulatory proteins.

Insight into the mechanisms used by *L. majuscula* in the transcription of secondary metabolite gene clusters could be of significant value in enhancing the overproduction of potential drug leads in laboratory cultures. Increased compound yield would reduce the need and environmental impact of repeated large scale field collections or the time and expense of chemical synthesis. Additionally, because the secondary metabolite biosynthetic gene clusters identified thus far from *L. majuscula* have been obtained from different strains of the same species, transcription of each pathway could be under similar mechanisms of regulation.

This chapter provides a transcriptional analysis of the jamaicamide gene cluster from *Lyngbya majuscula* JHB, and is the first such effort for a secondary metabolite gene cluster from a marine cyanobacterium. The jamaicamides are mixed PKS/NRPS neurotoxins that exhibit sodium channel blocking activity and fish toxicity. The molecules contain unusual structural features including a vinyl chloride and alkynyl bromide (Edwards et al. 2004). The gene cluster encoding jamaicamide biosynthesis is 57 kbp in length, and is composed of 17 ORFs that encode proteins ranging in length from 80 to 3936 amino acids. Intergenic regions between 5 and 442 bp are located between all but two of the ORFs, and a region of approximately 1700 bp exists between the first jamaicamide ORF (*jamA*, a hexanoyl ACP synthetase) and the closest upstream (5') ORF outside of the cluster (a putative transposase). In this study, an RT-PCR experiment was used to locate the transcriptional start site (TSS) of the jamaicamide gene cluster. Because it is not yet possible to perform genetics in filamentous marine cyanobacteria such as *Lyngbya*, a series of reporter gene assays were used to validate the activity of the primary jamaicamide pathway promoter and identify several possible internal pathway promoters.

2.3. Results

2.3.1 RT-PCR using *L. majuscula* RNA to search for the transcriptional start site (TSS) and promoter regions in the jamaicamide pathway

The initial characterization of the jamaicamide gene cluster (Edwards et al. 2004) described that the first 16 ORFs of the gene cluster (*jamA*-*jamP*) are all transcribed in the forward direction, while the last ORF (*jamQ*, a putative

condensation domain thought to be involved in the cyclization of the pyrrolinone ring of the molecule) is transcribed in the reverse direction (Figure 2.1).

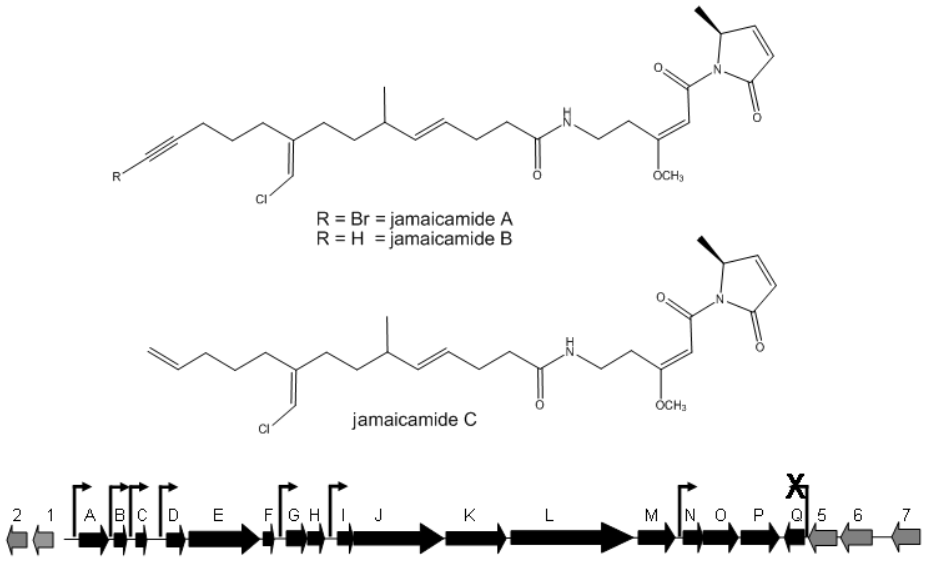


Figure 2.1: Structures of the jamaicamides and the jamaicamide biosynthetic gene cluster (Edwards et al. 2004). Genes associated with the pathway are represented by black arrows, and genes flanking the pathway are represented in gray. Elevated arrows above the upstream regions of selected open reading frames indicate where promoter activity was detected using the β-galactosidase reporter assay. The region upstream of *jamQ* did not have any detectable promoter activity in the assay (Jones et al. 2009).

In order to determine the location of the transcriptional start site (TSS) of the gene cluster, RNA was isolated from the jamaicamide producing strain of *Lyngbya majuscula* (JHB). First strand cDNA was synthesized using reverse transcriptase and a reverse primer designed as a complement to the 5' end of the *jamA* gene (Table S1). Initial experiments creating second strand cDNA using the first strand cDNA as template found that an unusually long untranslated leader region of at least 500 bp preceded *jamA.* A primer extension experiment was conducted in which second strand cDNA was amplified in 50 bp increments beyond this 500 bp location. The

experiment indicated that transcription of RNA began between 850 bp and 902 bp upstream of the *jamA* ORF start site (Figure 2.2).

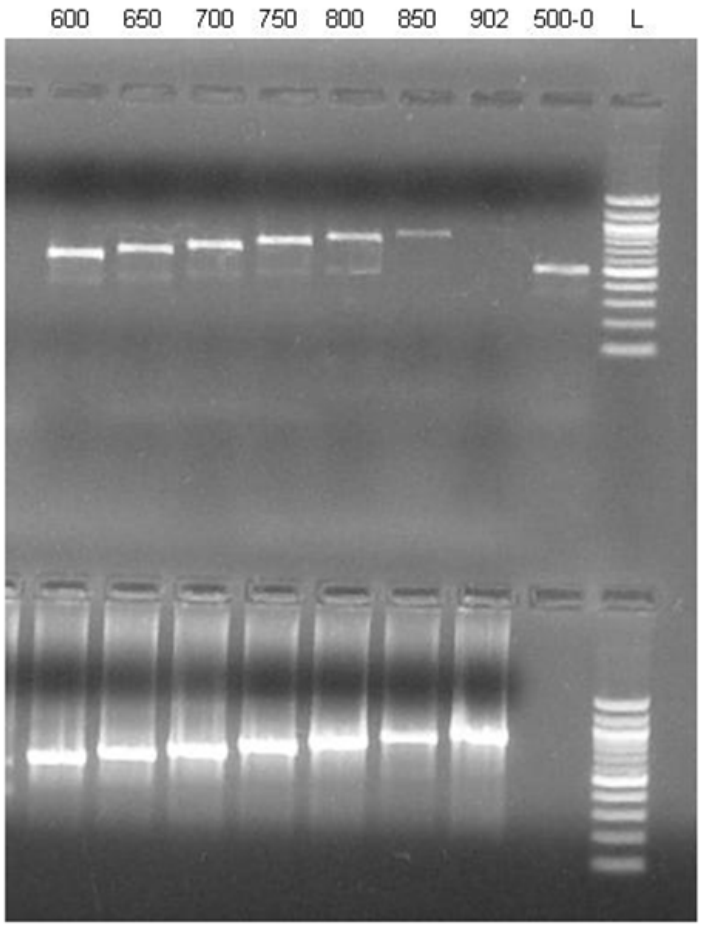


Figure 2.2: Transcription start site (TSS) primer extension experiment using first strand cDNA upstream of *jamA* (top) or jam fosmid (bottom) as PCR templates**.** The upstream region sizes (e.g., 600-0, 650-0) are indicated above each lane (Jones et al. 2009).

Using comparisons to consensus promoter and transcription start regions in *E. coli* (Aoyama et al. 1983; Hawley and McClure 1983; Lisser and Margalit 1993), a putative promoter was identified which, relative to a probable TSS (844 bp upstream

of*jamA*), included conserved hexamer RNA polymerase (RNAP) binding sites at -35 and -10 bp, a conserved extended -10 TGn region upstream of the -10 box, and an optimal DNA length between the hexamers (17 bp) (Figure 2.3).

Figure 2.3: Location of identified promoter regions and transcription start site (TSS) upstream of *jamA*. The consensus -35 and -10 boxes of each region are underlined. The conserved extended -10 TGn box of the primary pathway promoter is double underlined. The putative TSS is noted at +1, and was chosen based on similarities to the consensus *E. coli* TSS nucleotide region (Hawley and McClure 1983). The first four codons of the *jamA* gene are noted at the end of the sequence (Jones et al. 2009).

The next series of experiments evaluated whether the jamaicamide gene cluster contained non-transcribed intergenic regions between ORFs that could indicate the presence of breaks in the transcripts. Primers designed for those intergenic regions in the pathway 20 bp in size or larger (all but 2 intergenic regions) were used for synthesizing and amplifying cDNA to determine whether they were transcribed. All of

the intergenic regions in the jamaicamide pathway tested were amplified into second strand cDNA, including the intergenic region between *jamP* and *jamQ*. Intergenic regions between the two ORFs downstream of *jamQ* (putative transposases) were also transcribed. These results indicated that the majority of the jamaicamide gene cluster is composed of the operon *jamABCDEFGHIJKLMNOP*. Because no apparent breaks in transcription occurred between *jamQ* and at least the two neighboring downstream transposases (*ORF5* and *ORF6*) and a hypothetical protein (*ORF7*), one contiguous transcript may encode the translation of all of these proteins. Transcription of the intergenic region between *jamP* and *jamQ* indicated that a transcript including *jamP* must extend at least into the complementary strand of *jamQ* before termination, although transcription in the opposite direction would be necessary to generate *jamQ* mRNA.

2.3.2 Use of promoter prediction and β-galactosidase reporter gene assays to search for promoter activity

The large size (approximately 55 kbp) of the main jamaicamide operon (*jamA-P*) suggested that multiple promoters would likely be needed for efficient jamaicamide transcription. Because transcripts were found for each of the intergenic regions between the ORFs, these promoters may function intermittently and could be subject to promoter occlusion (Kaebernick et al. 2002). A software prediction program (BPROM, http://www.softberry.com) was used to predict whether the intergenic regions from the jamaicamide pathway contained conserved promoter binding regions.

Several of these regions were predicted to contain at least one potential pair of -35 and -10 binding sites (Table 2.1).

Table 2.1: Predicted -35 and -10 promoter regions (italics) and transcription start sites (TSS; nucleotides under +1 in table) for intergenic regions in the jamaicamide gene cluster (accession #AY522504). The bold ATG for the second *upjamI* indicates the start codon for the *jamI* gene (Jones et al. 2009).

Upstream region of gene	Predicted TSS location (bp)	ORF start (bp)	-35	-10	+1
up*jamA*	6626	6630	CTGAC*TTTCCA*CGACATGGGACTGATGGGAAATG	*TATATT*	TATTTGA
up*jamB*	8464	8591	GTGGG*TTGATT*TGATCAAGTTTGATGA	*TATAAT*	TTGATTTA
up*jamB*	8501	8591	TTTAA*TTTACA*GGGATACCGCCAATTCGG	*TAACCT*	GGAAAA
up*jamC*	9614	9718	AAAAC*TTGTCA*ACCTGAACAAGATCCTGAA	*CAAAAT*	ATTGTTG
up*jamD*	10433	10463	ACAGT*TTGATG*GTGCCGCTATTTTGAAGTTG	*GAAAAT*	TTTTTA
up*jamG*	18145	18222	ATTTG*TTGTTT*GGGAATCGGGAATTGG	*TATTAG*	TAGTGGAA
up*jamI*	20776	20982	CGGAA*TTCAAA*ATTCAAAATTCAAAATGCTTATG	*GATTAT*	GGAGTAAA
up*jamI*	20989	20982	CCAGG*TTGACA*AACCATTGATAAAGC	*TATAGT*	**ATG**TATTA
up*jamN*	51787	51811	TGGAG*TATAAA*AAACAGAGCCTGGTG	*ATAGT*	TAATTAA
up*jamQ*	63710[a]	63646[a]	GAACT*TTGAAT*CCTCTATTTTGAT	*TAAATT*	TGGAGA
	E. coli σ^{70} consensus promoter		TTGACA		TATAAT

[a]: Numbers correspond to bp in complementary 3' – 5' direction.

Because transformation methods into *L. majuscula* have not yet been developed, we used a reporter gene assay in *E. coli* to determine whether any of these upstream (up-) regions could function as promoters. Each region predicted to contain a promoter (up*jamA*, up*jamB*, up*jamC*, up*jamD*, up*jamG*, up*jamI*, up*jamN*, and up*jamQ*), as well as the promoter upstream of the jamaicamide TSS, was amplified with specific primers from fosmids containing different portions of the jamaicamide biosynthetic pathway (Edwards et al. 2004; Table S1). Each of these regions were individually ligated into the pBLUE TOPO vector (Invitrogen) and transformed into TOP-10 *E. coli*. The resulting constructs were evaluated for relative promoter activity using the β-galactosidase reporter gene assay (Invitrogen), standardized against total soluble protein content measured by BCA assay (Pierce). For up*jamA*, two regions were

evaluated, including the region predicted to contain the initial promoter, as well as immediately upstream of the *jamA* gene (a region with high activity in preliminary assays). The arabinose promoter from *E. coli* was amplified from the pBAD vector (Invitrogen) and ligated into the pBLUE vector as a positive control, while a 49 bp segment of a jamaicamide pathway gene (*jamG*) ligated into pBLUE vector was used as a negative control.

Several of the tested intergenic regions exhibited significantly stronger promoter activity than the positive control, including the promoter identified from the primer extension experiment (up*jamA*-902 - -832 bp), as well as up*jamB*, up*jamD*, and up*jamI* (Figure 2.4). The intergenic regions up*jamG* and up*jamN* both had some promoter activity, although lower than the positive control. The region upstream of *jamQ* did not have any detectable promoter activity in the assay, which suggested that the promoter for this transcript may be located upstream of an adjacent ORF. To more precisely localize the promoter regions upstream of two of these genes, a series of additional assays were conducted using truncated regions of upjam*A* (immediately upstream of the *jamA* gene) and up*jamI*. For up*jamA*, promoter activity was comparable relative to the positive control when testing longer upstream fragments (-500 - 0 and -200 - 0 bp; Figure 2.5).

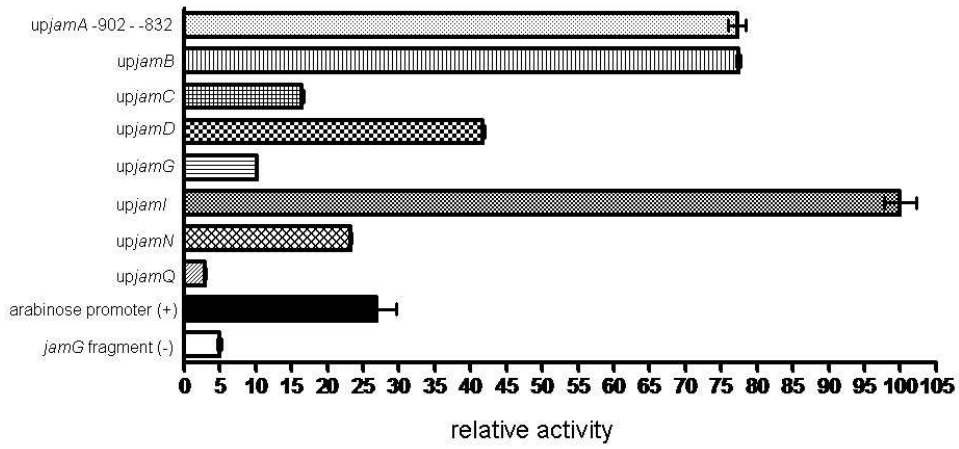


Figure 2.4: Relative activity of the primary promoter upstream of *jamA* and predicted promoters in jamaicamide intergenic regions in the β-galactosidase reporter assay**.** Standard error is represented by error bars (Jones et al. 2009).

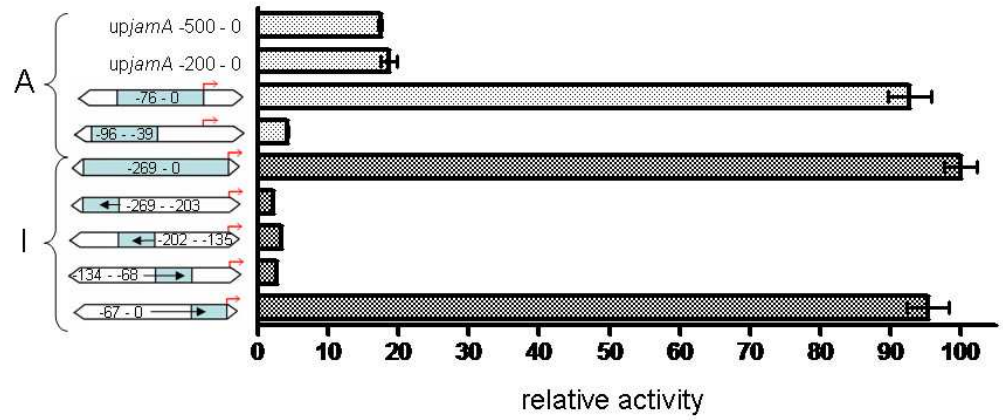


Figure 2.5: Activity of truncated up*jamA* and up*jamI* regions in the β-galactosidase assay**.** Trimmed regions are represented by blue shaded figures with associated base pair numbers. Red arrows indicate the start codon of the downstream ORF (*jamA* or *jamI*). Relative activity was calculated on same scale as Figure 4. Standard error is represented by error bars (Jones et al. 2009).

However, when small fragments closer to the *jamA* ORF start site were used, the promoter activity increased significantly, with maximal activity observed for the fragment -76 - 0 bp upstream of *jamA*. The promoter in the -76 - 0 region appeared to

require the sequence fragment -38 - 0, as another construct containing the region up*jamA*-96 - -38 did not have any promoter activity. The entire 269 bp up*jamI* upstream region also displayed strong promoter activity relative to the positive control. Promoter activity was lost using fragments encompassing -269 - -68 bp, but restored using the fragment -67 - 0 bp (Figure 2.5). Inspection of the sequences included in these active, truncated regions of up*jamA* and up*jamI* led to the identification of possible conserved promoter elements in close proximity to the ORF start sites for both genes (Table 2.1).

To quantitatively determine the promoter activities of the DNA fragments, a series of β-galactosidase assays incorporating a serial dilution of *E. coli* soluble protein lysate was also used in order to avoid saturation problems in color development (Figure 2.6). These data were used to calculate β-galactosidase activity in terms of nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$ for each of the upstream fragments with any detectable promoter activity. The strongest promoter was the section upstream of the jamaicamide TSS (-902 - -832 upstream of *jamA*), with an average of approximately 950 nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$. The promoter immediately upstream of *jamA* (-76 - 0) and those upstream of *jamB, jamD,* and *jamI* yielded lower values, with up*jamA,* up*jamB* and up*jamI* between 500-700 nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$, and up*jamD* at approximately 265 nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$. Reduced activity was found for promoters upstream of *jamC, jamG,* and *jamN,* with values ranging from approximately 75 to 150 nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$. The

arabinose promoter positive control construct yielded an average value of 170 nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$.

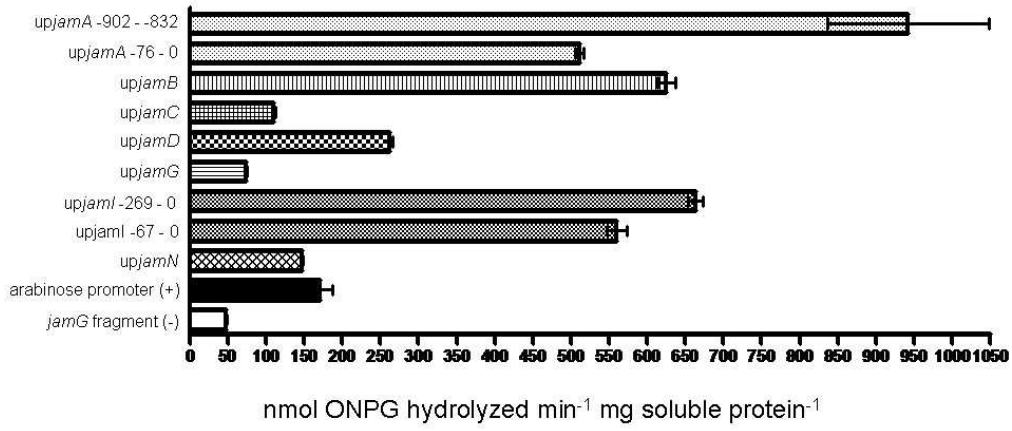


Figure 2.6: Specific activity of the strongest promoters in the β-galactosidase assay**.** Base pair number relative to gene ORF start site is provided when necessary. Standard error is represented by error bars (Jones et al. 2009).

2.4 Discussion

In this chapter, the transcription of the jamaicamide biosynthetic gene cluster in *Lyngbya majuscula* was explored. The jamaicamide cluster was chosen because it possesses a number of features commonly seen in other secondary metabolites isolated from marine cyanobacteria (Tidgewell et al. 2010). The jamaicamides are produced by the most prolific cyanobacterial natural product producer yet known (*L. majuscula*), are bioactive (ichthyotoxic, neurotoxic), are composed of mixed PKS/NRPS derived subunits, and contain unusual structural features such as a vinyl chloride and alkynyl bromide rarely seen in natural products from other organisms.

The first description of the jamaicamides (Edwards et al. 2004) demonstrated that the cluster is composed of 17 ORFs, with 16 transcribed in the same direction. The cluster is flanked on the 5' and the 3' ends by transposases and hypothetical

proteins. From the results of RT-PCR experiments, it appears that the gene cluster is preceded by an unusually long untranslated leader region (at least 844 bp), one that may be unprecedented in size for a secondary metabolite gene cluster. The function of having such a long region between the TSS and the start codon of *jamA* is unclear at this time, but may be important for overall regulation of the pathway.

In *Synechococcus* PCC 7942, the psBAII and psBAIII genes encoding the photosystem II reaction center D1 protein have *cis* regulatory elements in addition to basal promoters. Contained in the untranslated leader region downstream of the psB TSS are light responsive elements that were found to be responsible for increased expression of the genes under high light conditions (Li and Golden 1993). In the jamaicamide pathway, the fact that another region of DNA immediately upstream of *jamA* can function as a strong promoter indicates that although transcription may initiate well before the ORF start site, there could be a supplemental means of boosting transcription closer to the first protein in the cluster.

The amplification of second strand cDNA from JHB RNA corresponding to all of the intergenic regions between the jamaicamide ORFs tested indicated that the pathway is transcribed in at least two pieces. The first, *jamABCDEFGHIJKLMNOP*, is sufficiently large (~55 kb) to assume that multiple transcripts could be needed to process this portion of the gene cluster. A similar situation was found with the microcystin gene cluster (Kaebernick et al. 2002), in which all of the intergenic regions of the pathway aside from the bidirectional promoter were transcribed, and RACE experiments with several of these regions detected variations in intergenic TSS locations. As with microcystin, the jamaicamide pathway could contain internal

promoters which, while not representing true breaks in the transcription of the pathway, can function independently if not overwritten by RNAP acting from an upstream promoter (promoter occlusion; Gonzalez-y-Merchand 1998). Indeed, several of these regions were able to function as promoters in a reporter assay (see below).

A second transcript in the direction complementary to the large transcript in the jamaicamide pathway is probably needed to include *jamQ*, a gene encoding a condensation like protein that may be involved with the creation of the pyrrolinone ring of the molecule. According to RT-PCR experiments, the regions between *jamQ* and the three genes closest upstream (*ORF5* and *ORF6*, both transposases, and *ORF7*, a hypothetical protein), are all transcribed. In addition, the upstream region of *jamQ* does not appear to serve as a strong promoter in β-galactosidase reporter assays (see below), despite the presence of possible conserved promoter domains (Table 2.1). From these data, it appears that *jamQ* could be part of a larger transcript including these transposases. A larger intergenic region (approximately 1070 bp) lies upstream of *ORF7*, which could contain the TSS and a promoter for this transcript. The reason for including at least one transposase in the *jamQ* transcript is unclear, but this may be a way of ensuring transposable elements have remained associated with the cluster so as to facilitate horizontal gene transfer and pathway evolution. The hectochlorin biosynthetic gene cluster from *L. majuscula* JHB (Ramaswamy et al. 2007) contains a transposase (encoded by *hctC*) located between two of the initial genes (*hctB and hctD*) in the pathway, which is also thought to contribute to the plasticity of the cluster.

Biosynthetic investigations using *Lyngbya majuscula* strains have been highly successful in identifying secondary metabolite gene clusters, in part because *L. majuscula* readily incorporates isotopically labeled precursors in feeding studies (Chang et al. 2004; Edwards et al. 2004). However, further experimentation by way of gene knockout or overexpression in *L. majuscula* is not yet possible because a viable means of genetic transformation has not been developed. Due to this limitation, we used genetic constructs in *E. coli* to determine whether the promoters identified in this study, including the primary pathway promoter upstream of the TSS and those predicted in intergenic regions, were functional. Although some differences exist in the structure of RNAP between the two bacteria (Xie et al. 1989), promoter structures in cyanobacteria are often compared to consensus sequences in *E. coli* (Kaebernick et al. 2002; Shibato et al. 2002). Furthermore, a strong *E. coli* promoter has been shown to function in the cyanobacterium *Synechococcus* (Li and Golden 1993) and the psb2 promoter from *Microcystis* can be used in *E. coli* to drive β-galactosidase production (Shibato et al. 1998). The reporter assay proved effective in verifying the promoter identified upstream of the jamaicamide pathway TSS, as well as several internal promoters located at various regions throughout the gene cluster (Figures 2.4, 2.5 and 2.6). Previous studies with β-galactosidase reporter assays have been useful in demonstrating how enhancer elements within untranslated leader regions and intergenic promoters are important in driving transcription (Li and Golden 1993; Nakano et al. 1996). The strongest promoter in the assay was that identified upstream of the *jamA* TSS, but several other promoters were either equal to or greater in strength than the positive control in the assay. One of the regions predicted to contain

a strong promoter (up*jamI*) is located in front of a large set of ORFs. The ORF *jamI*, encoding an enoyl-CoA hydratase/isomerase, forms a di-domain with *jamJ*, which encodes for an enoyl reductase and a large PKS (Edwards et al. 2004). In addition, the subsequent ORFs in the pathway (*jamK* - *M*) are separated by small intergenic regions and do not appear to contain promoters. If *jamI* - *M* form one contiguous transcript (~30 kb), a promoter in front of *jamI* could be needed for efficient transcription. The identification of functional promoters in several other intergenic regions suggests that they could also be used to boost transcription beyond the capacity of the initial promoter located before the TSS upstream of *jamA*.

One intriguing finding from using truncated intergenic regions in the β-galactosidase assay was the detection of strong activity immediately upstream of *jamA* (-76 - 0) and *jamI* (-67 - 0) (Figure 2.5). An additional promoter was predicted in a region of up*jamI* (-269 - -203) farther upstream in the 5' direction (Table 2.1), but this region was not active when used in truncated form (Figure 2.5). If these active regions upstream of *jamA* (-76 - 0) and *jamI* (-67 - 0) are able to act as internal promoters to supplement overall transcription of the jamaicamide pathway, their close proximity to *jamA* and *jamI* may compromise the ability of transcripts initiating at these positions to subsequently allow for proper translation of the JamA and JamI proteins (although transcription could take place normally downstream of each location). This could occur as a result of insufficient room for a ribosome binding site, although translation of mRNA in cyanobacteria may not require the use of Shine-Dalgarno sequences (Mutsuda and Sugiura 2006) and some evidence exists for translation of leaderless mRNA in bacteria (Moll et al. 2002). It is possible that

heterologous use of these up*jamA* (-76 - 0) and up*jamI* (-67 - 0) regions in *E. coli* could lead to false positive identification of promoters in some instances. However, as previously discussed, the organization of the gene cluster supports the utility of functional promoters in both locations. The untranslated leader region of *jamA* is long enough for the presence of additional regulatory elements, and up*jamI* is a probable location for a promoter because of the long *jamI - M* transcript. Further evaluation of these two possible promoters will be necessary to determine how transcription from their locations could affect subsequent protein translation.

The work presented in this chapter represents the first transcriptional analysis of a natural product gene cluster from a filamentous marine cyanobacterium. Chapter 3 of this dissertation includes experiments focused on transcriptional regulation of the jamaicamide pathway and development of a mass spectrometry based method to measure *in vivo* turnover of the jamaicamides and other natural products from *Lyngbya* strains.

2.5 Materials and Methods

2.5.1 Bacterial strains, culture conditions, PCR reactions, and DNA measurements

Lyngbya majuscula JHB was originally collected from Hector's Bay, Jamaica (Edwards et al. 2004) and was maintained in a culture facility at Scripps Institution of Oceanography. Cultures were grown in BG-11 saltwater media at 29°C under a light intensity of approximately 5 μE m^{-2} s^{-1} and under 16 h light/8 h dark cycles. *E. coli* TOP-10 and BL-21 (DE3) were grown in Luria-Bertani (LB) media. *E.*

coli cultures were grown with ampicillin (100 μg ml^{-1}), or kanamycin (50 μg ml^{-1}) when necessary. PCR reactions were conducted using either PCR Master Mix (Promega) or Pfx50 proofreading Taq Polymerase (Invitrogen). DNA concentrations were measured using either Beckman-Coulter DU800 or NanoDrop 1000 (Thermo Scientific) spectrophotometers. Protein concentrations for recombinant JHB proteins were determined using the BCA assay (Pierce). Ladders for DNA (Fermentas and New England Biolabs) and protein (Bio-Rad) were used for size estimations when necessary.

2.5.2 RT-PCR using *L. majuscula* RNA to search for the transcription start site (TSS) and promoter regions in the jamaicamide pathway

Cyanobacterial filaments (approximately 2 g wet weight) from a culture of the jamaicamide producing strain of *L. majuscula* JHB were harvested and subjected to RNA isolation using TRIzol reagent (Invitrogen) and procedures based on those recommended by the manufacturer with minor modifications. RNA was treated with TURBO DNAse (Ambion) for 2 h at 38°C before use in cDNA reactions. To verify that genomic DNA contamination was not present, in selected cases negative control reactions were run in parallel with cDNA reactions in which reverse transcriptase enzyme was omitted. For the primer extension experiment, first strand cDNA was synthesized from the RNA using the primer up*jamA* 20-0 R (Sigma Genosys; Table S1) and the Superscript III Reverse Transcriptase Protocol (Invitrogen) with minor modifications. Second strand reactions were conducted with primers ranging from 500-902 bp upstream in 50 bp increments to determine where RNA transcription

upstream of *jamA* initiated. For cDNA synthesis of jamaicamide intergenic regions, first strand cDNA was generated using either random (Invitrogen and Frias-Lopez et al. 2004; Table S1) or specific jamaicamide upstream intergenic region reverse primers (Table S1). Forward and reverse oligonucleotide primers for each upstream intergenic region of the cluster were used to PCR amplify regions from the first strand cDNA to create second strand cDNA. In some instances, sequencing was used to confirm the correct amplification of second strand cDNA, either by direct sequencing of PCR products or by sequencing of TOPO TA cloning vectors (Invitrogen) containing the intergenic region. All cDNA PCR products were visualized on agarose gels.

2.5.3 Use of promoter prediction and β-galactosidase reporter gene assays to search for promoter activity

Each intergenic region upstream of the genes in the jamaicamide pathway was examined for conserved binding regions (in comparison to the σ^{70} *E. coli* consensus promoter) using the BPROM predictor (http://www.softberry.com; Table 2.1). The upstream (up-) regions of genes predicted to contain a promoter (*jamA, jamB, jamC, jamD, jamG, jamI, jamN, and jamQ*) were amplified with specific primers (Table S1) from fosmids produced previously (Edwards et al. 2004). Each upstream section was individually cloned into the pBLUE TOPO vector (Invitrogen) and transformed into TOP-10 *E. coli*. Plasmid purification (Qiagen) and sequencing (Seqxcel, Inc., La Jolla, CA) were used to confirm the sequence and direction of the inserts. Verified clones were used to measure relative promoter activity using the β-galactosidase reporter

gene assay (Invitrogen), standardized against total soluble protein content as measured by BCA assay (Pierce). For this protocol, colonies from freshly streaked selective (100 μg ml^{-1} ampicillin) plates were grown overnight in LB media (5 ml). Cultures were pelleted at 4000 RPM, and the pellets were washed once with 3 ml of chilled PBS. After an additional centrifugation step, the pellets were resuspended in 1 ml PBS. The tube contents were centrifuged at 14,000 RPM for 5 min, and subsequently resuspended in 110 μl lysis buffer (0.25 mM Tris, pH 8.0). The cells were lysed by freeze-thawing the suspensions 4 times with dry ice and a 37°C water bath, followed by another centrifugation period. The supernatant was removed from each tube for use in the β-galactosidase and BCA assays in a 96 well plate. O-nitrophenol measurements from ONPG cleavage in the assay were taken at 420 nm, while BCA readings were taken at 570 nm (Thermo-Electron Multiskan Ascent plate reader). Serial dilutions of lysis buffer from each culture were made to find an optimal range for colorimetric detection of o-nitrophenol and for a comparison of relative activity (identified as a 3-fold dilution). Relative activity calculations were made by determining the average nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$ value for each vector insert and dividing each value by the highest overall average value to obtain a percentage. A fragment of the pBAD TOPO vector (Invitrogen) containing the arabinose promoter and vector ribosome binding site (upstream of the TOPO cloning site) from *E. coli* was found to consistently produce β-galactosidase in the pBLUE TOPO vector in preliminary experiments, and was used as a positive control. Because the arabinose operator was not included in the positive control, the addition of arabinose was not required to produce β-galactosidase. A 49 bp segment of the jamaicamide *jamG* gene

was used as a negative control. [Note: the pBLUE vector contains a cryptic promoter that is reported to possibly limit the efficacy of assaying other promoter fragments in a prokaryotic host (Invitrogen). However, a series of preliminary assays indicated significant and repeatable differences in promoter activity between possible promoter regions, and baseline activity in the negative control was sufficiently low as to not conflict with the assay results. The BPROM prediction software was used to verify that the vector constructs did not introduce any artificial promoters]. Those regions found to have promoter activity were assayed again with additional dilution (10 fold) to quantify promoter strength, expressed as specific activity (nmol ONPG hydrolyzed min^{-1} mg soluble $protein^{-1}$).

2.5 Acknowledgements

The chapter author thanks Dr. Carla S. Jones for assistance with RNA extraction procedures, Dr. Sheila Podell for assistance with bioinformatics, and R. Cameron Coates and Tara Byrum for maintenance of JHB cultures. This work was supported by grants from the NIH (GM075832 and NS053398), and NOAA Grant NA08OAR4170669, California SeaGrant College Program Project SG-100-TECH-N, through NOAA's National Sea Grant College Program, US Department of Commerce. Chapter 2 of this dissertation includes a portion of a manuscript: Jones, Adam C.; Gerwick, Lena; Gonzalez, David; Dorrestein, Pieter C.; Gerwick, William H. “Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory

proteins." *BMC Microbiology* 2009, 9:247. The dissertation author is the primary investigator and author of this paper.

2.6. References

Aoyama T, Takanami M, Ohtsuka E, Taniyama Y, Marumoto R, Sato H, Ikehara M (1983) Essential structure of *E. coli* promoter: Effect of spacer length between the two consensus sequences on promoter function. *Nuc Acids Res* 11:5855-5864.

Bibb MJ (2005) Regulation of secondary metabolism in Streptomycetes. *Curr Op Microbiol* 8:208-215.

Blunt JW, Copp BR, Hu W-P, Munro MHG, Northcote PT, Prinsep MR (2009) Marine natural products. *Nat Prod Rep* 26:170-244.

Chang Z, Flatt P, Gerwick WH, Nguyen V-A, Willis CL, Sherman DH (2002) The barbamide biosynthetic gene cluster: A novel cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296:235-247.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 67:1356-1367.

Cragg GM, Newman, DJ, Snader KM (1999) Natural products in drug discovery and development: The United States National Cancer Institute Role. In *Phytochemicals in Human Health Protection, Nutrition, and Plant Defense*. Romeo JT, Ed. New York: Kluwer Academic/Plenum Publishers; 1-29.

de Figueiredo DR, Azeiteiro UM, Esteves SM, Goncalves F, Pereira MJ (2004) Microcystin-producing blooms-a serious global public health issue. *Ecotoxicol Environ Safety* 59:151-163.

Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* 11:817-833.

Frias-Lopez J, Bonheyo GT, Fouke BW (2004) Identification of differential gene expression in bacteria associated with coral black band disease by using RNA-arbitrarily primed PCR. *Appl Environ Microbiol* 70:3687-3694.

Galm U, Shen B (2006) Expression of biosynthetic gene clusters in heterologous hosts for natural product production and combinatorial biosynthesis. *Exp Op Drug Discov* 1:409-437.

Gonzalez-y-Merchand JA, Colston MJ, Cox RA (1998) Roles of multiple promoters in transcription of ribosomal DNA: Effects of growth conditions on precursor rRNA synthesis in mycobacteria. *J Bacteriol* 180:5756-5761.

Gu L, Geders TW, Wang B, Gerwick WH, Håkansson K, Smith JL, Sherman DH (2007): GNAT-like strategy for polyketide chain initiation. *Science* 318:970-974.

Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nuc Acids Res* 11:2237-2255.

Ishiura M, Kutsuna S, Aoki S, Iwasaki H, Andersson CR, Tanabe A, Golden SS, Johnson CH, Kondo T (1998) Expression of a gene cluster *kaiABC* as a circadian feedback process in cyanobacteria. *Science* 281:1519-1523.

Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.

Kaebernick M, Dittmann E, Börner T, Neilan BA (2002) Multiple alternate transcripts direct the biosynthesis of microcystin, a cyanobacterial nonribosomal peptide. *Appl Environ Microbiol* 68:449-455.

Kaebernick M, Neilan BA, Börner T, Dittmann E (2000) Light and the transcriptional response of the microcystin biosynthesis gene cluster. *Appl Environ Microbiol* 66:3387-3392.

Li R, Golden SS (1993) Enhancer activity of light-responsive regulatory elements in the untranslated leader regions of cyanobacterial *psbA* genes. *Proc Natl Acad Sci USA* 90:11678-11682.

Lindell D, Padan E, Post AF (1998): Regulation of ntcA expression and nitrite uptake in the marine *Synechococcus* sp. strain WH 7803. *J Bacteriol* 180:1878-1886.

Lisser S, Margalit H (1993) Compilation of *E. coli* mRNA promoter sequences. *Nuc Acid Res* 21:1507-1516.

Martin-Luna B, Sevilla E, Hernandez JA, Bes MT, Fillat MF, Peleato ML (2006) Fur from *Microcystis aeruginosa* binds *in vitro* promoter regions of the microcystin biosynthesis gene cluster. *Phytochem* 67:876-881.

Moll I, Grill S, Gualerzi CO, Blasi U (2002) Leaderless mRNAs in bacteria: Surprises in ribosomal recruitment and translational control. *Mol Microbiol* 43:239-246.

Mutsuda M, Sugiura M (2006) Translation initiation of cyanobacterial rbcS mRNAs requires the 38-kDa ribosomal protein S1 but not the Shine-Dalgarno sequence: Development of a cyanobacterial *in vitro* translation system. *J Biol Chem* 281:38314-38321.

Nakano MM, Zuber P, Glaser P, Danchin A, Hulett FM (1996) Two-component regulatory proteins ResD-ResE are required for transcriptional activation of *fnr* upon oxygen limitation in *Bacillus subtilis*. *J Bacteriol* 178:3796-3802.

Pfeifer BA, Wang CCC, Walsh CT, Khosla C (2003) Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host. *Appl Environ Microbiol* 69:6698-6702.

Ramaswamy AV, Sorrels CM, Gerwick WH (2007) Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products* 70:1977-1986.

Rossi JV, Roberts MA, Yoo H-D, Gerwick WH (1997) Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically-useful natural metabolite curacin A. *J Appl Phycol* 9:195-204.

Saito K, Sei Y, Miki S, Yamaguchi K (2008) Detection of microcystin-metal complexes by using cryospray ionization-Fourier transform ion cyclotron resonance mass spectrometry. *Toxicon* 51:1496-1498.

Schatz D, Keren Y, Vardi A, Sukenik A, Carmeli S, Börner T, Dittmann E, Kaplan A (2007) Towards clarification of the biological role of microcystins, a family of cyanobacterial toxins. *Environ Microbiol* 9:965-970.

Schmidt EW, Nelson JT, Rasko DA, Sudek S, Eisen JA, Haygood MG, Ravel J (2005) Patellamide A and C biosynthesis by a microcin-like pathway in *Prochloron didemni*, the cyanobacterial symbiont of *Lissoclinum patella*. *Proc Natl Acad Sci USA* 102:7315-7320.

Sevilla E, Martin-Luna B, Vela L, Bes MT, Fillat MF, Peleato ML (2008) Iron availability affects mcyD expression and microcystin-LR synthesis in *Microcystis aeruginosa* PCC7806. *Env Microbiol* 10:2476-2483.

Shibato J, Agrawal GK, Kato H, Asayama M, Shirai M (2002) The 5'-upstream cis-acting sequences of a cyanobacterial psbA gene: Analysis of their roles in basal, light-dependent and circadian transcription. *Mol Genet Genom* 267:684-694.

Shibato J, Asayama M, Shirai M (1998) Specific recognition of the cyanobacterial *psbA* promoter by RNA polymerases containing principal sigma factors. *Biochim Biophys Acta* 1442:296-303.

Suyama TL, Gerwick WH (2008): Stereospecific total synthesis of somocystinamide A. *Org Lett* 10:4449-4452.

Tan LT (2007) Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochem* 68:954-979.

Tidgewell K, Clark BR, Gerwick WH (2010) The Natural Products Chemistry of Cyanobacteria. In *Comprehensive Natural Products II Chemistry and Biology*; Mander L, Lui H-W, Eds. Elsevier: Oxford: 2:141-188.

Tillett D, Dittmann E, Erhard M, von Döhren H, Börner T, Neilan BA (2000): Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: An integrated peptide-polyketide synthetase system. *Chem Biol* 7:753-764.

Watanabe K, Hotta K, Praseuth AP, Koketsu K, Migita A, Boddy CN, Wang CCC, Oguri H, Oikawa H (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat Chem Biol* 2:423-428.

Wilkinson B, Micklefield J: Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* 2007, 3:379-386.

Xie WQ, Jager K, Potts M (1989) Cyanobacterial RNA polymerase genes rpoC1 and rpoC2 correspond to rpoC of *Escherichia coli*. *J Bacteriol* 171:1967-1973.

Yoon H-S, Golden JW (2001) PatS and products of nitrogen fixation control heterocyst pattern. *J Bacteriol* 183:2605-2613.

2.7. Supplemental information

Table 2.S1: Primers used in RT-PCR experiments.

Primers used for RT-PCR experiments		
Name	**Gene cluster position; accession # AY522504**	**Primer sequence**
up*jamA* F 902	5728-5752	AAGCAGAATTTATTTACGATTAGGG
up*jamA* F 500	6130-6151	GCGCGAGCTAGCTAGTCTATCC
up*jamA* 550 F	6080-6094	CAAAGGCGTACGCCT
up*jamA* 600 F	6030-6050	AGGGGAAAAAGAAACTGATAT
up*jamA* 650 F	5980-5997	TTGAGAGCCCATTCTCAG
up*jamA* 700 F	5930-5947	GATGTTCAGACCTTGACG
up*jamA* 750 F	5880-5898	AGCCTTGAAAAACGCTATT
up*jamA* 800 F	5830-5857	ACACCTAAGTATAGTTAAAACAGTAAAA
up*jamA* 850 F	5780-5798	TTTCTCATGGTGCCTAATT
up*jamA* 20-0 R	6630-6649	GAAAATTCTGGCTTGCTCAT
up*jamA* R	6604-6629	ATATCAAATAAATATACATTTCCCAT
up*jamB* F	8410-8428	TGGAGGGATTTGTGGTGGG
up*jamB* R	8575-8596	TGACATCAAACTGTCTCCTTGG
up*jamC* F	9557-9581	TTGTTCAGAATAAATAAAACTTGTC
up*jamC* R	9696-9717	AATTTATTGCTTGATGGGTACT
up*jamD* F	10021-10047	TATTTGTTGAAGATTATATCAAATCCG
up*jamD* R	10439-10462	AATATTTTGTCTCACACTTTTATC
up*jamE* F	12499-12518	GTATTTGAAGCCCTAGAGTG
up*jamE* R	12593-12618	GCTATTGGTTCTTTATTCATCATTTT
up*jamF* F	17765-17788	CTCAAACAACAGGTTACAGTTTAG
up*jamF* R	17811-17835	GTTTGAGTACTTGTTCTTTGCTCAT
up*jamG* F	18026-18050	TGATTTATTTGCGGCGAAATTATAG
up*jamG* R	18222-18246	TACCAGTGATATTGATGTCATTCAT
up*jamI* F	20713-20728	CTCTAGTACCGCTGCG
up*jamI* R	20957-20981	ACTATAGCTTTATCAATGGTTTGTC
up*jamL* F	36729-36752	GAAAAAATCAAGAGTATGCTTTGA
up*jamL* R	36820-36844	GATCTTTGTTGGTAGTCGGTTCCAT
up*jamM* F	48601-48627	CGTTCAAAAAATAAAATTAGAGAATAA
up*jamM* R	48697-48721	AAGTTTGTTCATCTTGTCTTTCCAT
up*jamN* F	51672-51696	AGAGAATAGTTCTGAAAATGGTTGA
up*jamN* R	51811-51836	GCAAGACGTTCTTGAATATCACTCAT
up*jamP* F	56726-56750	ACAAAAGAATCGTAGAAATAAATAA
up*jamP* R	56771-56795	CATCAATACTATTTGTGGTAATCAT
P-Q F	62173-62197	ATTTTACCTTGAACAGCATAATTAA
P-Q R	62318-62342	TGATGCTAACAAGGGAAAAATATAA
up*jamQ* F	64031-64049	TTGAAATTTTCGGCTCTTC
up*jamQ* R	63647-63666	ATCGGTGTTTTGTTACGTAC
up*Orf5* F	64874-64893	AAAGACGCACGGAAGACATC
up*Orf5* R	64767-64782	AAGCTGCACGGCCTGT
up*Orf6* F	66154-66174	GGATAGATAGATTTGCGTTTG
up*Orf6* R	65518-65541	TATTCGGATCTTTCTTAATATTTG

Table 2.S1(continued): Primers used in RT-PCR experiments.

Invitrogen	catalog #48190-011	random primers (mostly hexamers)
Lnr81	Frias-Lopez et al. 2004	TGAGCGGACA
Lnr95	Frias-Lopez et al. 2004	CAGCCCAGAG
Lnr99	Frias-Lopez et al. 2004	TCGTGCGGGT

Table 2.S2: Primers used in β-galactosidase reporter assays.

pBLUE Insert	**Gene cluster position; accession # AY522504**	**Description**	**F primer name**	**Forward primer (5')**	**R primer name**	**Reverse primer (3')**
Arabinose promoter	Invitrogen pBAD vector, catalog #K4300-01, 181-386 bp	upstream region of pBAD cloning site	ARAC F	ATTATTTGCACGGCGTCACAC	ARAC R	AAGGGCGAGCTTGTCATCGTC
jamG	19127-19175	906 bp into *jam*G - 954 bp into *jam*G	*jamG* F	TACTCATGGTACCAGTTCCC	*jamG* R	AGCAGATACTTCCGTATTG
up*jamA* 902-832	5728-5798	902-832 upstream of jamA	up*jamA* 902 F	AAGCAGAATTTATTTACGATTAGGG	up*jamA* 832 R	AATTAGGCACCATGAGAAA
up*jamA* 500-0	6130-6629	500 bp upstream - 0 bp upstream	up*jamA* 500 F	GCGCGAGCTAGCTAGTCTATCC	up*jamA* R	ATATCAAATAAATATACATTTCCCAT
up*jamA* 200-0	6430-6629	200 bp upstream - 0 bp upstream	up*jamA* 200 F	AGCAGTAGCCATGCCTAACTCC	up*jamA* R	ATATCAAATAAATATACATTTCCCAT
up*jamA* 76-0	6554-6629	76 bp upstream - 0 bp upstream	up*jamA* 76	AGCAAACAATTGTAGTAGGTTGGTTGC	up*jamA* R	ATATCAAATAAATATACATTTCCCAT
up*jamA* 96-39	6534-6591	96 bp upstream - 39 bp upstream	up*jamA* 96_39 F	GCAAGCGTTCGGACATAGCC	up*jamA* 96-39 R	GTGGAAAGTCAGCAACCAAC
up*jamB*	8410-8596		up*jamB* F	TGGAGGGATTTGTGGTGGG	up*jamB* R	TGACATCAAACTGTCTCCTTGG
up*jamC*	9557-9717		up*jamC* F	TTGTTCAGAATAAATAAAACTTGTC	up*jamC* R	AATTTATTGCTTGATGGGTACT
up*jamD*	10021-10462		up*jamD* F	TATTTGTTGAAGATTATATCAAATCCG	up*jamD* R	AATATTTTGTCTCACACTTTTATC
up*jamG*	18051-18221		up*jamG* F bgal	CATTTGTTTGTCATTTGTCATTT	up*jamG* R bgal	ACCAGTTTTCCCTGGTGTT

Table 2.S2 (continued): Primers used in β-galactosidase reporter assays.

pBLUE Insert	**Gene cluster position; accession # AY522504**	**Description**	**F primer name**	**Forward primer (5')**	**R primer name**	**Reverse primer (3')**
up*jamI* - 269 - 0	20713-20981	269 bp upstream - 0 bp upstream	up*jamI* F bgal	CTCTAGTAC CGCTGCGCG GA	up*jamI* R	ACTATAG CTTTATCA ATGGTTTG TC
up*jamI* 1 (-269- - 203)	20713-20779	269 bp upstream - 203 bp upstream	up*jamI* F bgal	CTCTAGTAC CGCTGCGCG GA	up*jamI* 222-203 R	ATCTTTAC TCCATAAT CCAT
up*jamI* 2 (-202 - -135)	20780-20847	202 bp upstream - 135 bp upstream	up*jamI* 202-183 F	TCTATCGAC TCAAACTGG TG	up*jamI* 154-135 R	TGGTTGCT CACCCAG ATCAA
up*jamI* 3 (-134- -68)	20848-20914	134 bp upstream - 68 bp upstream	up*jamI* 134-115 F	TGAAGACTG TATAAAAGA GA	up*jamI* 87-68 R	GCTTTTAG CTATCAGC TATC
up*jamI* 4 (-67 - 0)	20901-20981	67 bp upstream - 0 bp upstream	up*jamI* 67-48 F	TGATAGCTA AAAGCTGAT AG	up*jamI* R	ACTATAG CTTTATCA ATGGTTTG TC
up*jamN*	51697-51810		up*jamN* F bgal	GATAGAAA AAATAATAC CCCTATGC	up*jam N* R bgal	GTCAATTC CTCCTGAT TTATTTAT T
up*jamQ*	63647-64049		up*jamQ* F	TTGAAATTT TCGGCTCTT C	up*jam Q* R	ATCGGTGT TTTGTTAC GTAC

CHAPTER 3

Investigations into jamaicamide regulation and the temporal dynamics of jamaicamide biosynthesis in *Lyngbya majuscula* JHB

3.1. Abstract

Understanding regulatory factors controlling natural product expression in marine cyanobacteria is one of the most important ways by which to improve yields from laboratory cultures as well as potentially remediate the effects of harmful bloom events in the environment. Research in this chapter builds on the transcriptional analysis of the jamaicamide gene cluster presented in chapter 2. Protein pulldown assays were used to isolate proteins that may regulate the jamaicamide pathway. Pulldown experiments using the intergenic region upstream of *jamA* as a DNA probe isolated two proteins that were identified by LC-MS/MS. By BLAST analysis, one of these had close sequence identity to a regulatory protein in another cyanobacterial species. Protein comparisons suggest a possible correlation between secondary metabolism regulation and light dependent complementary chromatic adaptation. Electromobility shift assays were used to evaluate binding of the recombinant proteins to the jamaicamide promoter region. Additional experiments were performed in the development of a novel mass spectrometry based approach using stable isotope feeding with *Lyngbya* strains to monitor isotope incorporation and thus measure biosynthetic turnover. This technique afforded the production of the jamaicamides to be more carefully studied, including an assessment of how jamaicamide turnover compares with filament growth rate and primary metabolism, how production is affected under different growth conditions, and provided new insights into the

biosynthetic timing of jamaicamide A bromination, including support for the hypothesis that jamaicamide production is affected by light. This approach should also be valuable in determining how environmental factors affect secondary metabolite production, ultimately yielding insight into the energetic balance among growth, primary production, and secondary metabolism, and thus aid in the development of methods to improve compound yields for biomedical or biotechnological applications.

3.2 Introduction

The secondary metabolites of marine organisms are a valuable and inspirational source for a host of biomedical and technological applications. Of prokaryotic marine natural products identified and evaluated for their biomedical potential, it is estimated that approximately 40% are of cyanobacterial origin (Blunt et al. 2009). Natural products from marine cyanobacteria have a wide range of bioactivities (Tan 2007), and in the last decade, considerable advances have been made in elucidating the enzymatic mechanisms used in compound biosynthesis (Gu et al. 2009, Jones et al. 2010). However, the natural functions these molecules serve in the producing organism are comparatively less well understood. Some cyanobacterial secondary metabolites appear to have a protective role (Nagle and Paul 2002), acting as feeding deterrents (Cruz-Rivera and Paul 2007) or toxins (Cardellina II et al. 1979), while others may be involved in microbial communication as quorum sensors (Sharif et al. 2008). As with plants or algae (Herms and Mattson 1992, Cronin and Hay 1996), the production rate of a specific cyanobacterial compound may vary based on its

effective concentration, genetic or environmental regulatory factors, or the presence of an inducing action. The energetic expense of secondary metabolite production (Herms and Mattson 1992) may also lead to variability in its biosynthesis depending on available resources or specific ecological conditions. This variability is often observed during laboratory secondary metabolite isolation efforts, where compound amounts can range from trace (submilligram) levels (Volk 2007) to instances where a single natural product is the major component of a given extract (Gerwick et al. 1994). Low compound yields, in addition to very slow growth in culture (with doubling times of 6 days or longer in some cases) (Rossi et al. 1997), are among the most significant impediments to further pursuit of marine cyanobacterial lead compounds for biotechnological applications.

This chapter includes several experiments focused on the jamaicamide gene cluster from *Lyngbya majuscula* JHB, with the aim of developing methods to improve natural product yields of *Lyngbya* natural products in laboratory culture. Protein pulldown experiments were used to isolate at least one possible jamaicamide regulatory protein that is able to bind to the region upstream of the jamaicamide transcription start site in gel shift assays. Bioinformatic analyses conducted with the protein sequence suggest a correlation between secondary metabolite production and complementary chromatic adaptation (CCA) in cyanobacteria. A mass spectrometry/stable isotope approach developed in collaboration with Eduardo Esquenazi (Gerwick Lab, SIO/UCSD; Esquenazi et al. 2011) revealed the temporal dynamics of jamaicamide biosynthesis and allowed us to test the effect of a number of

environmental parameters on jamaicamide production, ultimately confirming that jamaicamide production is affected by available light levels.

3.3 Results

3.3.1 Isolation and characterization of possible transcription factors from a pulldown assay

To determine whether jamaicamide regulatory proteins are encoded in the *L. majuscula* JHB genome, DNA - protein "pulldown" experiments were performed to isolate proteins with affinity to the upstream region of *jamA*. A biotinylated, DNA probe extending from 1000 bp upstream of *jamA* to 20 bp into the *jamA* gene (encompassing both the putative promoter region and entire untranslated leader region; primers in Table 3.1) was used to label streptavidin coated magnetic Dynabeads (Invitrogen), which were then incubated with a soluble protein lysate from *L. majuscula* JHB.

Table 3.1: Primers used for isolation and characterization of regulatory proteins (Jones et al. 2009).

Name	Primer sequence
upjamA 1000 biotin	biotin-AGGATGAAAGTACCTTAATCAATGG
upjamA 20-0 R	GAAAATTCTGGCTTGCTCAT
5335 F	TTGGTTAGGCTATCACTGGATAGAGA
5335 R	TAGCCGCTTAGCATAGCAGCGTAATC
7968 F	GTGGTTATAGGGAATACAGTTAGGT
7968 R	CTGCACCATAATGAAAAATAGCCTC
5335 FB	ACTCAATCGTGCCGAATTGGCATCA
5335 RB	TATTACAATAACCAGGCCAGCCTAG
7968 FB	GAATCTATTGCGGAAATCCTTGACT
7968 RB	CTTTGTATCCTACTTTAGGATCATA
5335 Nco1F	GGGGCCATGGTTAGGCTATCACTGGATA
5335 Not1R	CCCCGCGGCCGCTTATAGCCGCTTAGCATAGC
7968 Nde1F	GGGGCATATGGTTATAGGGAAATACAGTT
7968 Xho1R	CCCCCTCGAGTTACTGCACCATAATGAAAA
upjamA 1000 F	AGGATGAAAGTACCTTAATCAATG
upjamA 832 R	AATTAGGCACCATGAGAAA

A series of wash steps were first conducted to remove proteins non-specifically bound, followed by elution of those proteins specifically bound to the probe. This elution was visualized using SDS-PAGE and revealed at least two bands of approximately 30-45 kDa in size (Figure 3.1). The protein bands from the gel, as well as crude fractions eluted from the magnetic beads in repeated experiments, were submitted to the Mass Spectrometry Center at The Scripps Research Institute for trypsin digest and LC-MS/MS analysis.

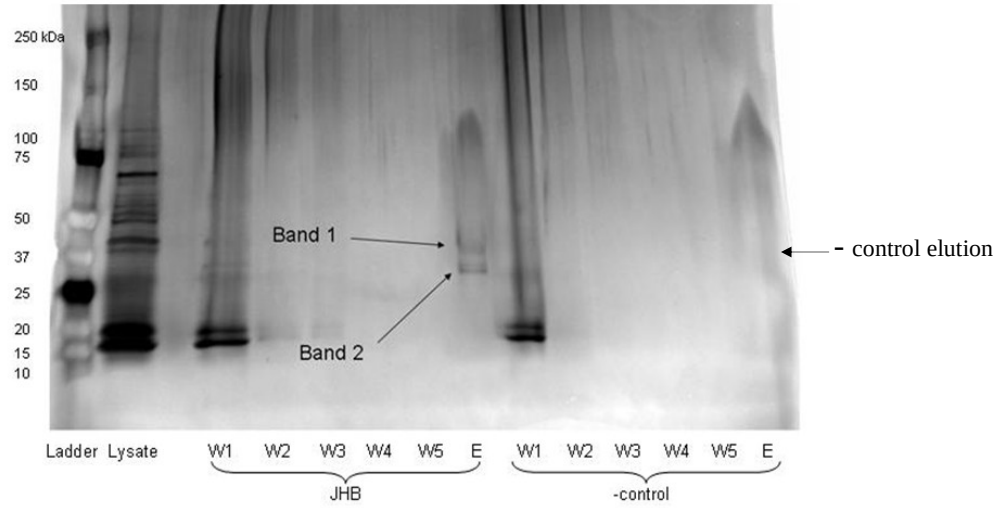


Figure 3.1: Results from JHB soluble protein pulldown experiment**.** From left to right: Ladder, JHB soluble protein lysate, wash fractions (W1 - 5) and elution (E) for incubations with the 1020 bp probe (labeled JHB) or without probe (labeled -control). Note the presence of two bands eluted from the beads containing the probe, indicating successful binding of possible regulatory proteins to the upstream region of *jamA*.

The LC-MS/MS analysis of the gel band and elution digests provided fragmented peptides that were used to query the unfinished *Lyngbya majuscula* 3L genome (a strain from Curaçao that produces several natural products, including barbamide and curacin A) using the MS/MS post-processing program InSpecT (Tanner et al. 2005). By this approach, the *L. majuscula* JHB peptide fragments were matched with to two *L. majuscula* 3L encoded proteins with high confidence from "band 2" (Figure 3.1; both proteins having global (N-terminal to C-terminal) distributions among the identified peptides): (i) All4300 protein (39.2% coverage and a molecular weight of 32 kDa), and (ii) hypothetical protein (35.9% coverage and a molecular weight of 33 kDa). Manual annotation of the most abundant peptide identified within the primary sequence of All4300 demonstrated the *b* and *y* ion series fell within a mass error of 5-400 ppm. Furthermore, the *b* and *y*-ion series

for this peptide showed 22/30 possible fragmentations covered with several contingent ion series. The ion series for the hypothetical protein showed similar results to the All4300 protein. Results from the LC-MS/MS of the PAGE gel "band 1" (Figure 3.1) were inconclusive. Separate analyses of the elution fractions identified with high confidence the same All4300 and hypothetical protein from band 2, as well as a number of putative proteins in the 3L genome such as a peptidase (~45 kDa) and an AP endonuclease (~30 kDa). Several pigment related proteins were also identified that were not visually apparent by SDS-PAGE (smaller than the two main bands indicated on Figure 3.1), including C-phycoerythrin class 1 subunit alpha (~19 kDa), allophycocyanin alpha subunit (~17 kDa), and photosystem I (PsaD) (~16 kDa). BLAST analyses of the All4300 and the hypothetical protein (referred to subsequently as protein 5335 and protein 7968, respectively, from initial annotations of the *L. majuscula* 3L genome) both yielded a number of hypothetical protein matches in other cyanobacteria including *Anabaena variabilis, Microcoleus chthonoplastes*, *Nostoc punctiforme*, and *Trichodesmium erythraeum* (JHB protein BLAST hits in Table 3.2; see below). Interestingly, both proteins also matched (although significantly better for 7968) with the protein RcaD, an activator protein from the cyanobacterium *Calothrix* (= *Fremyella diplosiphon* or *Tolypothrix*) known to regulate complementary chromatic adaptation (Sobczyk et al. 1994; Schyns et al. 1998; Noubir et al. 2002; Kehoe and Gutu 2006).

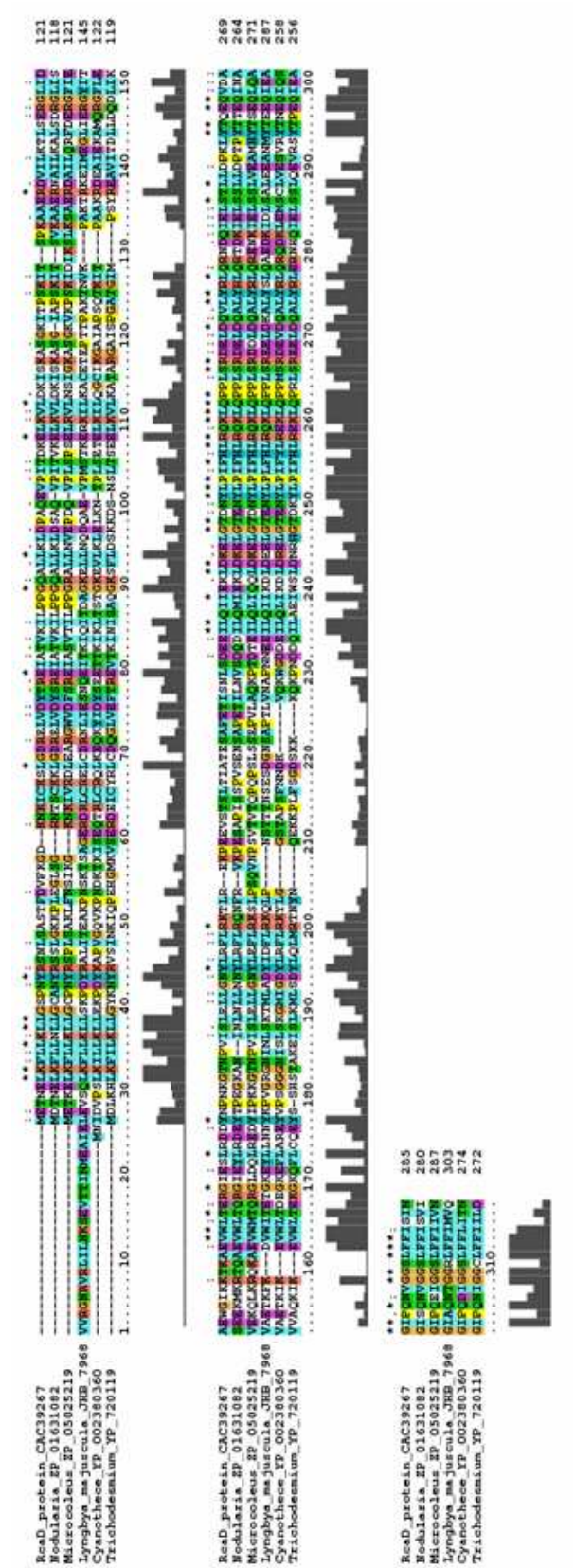
Table 3.2: BLAST results with *Lyngbya majuscula* JHB proteins 5335 and 7968 (Jones et al. 2009).

5335 (279 aa)						
Best BLAST hit	**BLAST organism**	**Size (aa)**	**identity**	**similarity**	**e value**	**accession #**
hypothetical protein	*Nostoc punctiforme* PCC 73102	217	56	70	8.00E-62	YP_001867255
hypothetical protein	*Microcoleus chthonoplastes* PCC 7420	245	56	71	2.00E-59	ZP_05025825
hypothetical protein all4300	*Nostoc* sp. PCC 7120	227	49	68	4.00E-54	NP_488340
hypothetical protein	*Anabaena variabilis* ATCC 29413	221	49	65	1.00E-51	YP_321771
hypothetical protein	*Lyngbya* sp. PCC 8106	224	47	64	9.00E-47	ZP_01623947
hypothetical protein	*Lyngbya* sp. PCC 8106	156	33	56	2.00E-11	ZP_01621638
hypothetical protein	*Nodularia spumigena* CCY9414	100	41	61	3.00E-11	ZP_01628571
hypothetical protein	*Arthrospira maxima* CS-328	131	32	60	1.00E-08	ZP_03271683
RcaD protein	***Tolypothrix* sp. PCC 7601**	**285**	**22**	**48**	**0.2**	CAC39267
7968 (304 aa)						
Best BLAST hit	**BLAST organism**	**Size (aa)**	**identity**	**similarity**	**e value**	**accession #**
hypothetical protein	*Cyanothece* sp. PCC 7424	274	49	69	2.00E-68	YP_002380360
RcaD protein	***Tolypothrix* sp. PCC 7601**	**285**	**43**	**63**	**3.00E-54**	CAC39267
hypothetical protein	*Trichodesmium erythraeum* IMS101	272	40	59	3.00E-52	YP_720119
hypothetical protein	*Nodularia spumigena* CCY9414	280	44	62	1.00E-50	ZP_01631082
hypothetical protein	*Microcoleus chthonoplastes* PCC 7420	287	41	62	7.00E-50	ZP_05025219
hypothetical protein	*Synechococcus* sp. PCC 7335	199	33	57	3.00E-22	ZP_05035072

Complementary chromatic adaptation (CCA) is a phenomenon exhibited by many cyanobacteria in response to changes in light wavelength and intensity. CCA allows cyanobacteria to alter pigment levels so as to optimize their capacity for photosynthesis, and usually involves variation between green and red phenotypes (Li et al. 2008). RcaD is a protein that binds to the promoter for phycocyanin 2 (*cpc2*) and alters the expression of several red light operons in the acclimation phase of CCA (Noubir et al. 2002; Kehoe and Gutu 2006). Another protein, RcaG, is located downstream of RcaD and has been identified as a putative ATPase. RcaG may facilitate binding of RcaD to DNA, and could require phosphorylation to complete this task (Noubir et al. 2002). Bioinformatic analysis of the *L. majuscula* 3L genome revealed that the proteins immediately downstream of 5335 and 7968 both resulted in BLAST hits with RcaG, although as with RcaD, the protein neighboring 7968 (7969) had much stronger sequence identity than the neighboring protein to 5335 (5336).

Primers were designed from each of the gene sequences for the two proteins identified above using the *L. majuscula* 3L unfinished genome, and were successful in amplifying homologous gene sequences from *L. majuscula* JHB genomic DNA. The JHB homolog to 5335 encodes for a protein that differs from the 3L protein by only one amino acid (99.6% identical), while the 7968 homolog in JHB encodes for a protein 89.5% identical to the 7968 protein in 3L. Alignments of each JHB protein with their nearest respective BLAST hits (alignment of protein 7968 shown in Figure 3.2.) indicated several conserved sequence regions, with the highest level of conservation found toward the C terminal end of the proteins (a region in the RcaD protein thought to be involved in DNA binding) (Noubir et al. 2002).

Figure 3.2: Sequence alignment with *Lyngbya majuscula* JHB protein 7968 and 5 proteins with highest identity matches from NCBI BLAST analyses, performed in ClustalX2 (Jones et al. 2009).



3.3.2 Recombinant expression of identified proteins and Electromobility Shift Assays (EMSAs)

The sequences encoding the 5335 and 7968 proteins in JHB were used in creating constructs for recombinant expression in *E. coli* (Figure 3.3).

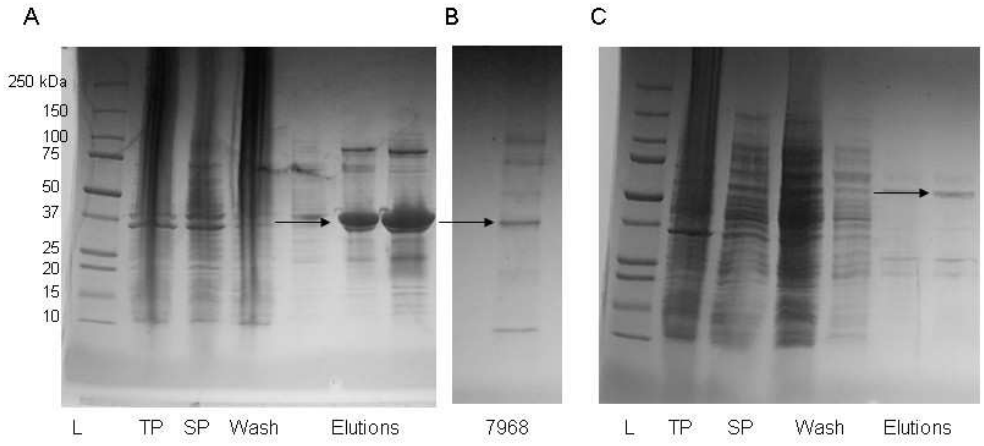


Figure 3.3: Recombinant expression of JHB proteins**.** A: Protein expression from *L. majuscula* JHB 7968 (His+protein: ~37 kDa). Arrow indicates eluted protein. B: Protein 7968 after thrombin His tag cleavage and concentration. Arrow indicates cleaved protein. C: Protein expression from *L. majuscula* JHB 5335 GST fusion vector (GST+protein: ~60 kDa). Arrow indicates eluted GST+5335 protein (Jones et al. 2009).

After expression and purification of each protein, both were used in **E**lectro**m**obility **S**hift **A**ssays (EMSAs). In these assays, protein and a fragment of DNA amplified from a region that included both the sequence of the primary jamaicamide promoter and the region upstream from the original probe (1000 - 832 bp upstream of *jamA*) were incubated and visualized on native PAGE gels. Recombinant 7968 was found to bind this putative transcription factor binding region upstream of *jamA* after His tag removal with thrombin cleavage (Figure 3.4a), although

promiscuous binding was also observed with other control DNA fragments (data not shown).

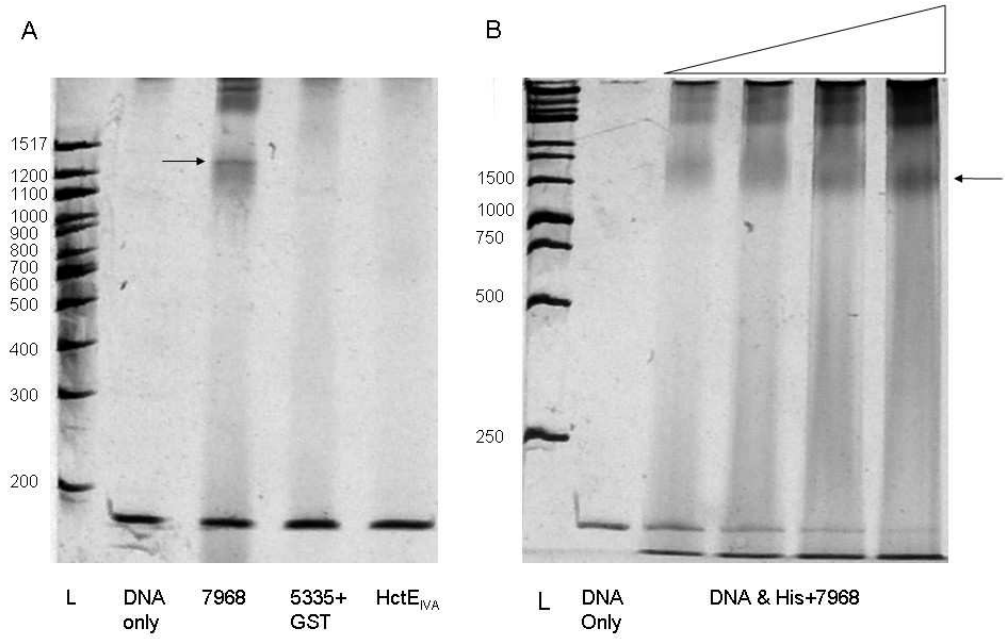


Figure 3.4: Electromobility shift assays**.** A) EMSA gel shift assay with DNA region -1000 - -832 bp upstream of *jamA*. DNA [270 fmol (= 30 ng)] was assayed with (from left to right) no protein, 7.3 pmol of 7968, 8.4 pmol of GST+5335, or 31 pmol of $HctE_{IVA}$. Arrow indicates DNA + protein shift for 7968. B) Serial titration experiment with 45 fmol (= 5 ng) of the same DNA region with (from left to right) no protein, 6.8 pmol, 13.7 pmol, 27.3 pmol, or 54.8 pmol His+7968. Arrow indicates DNA + protein shift (Jones et al. 2009).

A serial titration of 7968 with the N-terminal His tag still attached showed increased DNA binding with larger amounts of protein (Figure 3.4b). Recombinant protein 5335 was expressed and purified with a GST-tag on the N-terminus of the protein. However, attempts to remove the GST tag were unsuccessful, and thus we assayed protein 5335 with the GST tag still attached (Figure 3.3c). This version of 5335 did not bind to the *upjamA*-1000 - -832 bp region (Figure 3.4a), even with elevated protein concentrations (Figure 3.5).

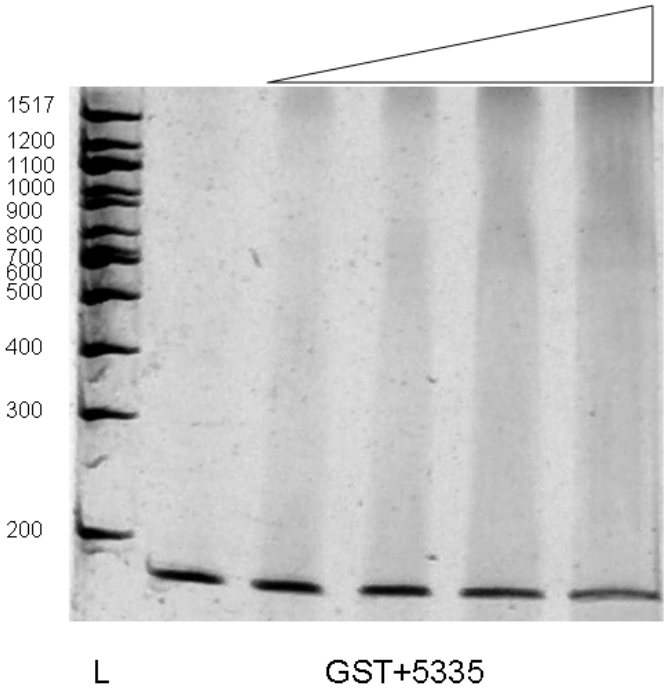


Figure 3.5: EMSA with DNA region -1000 - -832 bp upstream of *jamA* and protein GST+5335. From left to right: 270 fmol DNA only, 8.4 pmol, 16.4 pmol, 33.5 pmol, and 67.0 pmol of GST+5335 combined with 270 fmol DNA (Jones et al. 2009).

3.3.3 Temporal dynamics of jamaicamide biosynthesis in *Lyngbya majuscula* JHB

To further test the possible relationship between light and jamaicamide production, additional experiments were conducted as part of a larger effort to design a method for monitoring the *in vivo* turnover of natural products in *Lyngbya* strains (Esquenazi et al. 2011). In growing *Lyngbya* cultures in ^{15}N-enriched growth media and probing over time using micro-extraction and Matrix Assisted Laser Desorption Ionization (MALDI) mass spectrometry, we found that the dynamics of cyanobacterial natural product biosynthesis can be observed, and conditions that may influence the turnover of these molecules can be tested. By comparing and monitoring the changes

in the isotopic profiles of metabolites arising from [^{15}N] $NaNO_3$-enriched cultures of *Lyngbya*, such as pheophytin *a*, a stable breakdown product of chlorophyll *a* (Kahn et al. 2002) and various other metabolites, it is possible to determine the percentage of ^{15}N labeling (representing new biosynthesis) of each compound at various time points and, in turn, gain insight into their turnover rates *in vivo*.

3.3.4 Growth rate, and turnover of pheophytin *a* and jamaicamide B in *Lyngbya majuscula* JHB

Jamaicamide B from *Lyngbya majuscula* JHB was originally chosen for development of this approach due to its robust ionization and detection by MALDI. Pheophytin *a*, a more stable breakdown product of chlorophyll *a*, was chosen for its role as a primary metabolite critical to energy production in these organisms (it differs from chlorophyll *a* in the lack of a central magnesium atom, removed during the acidic MALDI preparation (Kahn et al. 2002). We performed a series of experiments using typical culture conditions to show that the rate of ^{15}N incorporation into jamaicamide B (representing new biosynthesis) proceeds approximately 2 times faster than that of pheophytin *a*, and outpaces filament growth (Esquenazi et al. 2011). Extractions of culture media did not reveal any detectable amounts of secreted jamaicamide B, suggesting the molecule is either catabolized or serving as a biosynthetic intermediate (see below).

3.3.5 The effect of environmental parameters on jamaicamide B and pheophytin *a* production

This MALDI-TOF approach was used to measure the effect of manipulating culture conditions on jamaicamide B and pheophytin *a* turnover (see materials and methods for MALDI analysis procedure). The first series of experiments evaluated different light wavelengths on these two compounds to determine if there was any CCA related effect based on the homology of the 5335 and 7968 proteins discussed above to the RcaD protein (Noubir et al. 2002). *Lyngbya majuscula* JHB filaments were grown in 6-well plates in [^{15}N] $NaNO_3$ media for 5 days under different light filters (clear, green, red, and blue) to select for various light wavelengths. Each plate was positioned in the Gerwick culture laboratory to receive approximately the same overhead light intensity during the course of the experiment (~4 μE m^{-2} sec^{-1}). Initial experiments allowed ambient light to enter the plates from the sides and bottom of the plates, while additional experiments used plates wrapped in aluminum foil to block all light except that entering through the light filter. The initial experiments (Figure 3.6) found that no significant differences were apparent in jamaicamide turnover due to the green, red, or blue filters in comparison to the control, clear filter; however, pheophytin *a* turnover did decrease under the green filter.

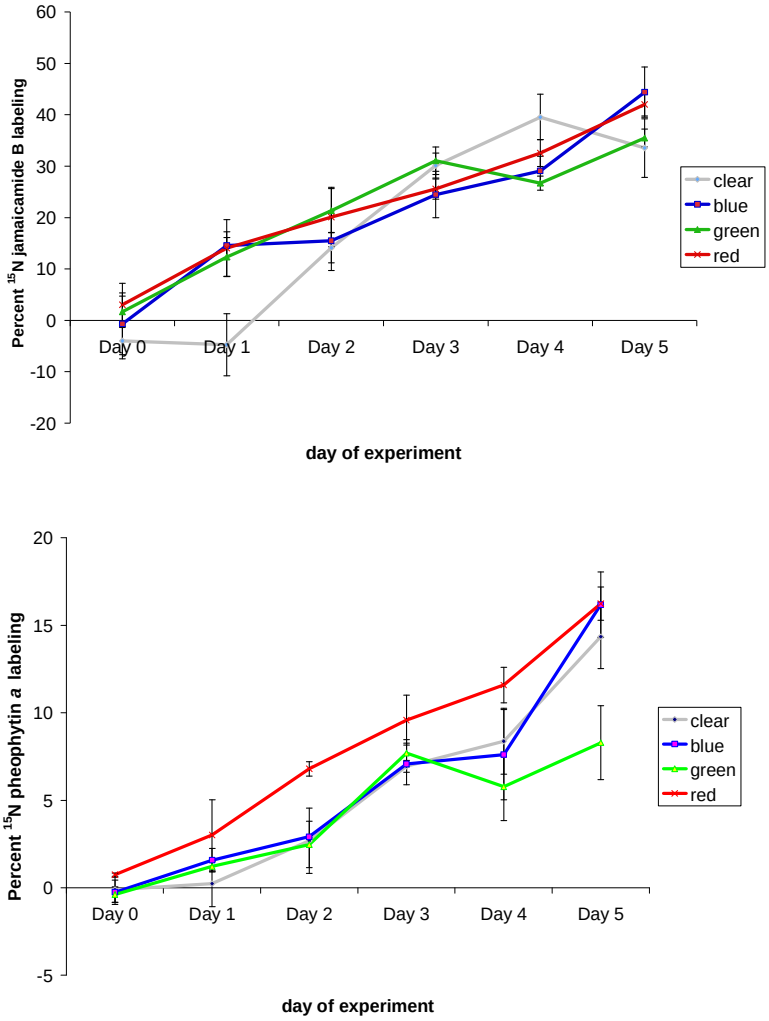
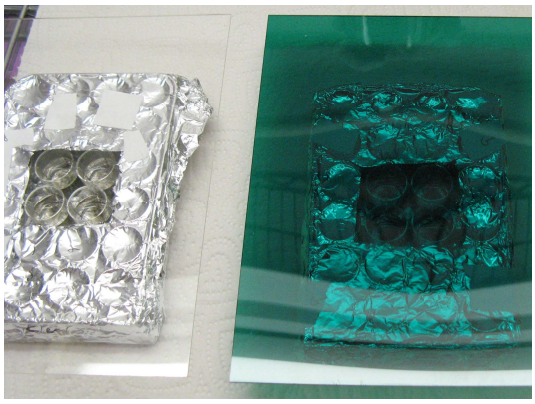


Figure 3.6: Initial experiments testing the effect of different visible wavelengths (all at ~4 μE m^{-2} sec^{-1}) on jamaicamide B (top) and pheophytin *a* (bottom) turnover, as measured by MALDI-TOF/stable isotope feeding. N = 3; Error bars are SEM.

The decrease in pheophytin *a* turnover under green light in these initial experiments, coupled with no measurable effect on jamaicamide B turnover, led us to speculate that green light may activate jamaicamide transcription. Pheophytin *a* was found to be a reasonable proxy for *Lyngbya* growth (Esquenazi et al. 2011) and possible slower growth rates under green light may indicate that jamaicamide B is

being produced faster in proportion to normal growth conditions. Additional experiments testing the green and clear filters using foil wrapped plates (Figure 3.7a) revealed more clearly that green light actually decreases jamaicamide B turnover as well as pheophytin *a* turnover, suggesting that ambient light in the initial experiments may have been sufficient to augment turnover of jamaicamide B (Figure 3.7b).

A



B

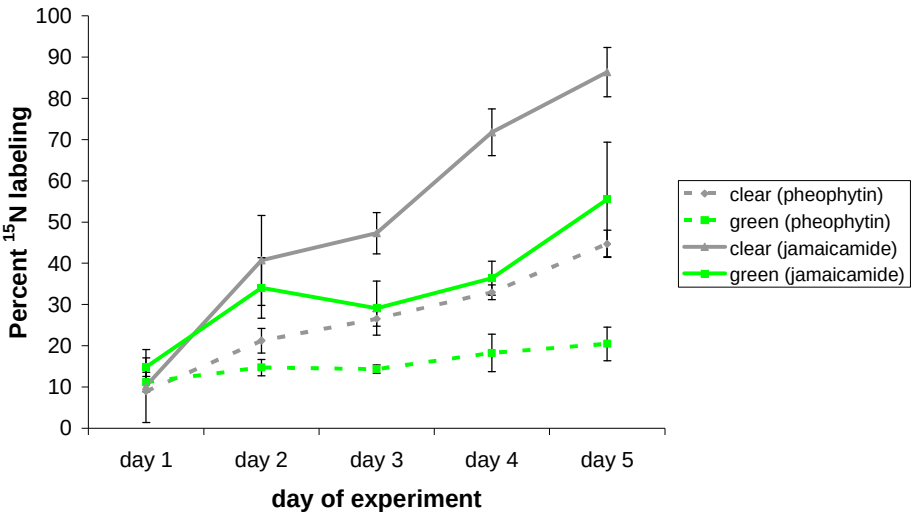


Figure 3.7: Experiments testing the effect of green light on jamaicamide B and pheophytin *a* production. A: Clear and green optical filters with foil wrapped plates containing *L. majuscula* JHB filaments. B: Percent ^{15}N labeling over time of jamaicamide B and pheophytin *a* in each treatment. N = 3; Error bars are SEM.

The next light based experiment tested the effect of ultraviolet light on jamaicamide B and pheophytin *a* turnover. Following a similar format, 6 well plates with *L. majuscula* JHB [^{15}N] $NaNO_3$ were grown at comparable light intensities under either normal grow lights or grow lights in combination with ultraviolet lights for 5 days (see materials and methods). The turnover of both molecules was significantly lower in UV light conditions, but the higher jamaicamide B turnover in relation to pheophytin *a* remained consistent (Figure 3.8; Modified from Esquenazi et al. 2011).

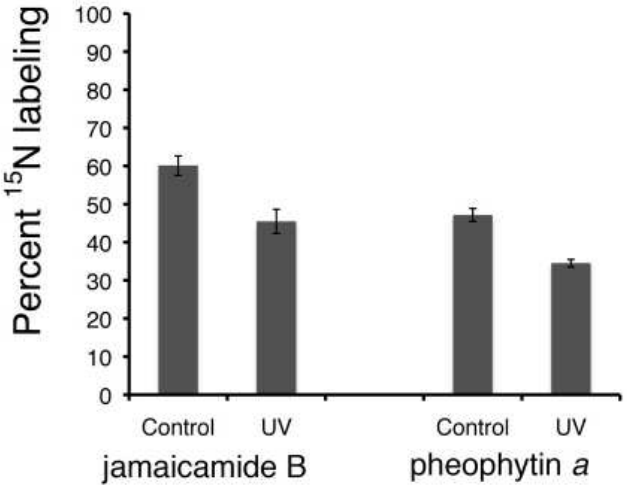


Figure 3.8: Effect of ultraviolet light on percent ^{15}N labeling of jamaicamide B and pheophytin *a* after 5 days in comparison to typical culture conditions. N = 5; Error bars are SEM (Esquenazi et al. 2011).

A final light experiment was performed to measure the effects of light intensity on jamaicamide B and pheophytin *a* turnover. Plates with [^{15}N] $NaNO_3$ media containing *L. majuscula* JHB filaments were placed adjacently to one another under the same grow lights. One of these plates was covered with a standard plastic plate

cover, while the others were covered with increasing layers of cheesecloth (serving as "neutral density filters") to decrease the amount of light reaching the *Lyngbya* filaments [Plate A, 25.0 μE m^{-2} sec^{-1}; Plate B, 14.9 μE m^{-2} sec^{-1}; plate C, 8.4 μE m^{-2} sec^{-1}; plate D, 4.9 μE m^{-2} sec^{-1}; plate E, 0 μE m^{-2} sec^{-1}; (wrapped in foil), Figure 3.9.] . An additional plate was completely wrapped in foil to block any available light. This experiment was run for 5 days. There was no significant difference between any jamaicamide B turnover measurements for those plates receiving at least some level of light (the same trend was true for pheophytin, data not shown). However, turnover of both jamaicamide B and pheophytin *a* turnover was completely abolished in the dark condition, which indicated that light is necessary for the biosynthesis of each compound to occur.

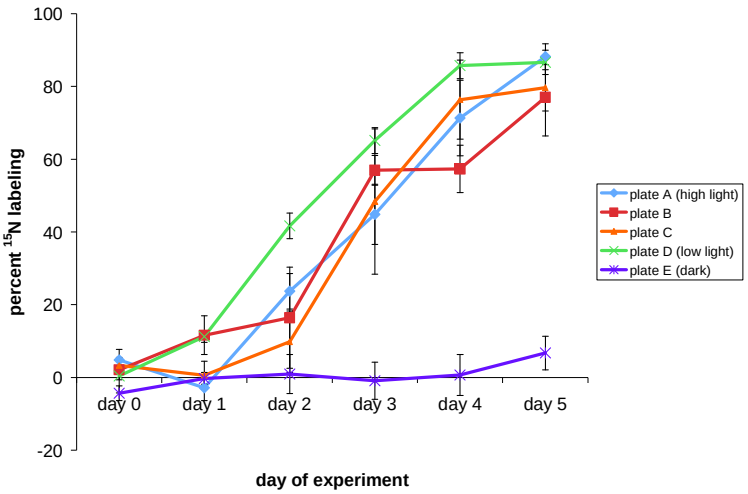


Figure 3.9: Effect of light intensity on percent ^{15}N labeling of jamaicamide B. Plate A = 25.0 μE m^{-2} sec^{-1}; plate B = 14.9 μE m^{-2} sec^{-1}, plate C = 8.4 μE m^{-2} sec^{-1}, plate D = 4.9 μE m^{-2} sec^{-1}, plate E = 0 μE m^{-2} sec^{-1} (wrapped in foil). N = 3; Error bars are SEM.

Further development of the MALDI-TOF/^{15}N feeding approach to measure jamaicamide B and pheophytin *a* turnover included several additional experiments,

including testing the effects of filament density and decreasing nitrate concentrations (Esquenazi 2010). With the exception of the completely dark condition where jamaicamide B and pheophytin *a* production was abrogated, the ratio of jamaicamide B turnover to pheophytin *a* remained surprisingly consistent (approximately 2 fold faster). This difference led us to examine the turnover of jamaicamide A in comparison to jamaicamide B (Esquenazi et al. 2011). Jamaicamide A is the brominated analog of jamaicamide B, and is halogenated by an unknown enzyme (although this is speculated to be the product of *jamD* in the jamaicamide gene cluster; Edwards et al. 2004). We found that jamaicamide A is turned over more slowly than jamaicamide B, although this turnover can be increased somewhat upon sodium bromide supplementation to the culture media. These findings, in addition to the observation that a lag in jamaicamide A turnover [time to 50% labeling (T_{50}) for jamaicamide B = 2.5 days, T_{50} for jamaicamide A = 4.0 days)] led us to hypothesize that jamaicamide B may serve as the substrate for halogenation to produce jamaicamide A.

To further explore the hypothesis that the bromination of jamaicamide A occurs after the assembly of jamaicamide B, an additional experiment was performed using a similar ^{15}N feeding time course with the inclusion of a dark phase (Esquenazi et al. 2011). In preliminary trials, it was found that *L. majuscula* JHB cultures grown in a completely and continuously dark environment effectively stopped turnover of jamaicamide B and pheophytin *a* (Figure 3.9), a finding that supports the notion of light regulation of these pathways (Jones et al. 2009). Comparison of ^{15}N incorporation into jamaicamide A and B during this dark phase revealed that

production of jamaicamide A continues in the dark, whereas production of jamaicamide B is completely abrogated (Esquenazi et al. 2011). These results again suggest that jamaicamide A bromination likely occurs after the assembly of jamaicamide B, and implicates jamaicamide B as the substrate for bromination. This prediction of the timing of bromination is consistent with previously reported experiments that demonstrated the inclusion of only non-brominated substrates during the initial steps of jamaicamide biosynthesis (Dorrestein et al. 2006). This experiment was repeated using an alternative design (see Esquenazi et al. 2011), and these latter results also support the initial experiment with ^{15}N labeling of jamaicamide A continuing in the dark condition, while ^{15}N labeling of jamaicamide B is inhibited.

These MALDI results were reinforced by those of a larger scale liquid chromatography – mass spectrometry (LC-MS) experiment. The relative amounts of jamaicamides A and B extracted from cultures grown for six days under either complete darkness or regular light cycle were compared. The ratios of total jamaicamide A to jamaicamide B in each extract (Esquenazi et al. 2011) indicate that while jamaicamide A biosynthesis continues in total darkness, jamaicamide B turnover drops considerably (T-test, $P = 0.067$), resulting in a reduced ratio of jamaicamide B to jamaicamide A.

3.4 Discussion

In concert with the transcriptional analysis of the jamaicamide gene cluster described in chapter 2, work presented in this chapter was focused on achieving a better understanding of the genetic and environmental factors that may play a role in

jamaicamide biosynthesis. Of particular interest was the successful isolation of proteins using "pulldown" experiments that could be involved in the regulation of jamaicamide expression. Gene clusters of marine cyanobacterial PKS/NRPS secondary metabolites identified to date lack any associated regulatory proteins that are imbedded in or in proximity to the main cluster, in contrast to antibiotic pathways in other prokaryotes such as actinobacteria (Bibb 2005). This absence has led to the suggestion that secondary metabolite pathways from *L. majuscula* could be constitutively expressed (Edwards et al. 2004). By using the upstream region of *jamA* as a DNA probe, we hoped to isolate putative regulatory proteins from the soluble protein fraction of JHB. This was predicted on the hypothesis that if the jamaicamide pathway does have associated regulatory proteins, they are located elsewhere in the genome. A biotinylated DNA sequence from the jamaicamide pathway (1000 bp upstream of *jamA* to 20 bp into the *jamA* gene) was incubated with protein lysate from *L. majuscula* JHB. The probe was long enough to encompass the entire untranslated leader region of the pathway, as well as the primary promoter and an additional 123 bp upstream of the promoter -35 hexamer. Because transcription factors commonly bind at either the -35 box of the promoter itself, or within 90 bp of the -35 box (Browning and Busby 2004) it is probable that the probe was long enough to capture proteins that might associate with the promoter. The probe also allowed for binding of regulatory proteins with affinity to the untranslated leader region (Li and Golden 1993; Hobl and Mack 2007). Analysis of protein samples isolated from both an excised SDS-PAGE gel band and elution fractions of several repeated pulldown assays consistently identified two proteins in three separate data sets using LC-

MS/MS. These proteins were partially identified using sequence data from the unfinished *L. majuscula* 3L genome (see Chapter 5), a strain from Curaçao responsible for the production of the anticancer compound curacin A (Chang et al. 2004; Gerwick et al. 1994). The *L. majuscula* 3L gene sequences corresponding to these proteins were used to design primers that were successful in amplifying the homologs from *L. majuscula* JHB genomic DNA. The two proteins (5335 and 7968) displayed strongest sequence identity to hypothetical proteins found in other cyanobacteria, but could not immediately be assigned a function. BLAST searches with both proteins resulted in hits with RcaD, a protein involved in complementary chromatic adaptation (CCA) in another species of cyanobacteria (Noubir et al. 2002). Interestingly, although the level of sequence identity of the two proteins with RcaD was quite different (Table 3.2), both proteins (in the 3L genome) had a similar gene neighborhood to RcaD, indicating probable synteny. The *L. majuscula* 3L proteins downstream of each (5336 and 7969) both had BLAST hits with RcaG, the ATPase associated with RcaD, although 7969 (49% identity) had a significantly higher identity than 5336 (23% identity).

Complementary chromatic adaptation has been identified in a number of freshwater (Li et al. 2008) and marine cyanobacteria (Palenik 2001). In the cyanobacterial CCA model organism *Fremyella* (= *Calothrix*, *Tolypothrix*), a photoreceptor circuit involving the Rca receptors and response regulators (RcaC, RcaE, RcaF, and RcaD) has been found to be responsible for pigment modifications under red and green light (Kehoe and Gutu 2006). RcaD appears to affect several operons during the acclimation phase of CCA (Noubir et al. 2002). Although *Lyngbya majuscula* strains have not been observed to undergo CCA in culture, there are several

color morphotypes known (for example, in our culture collection *L. majuscula* 3L is red under 16 h light/8 h dark cycles, while *L. majuscula* JHB is dark green). In addition, a microarray analysis of cyanobacteria undergoing CCA found that over 80 genes were upregulated, including many not involved in photosynthesis (Stowe-Evans et al. 2004). Considering the widespread effects that CCA regulatory proteins play in cyanobacteria, it is plausible that secondary metabolite production is regulated by homologous proteins. Regulation by light could also be in accordance with the mechanisms previously described for the microcystin biosynthetic pathway (Kaebernick et al. 2000; Kaebernick et al. 2002).

To further evaluate the function of the two possible regulatory proteins isolated in the pulldown assay, we overexpressed both proteins in *E. coli* to evaluate their respective binding affinities for the jamaicamide primary promoter region. Protein 7968 was found to bind to the proposed transcription factor binding region of the jamaicamide pathway (1000-832 bp upstream of *jamA*; Figure 3.4a), and this DNA binding activity was supported with serial protein titration (Figure 3.4b). Although we demonstrated that a control protein would not bind under the same conditions, we also found that protein 7968 was able to bind nonspecifically to several other unrelated pieces of DNA. Thus, we were unable to assign a specific sequence for 7968 binding. Attempts to cleave the GST tag from the 5335 protein were unsuccessful, and binding assays indicated that the GST+5335 fusion protein was not able to bind to the same intergenic region as 7968 (Figure 3.4a; Figure 3.5). Because of its strong binding affinity with DNA, 7968 is the better candidate protein for providing transcriptional regulation of the jamaicamide pathway. The presence of multiple intergenic promoters

in the pathway (chapter 2) could also offer other binding locations for additional regulation.

It is difficult to predict how the binding affinity of recombinant forms of 5335 or 7968 compares quantitatively with the native proteins. Noubir et al. (2002) found that native RcaD bound much more effectively to the phycocyanin 2 promoter than a recombinant version, and hypothesized that the reduced affinity may be from lack of ATPase RcaG, which facilitates binding, or from lack of phosphorylation. We attempted a dual-shift experiment with 7968 and the GST tagged 5335, but no shift differences compared to 7968 alone were observed (data not shown). It will be intriguing to determine whether 5335 and 7968 work in tandem to regulate the jamaicamide pathway, or if they require downstream neighbors (5336 or 7969) to assist in binding. Alternatively, it is possible that 7968 is the true regulator of the pathway, and 5335 was "pulled down" in the magnetic bead assay due to its sequence identity being minimally sufficient for recognition. Interestingly, protein 7968 was found to form dimers by PAGE analysis. Transcription factors often function as dimers in their association with DNA and RNAP (Browning and Busby 2004) and thus, this finding also supports 7968 as the best candidate regulatory protein identified in this study.

If transcription factors are in fact regulating the expression of secondary metabolites such as jamaicamide, it is useful to consider the potential pleiotropic role of proteins such as 7968 in regulating more than one biosynthetic pathway in *L. majuscula* JHB. There are a number of similarities in the secondary metabolite gene clusters of *L. majuscula*, such as those encoding for the jamaicamides, hectochlorin

(also produced by the JHB strain; (Ramaswamy et al. 2007) and curacin A (Chang et al. 2004; Gu et al. 2009). For example, the genes *jamA* and *hctA* are both ACP synthetases and are 58% identical, which might indicate that similar regulatory proteins associate with the upstream regions of each gene. If jamaicamide and hectochlorin are both used in defense of *L. majuscula* against predation or infection, their co-regulation would enhance the defense of the strain. It is also interesting to speculate that proteins in *L. majuscula* 3L homologous to jamaicamide regulatory proteins could be used to regulate production of curacin A. A comparison of the approximately 1700 bp that separate *jamA* from its upstream neighboring gene (a transposase) with the upstream region of *curA* from the curacin A pathway reveals that approximately 1550 bp of the up*jamA* region is 95% identical with the up*curA* region. This level of identity is not shared between the upstream regions of the jamaicamide and hectochlorin pathways. Moreover, proteins 5335 and 7968 are 99.6% and 89.5% identical with their respective homologs in *L. majuscula* 3L (the curacin A producer). If either of these two proteins functions as a pleiotropic regulator for natural products biosynthesis in *L. majuscula*, their use in overexpression efforts would be valuable in unlocking the full biosynthetic potential of these filamentous marine cyanobacteria.

Ultimately, quantitative co-transcriptional analyses of the two proteins with the rest of the jamaicamide pathway and gene knockouts will be necessary to conclusively link these proteins with jamaicamide regulation. Recent completion of the *Lyngbya majuscula* 3L draft genome (see chapter 5) may also provide additional information useful in determining how homologs of these proteins serve in pathway regulation.

The development and use of the MALDI-TOF/stable isotope feeding approach to monitor turnover of natural products in *Lyngbya* strains (Esquenazi et al. 2011) was very important in determining how light affects the turnover of jamaicamide B and ultimately gave support to the homology based predictions described above. This approach allowed for several environmental parameters to be readily tested using small amounts of *Lyngbya* biomass and enabled side-by-side measurements of jamaicamide B turnover with pheophytin *a* and jamaicamide A.

From the data presented here, it does not appear that light wavelength has a direct impact on jamaicamide B turnover. We determined that *in vivo* measurements of molecule turnover would be a more accurate metric of the effects of light on jamaicamide B than other types of experiments such as qRT-PCR experiments, since previous investigations with the microcystin pathway revealed that, despite variable amounts of microcystin transcripts generated from exposure to different colors of light, comparable microcystin concentrations were subsequently still present in laboratory culture (Kaebernick et al. 2000). The decrease in pheophytin *a* turnover upon exposure to green light is not unexpected, as green pigments are not as effective in absorbing green light as other complementary colors (Kehoe and Gutu 2006). The pigment change in these filaments was visually apparent as their color became much darker during the course of the experiment. The negative impact of ultraviolet light on both jamaicamide B and pheophytin *a* turnover is indicative that UV light may play an important role in limiting secondary metabolite production in *Lyngbya* field populations, as UV exposure in the environment is much greater than in laboratory cultures. This effect is also contrary to the cyanobacterial UV absorbing pigment

scytonemin (Sorrels et al. 2009) where expression levels increase upon UV exposure; however, when produced scytonemin provides cyanobacteria with UV protection.

The most revealing finding from the light-based MALDI experiments was that both pheophytin *a* and jamaicamide B turnover are completely inhibited in the dark. The light intensity gradient experiment ranged from light levels significantly higher than normal (25 μE m^{-2} sec^{-1}) to typical laboratory conditions (~5 μE m^{-2} sec^{-1}), in addition to the dark condition. The comparable turnover for these two molecules suggest increased amounts of visible light do not have a significant impact on biosynthesis (although increased mortality was observed at higher light levels), but that a minimal light level must be achieved; this is the same pattern observed for separate nitrate limiting experiments (Esquenazi et al. 2011). The lack of pheophytin *a* turnover in the dark is again not surprising as chlorophyll is used in photosynthesis. However, abolishment of jamaicamide B turnover in the dark was instrumental in designing subsequent experiments comparing jamaicamide B and A turnover.

As described separately (Esquenazi 2010, Esquenazi et al. 2011) other experiments found that although jamaicamide A turnover under typical culture conditions lags behind that of jamaicamide B, the turnover of jamaicamide A can be augmented with the addition of sodium bromide, showing that bromide is a limiting nutrient in jamaicamide A biosynthesis. This increased turnover still does not equal that of jamaicamide B, but this inequality also supports that jamaicamide B is biosynthesized first before the final bromination step (possibly *via* JamD encoded in the jamaicamide gene cluster). Inclusion of a dark phase allowed us to measure jamaicamide A turnover occurring in the dark while jamaicamide B turnover was

inhibited (Esquenazi et al. 2011). This measurement shows that a previously expressed halogenase could be acting on intact jamaicamide B, and is consistent with previous mass spectrometry data showing that only non-brominated substrates are accepted by jamaicamide enzymes in initial stages of jamaicamide biosynthesis (Dorrestein et al. 2006). Although the mechanism of light regulation of the jamaicamide pathway still requires further investigation, these data have provided the best understanding of natural product regulation in a marine cyanobacterium to date.

Understanding the regulation of natural product pathways that encode compounds with pharmaceutical potential is important to overcoming the "supply issue" that is so prevalent in natural products research (Cragg et al. 1999). While marine cyanobacteria are recognized as prolific producers of bioactive compounds, natural product yields from field collections are low, and slow culture growth severely limits the amount of compound that can be produced in this manner. Insights into jamaicamide biosynthetic gene cluster regulation presented here from transcription factor isolation and MALDI-based measurements provide evidence consistent with light playing a direct role in regulating compound production. Future advances in identifying how these types of gene clusters are regulated will be important to diverse applications in biotechnology, such as combinatorial biosynthesis and the heterologous expression of entire natural product pathways. Additionally, this information could also benefit ongoing efforts attempting to regulate the expression of cyanobacterial toxins that have deleterious environmental impacts.

3.5. Materials and methods

3.5.1 Bacterial strains, culture conditions, PCR reactions, and DNA measurements

Lyngbya majuscula JHB was originally collected from Hector's Bay, Jamaica (Edwards et al. 2004) and was maintained in a culture facility at Scripps Institution of Oceanography. Cultures were grown and maintained in BG-11 saltwater media (SW BG-11; Castenholz 1988) at 29°C under a light intensity of approximately 5 μE m^{-2} s^{-1} and under 16 h light/8 h dark cycles. For MALDI experiments, the $NaNO_3$ content of the SW BG-11 media (5 g/L) was replaced with ^{15}N $NaNO_3$ (98%, Cambridge Isotopes). *E. coli* TOP-10 and BL-21 (DE3) were grown in Luria-Bertani (LB) media. *E. coli* cultures were grown with ampicillin (100 μg ml^{-1}), or kanamycin (50 μg ml^{-1}) when necessary. PCR reactions were conducted using either PCR Master Mix (Promega) or Pfx50 proofreading Taq Polymerase (Invitrogen). DNA concentrations were measured using either Beckman-Coulter DU800 or NanoDrop 1000 (Thermo Scientific) spectrophotometers. Protein concentrations for recombinant JHB proteins were determined using the BCA assay (Pierce). Ladders for DNA (Fermentas and New England Biolabs) and protein (Bio-Rad) were used for size estimations when necessary.

3.5.2 Isolation of possible transcription factors from a pulldown assay

Protein pulldown experiments were based on methods similar to Rachid et al. 2007. A DNA probe that extended from 1000 bp upstream of *jamA* to 20 bp into the *jamA* gene was amplified by PCR from the jamaicamide fosmid described above using the primers up*jamA* 1000 biotin (biotinylated at the 5' end; Invitrogen) and

up*jamA* 20 - 0 R (Table 3.1). The PCR product was purified (MinElute PCR Purification Kit, Qiagen) and 10 pmol of the biotinylated DNA were incubated with 1 mg of magnetic M-270 streptavidin Dynabeads (Invitrogen), according to the manufacturer's instructions. *L. majuscula* JHB tissue was obtained from pan cultures that had been growing for 1-2 months. Approximately 2-3 ml of culture was measured by displacement in sterile, chilled binding buffer [10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 150 mM NaCl, and 5% (w/v) glycerol]. The binding buffer was also treated with a broad range protease inhibitor (Complete, EDTA free; Roche). The tissue was sonicated and kept on ice using a probe sonicator with six 10-s pulses, and insoluble material was pelleted at 13,200 RPM for 10 minutes. The soluble protein fraction (750 µl) was added to each mg of DNA coated beads. One µg of Poly DI-DC was also added to inhibit non-specific binding of protein to the DNA. Magnetic beads that were not treated with biotinylated DNA were incubated with JHB soluble protein as a negative control. The beads and soluble protein were incubated for 1 h using an end-over-end rotator at 4°C. The beads were subsequently washed twice using 200 µl of binding buffer containing 100 µl sheared salmon sperm DNA (Invitrogen; 5 mg ml^{-1}), three times with binding buffer, and eluted with 50 µl of binding buffer containing 1.0 M NaCl. Wash fractions and elutions were flash frozen in liquid N_2 and concentrated on a freeze dryer. After concentration, aliquots of each were mixed with protein sample buffer, denatured for 3 minutes at 95-100°C, and analyzed by SDS-PAGE. The gels were stained with either silver (Silverquest Kit, Invitrogen) or colloidal Coomassie brilliant blue G-250.

3.5.3 Identification of DNA binding proteins

Once gel bands were visible in the elution fraction from the binding assay, the assay was repeated on a larger scale using additional replicates of the procedure described above to isolate sufficient protein for mass spectrometry (visible by colloidal Coomassie staining). Both gel bands (excised using a scalpel) and whole elution fractions were submitted to The Scripps Research Institute (La Jolla, CA) Center for Mass Spectrometry for nano-LC MS/MS analysis. Raw spectrum data (mzdata format) was obtained and analyzed at UCSD by a DOS common-line version of InsPecT 20070712 (Tanner et al. 2005).

InsPecT search parameters for the mzdata files were the following: (i) *Lyngbya majuscula* 3L common database (see Chapter 5), common contaminants database, reverse or "phony" database, and NCBI nr database; (ii) parent ion $\Delta m = 1.5$ Da; (iii) b and y-ion $\Delta m = 0.5$ Da. Top protein identifications were verified by using two different database searches: (i) *Lyngbya majuscula* 3L genome alone; (ii) NCBI nr with *L. majuscula* 3L genome inserted. The mass spectral identifications of 5335 and 7968 were further verified by manual annotation of the N-terminal and C-terminal peptides, as well as the most abundant peptide identified.

3.5.4 Characterization of putative transcription factors from a pulldown assay

Protein sequences detected using InsPecT were compared with raw nucleotide sequences from the *L. majuscula* 3L genome to identify their corresponding ORFs. Forward and reverse primers (5335 F &R, 7968 F &R, Table 3.1) were designed from each sequence and used to amplify the corresponding genes from *L. majuscula* JHB.

The blunt PCR products were cloned (Z-Blunt TOPO vector, Invitrogen) and transformed into *E. coli* for sequencing to compare the gene sequences from JHB with those of 3L. Additional gene boundary primers (5335 FB, 5335 RB; 7968 FB, 7968 RB; Table 3.1) were used to amplify the JHB genes with priming sites 25 bp upstream and downstream in order to verify the sequences covered by 5335 and 7968 forward and reverse primers and avoid inclusion of sequences from *L. majuscula* 3L. Bioinformatic analyses of each gene sequence were conducted using BLAST programs available through the National Center for Biotechnology Information (NCBI; http://blast.ncbi.nlm.nih.gov).

3.5.5 Recombinant expression of identified proteins

Genes corresponding to identified proteins in the JHB protein pulldown assay were amplified from JHB genomic DNA using the primers 5335 Nco1F and 5335 Not1R or 7968 Nde1F and 7968 Xho1R (Table 3.1). Start codons in both gene sequences were changed from either leucine (5335) or valine (7968) to methionine for improvements in *E. coli* overexpression efforts. PCR products from both reactions were purified and digested with their corresponding restriction enzymes (New England Biolabs). Gene 5335 was ligated into the pGS21a vector (Genscript), which contains both an N-terminal 6× His tag and GST tag. The 5335 construct was verified via transformation, plasmid isolation from TOP-10 *E. coli*, and sequencing. The vector containing the gene sequence was then transformed into BL-21 (DE3) *E. coli*. Four liters of *E. coli* harboring the 5335 pGS21a vector were grown (using starter cultures) for 4 h at 37°C to an OD_{600} between 0.6 and 1.0, induced with 0.7 mM IPTG, and then

grown at 18°C overnight. Cultures were centrifuged at 4500 RPM for 15 min at 4°C, and the ensuing pellets from each liter of culture were resuspended in 5 ml protein lysis buffer (20 mM Tris, pH 8.0, 500 mM NaCl, and 20 mM imidazole) and lysed on ice using sonication (6-7 10-s pulses). Resuspended lysate was centrifuged, and supernatant containing soluble protein was incubated on an end-over-end rotator at 4°C with Nickel-Superflow resin (Qiagen) for 2 h. Following incubation, the recombinant GST+5335 fusion protein was purified using polypropylene columns (Qiagen). The nickel slurry from the incubation was washed twice with protein wash buffer (20 mM Tris, pH 8.0, 500 mM NaCl, and 50 mM imidazole), and protein was eluted with 5 × 1 mL aliquots of protein elution buffer (20 mM Tris, pH 8.0, 500 mM NaCl, and 750 mM imidazole). Purified protein was dialyzed against binding buffer (the same buffer used for the pulldown assay, but not containing DTT) overnight using a 50 kDa MWCO dialysis membrane (Spectrum Labs, Rancho Dominguez, CA). To express 7968, the corresponding gene sequence was ligated into the N-terminal His tag containing pET28b vector (Novagen), and grown and purified similarly to 5335 (although the 7968 wash buffer contained 40 mM imidazole). Spectra/Por Float-A Lyzer G2 dialysis membranes (20 kDa MWCO; Spectrum Labs) were used to dialyze protein 7968. Concentrations of each protein for use in assays were determined using the BCA assay (Pierce).

3.5.6 Electromobility shift assays (EMSAs)

Gel shift EMSAs were performed to verify binding of 5335 and 7968 to jamaicamide promoter regions. The region upstream of the jamaicamide TSS (1000 -

832 bp upstream of *jamA*) was amplified from a jamaicamide fosmid using Pfx50 Taq Polymerase (Invitrogen). Each PCR product was purified (MinElute kit, Qiagen) before being used in the assay. For the comparative binding assay (Figure 3.4a), the N-terminal His tag was cleaved from protein 7968 using the Thrombin Cleavage Capture Kit (Novagen). The cleaved 6× His tag was subsequently removed by concentrating the protein sample over a Microcon 10,000 MWCO column (Millipore). SDS-PAGE gels and western blotting were conducted to confirm the success of the cleavage reaction. Serial titration of the 7968 protein to verify DNA binding was performed using the recombinant version of the protein without His tag cleavage. GST+5335 elutions were also concentrated over Microcon 10,000 MWCO columns prior to use in shift assays. For a negative control, the purified recombinant protein $HctE_{IVA}$ from the hectochlorin pathway (purification described Ramaswamy et al. 2007 was used. The concentrations of $HctE_{IVA}$ protein used in the EMSA experiments were measured using a Bradford assay, and the purified $HctE_{IVA}$ included a 6× N-terminal His tag from its expression vector (pET28b; Novagen). Each gel shift assay reaction was performed with the indicated quantities of DNA and purified protein (Figure 3.4) in EMSA binding buffer adapted from the DIG Gel Shift Kit, 2nd generation protocol (Roche) [20 mM HEPES, pH 7.6, 1 mM EDTA, 10 mM $(NH_4)_2SO_4$, 1 mM DTT, Tween 20, 0.2% (w/v), 30 mM KCl] and water (total volume 20 μl) for 30 min at room temperature. Following the incubation period, 5μl of native loading dye containing bromophenol blue was added to each reaction, and the reaction contents were immediately transferred to a 10% native PAGE gel. The gels were electrophoresed at 85 V for ~3.0 h in 0.5× TBE buffer (44.5 mM Tris, 44.5 mM boric

acid, 1 mM EDTA), followed by staining for at least 10 min in SYBR Gold Nucleic Acid Gel Stain (Molecular Probes/Invitrogen) and visualization on a UV transilluminator.

3.5.7 Sequence information

DNA and amino acid sequences of the proteins 7968 and 5335 identified in this study have been deposited in Genbank under the accession numbers GQ860962 and GQ860963.

3.5.8 MALDI-based experiments measuring turnover of jamaicamides B and A and pheophytin *a*

Detailed methodology of use of the MALDI instrument and the MALDI instrument parameters for all studies is available separately (Esquenazi et al. 2011). Selected methods from this publication are reprinted below.

3.5.9 Filament sampling

Using aseptic technique and small tweezers, 2-4 filaments (~ 0.5 μg wet weight) from small cultures were patted dry on a Kimwipe, placed in a PCR tube and flash frozen with dry ice. Sampling intervals varied between experiments, as is explained in each experimental description.

3.5.10 MALDI MS Sample preparation

Approximately 1 μL of MALDI matrix solution (Per 1 mL: 35 mg α-cyano-4-

hydroxycinnamic acid (CHCA), 35 mg 2,5-Dihydroxybenzoic acid (DHB) (Universal MALDI matrix, Sigma Aldrich), 750 μL acetonitrile, 248 μL milliQ H_2O, 2 μL TFA) per 0.1 μg of biomass was mixed in a tube or well. After 20-30 seconds, 1 μL of this crude matrix solution was deposited on a well (spot) of the Bruker Microflex MSP 96 Stainless Steel Target Plate. After each spot had dried at room temperature, the plate was analyzed using a Bruker Microflex MALDI-TOF mass spectrometer equipped with flexControl 3.0.

3.5.11 Calculations

For a particular compound in a sample, the mean mass was computed using the weighted average of either the observed isotope cluster (light wavelength, UV, jamaicamide B vs. A comparisons; Kahn et al. 2002), with values obtained from peak measurements in ClinProTools software (Bruker) or by calculating the percent total, single, double and unlabeled peaks for jamaicamide and pheophytin in specific instances (light intensity experiments; Biemann 1962).

3.5.12 Light wavelength experiments with *Lyngbya majuscula* JHB

Initial wavelength experiments with *L. majuscula* JHB were conducted by placing filaments in three separate wells of 6-well plates containing ^{15}N SW BG-11, in addition to one control well with ^{14}N SW-BG-11. Each plate was covered by a different optical filter (Edmund Optics, 5 x 7 Optical Cast Plastic Filters) and sampled daily for 5 days before MALDI analysis. For further study of the effect of green light on jamaicamide B and pheophytin *a* turnover, *L. majuscula* JHB filaments were

placed in four central wells of a 24 well plate (3 wells containing ^{15}N SW BG-11, 1 control well with ^{14}N SW BG-11). The rest of the plate was wrapped completely in aluminum foil to prevent any ambient light from entering the plate. The filaments in each condition were sampled daily before MALDI analysis.

3.5.13 Light intensity experiments with *Lyngbya majuscula* JHB

Light intensity experiments were performed using 6 well plates containing 3 ^{15}N SW BG-11 wells with *L. majuscula* JHB and one control ^{14}N well as above. Light measurements through each of the plate covers were made with measured with an ILT-1400A light meter (International Light Technologies, Peabody, MA). Plates were covered with either a standard plate lid (25.0 μE m^{-2} sec^{-1}); or increasing layers of white cheesecloth (14.9 μE m^{-2} sec^{-1}, 8.4 μE m^{-2} sec^{-1}, 4.9 μE m^{-2} sec^{-1}, respectively). The dark condition was created by wrapping another plate completely in aluminum foil (0 μE m^{-2} sec^{-1}). Filaments from each well were sampled daily before MALDI analysis.

3.5.14 Effect of ultraviolet exposure of *Lyngbya majuscula* JHB on turnover of jamaicamide B and pheophytin *a*

Both UV and control plates were inoculated as described above for the 10 day experiment (N = 5). Both plates were kept at 33 μE m^{-2} sec^{-1} light intensity (measured with an ILT-1400A light meter). The UV levels were 165.7 μw cm^{-2} UVA (Omega Engineering, Stamford, CT) and 400 μw cm^{-2} UVA+B (Solarmeter 5.0, Solartech,

Inc., MI). Other than these light factors, cultures were kept in the same conditions described above and sampled on days 0 and 5.

3.5.15 Dark phase with sodium bromide supplementation in *Lyngbya majuscula* JHB:

Two 24-well culture plates were each inoculated with 8 small (2.5 mg) *L. majuscula* JHB (filaments cut to 0.5-1.0 cm length) cultures in 2 mL ^{15}N media containing 1.0 g/L NaBr, placed under the control culture conditions (above) and sampled on days 0-4. On the fourth day, one of the plates was wrapped completely with aluminum foil, while the other reference plate remained in control conditions. On the ninth day, the experimental plate was un-wrapped, with samples taken from the reference and experimental cultures on day 9 as well as days 10 and 14. At the conclusion of the study, the samples were analyzed by MALDI and the percentage of total ^{15}N labeled jamaicamide B and A was calculated for each sample at each time point.

3.5.16 LC-MS comparison of jamaicamide B and A in 16 h light/8 h dark vs. 24 h dark

Two 6-well plates were inoculated with 6 small *L. majuscula* JHB cultures (10 mg) in regular ^{14}N BG-11 media. The control plate was maintained in regular light conditions and the other was completely wrapped in aluminum foil. After six days the experimental plate was unwrapped and the vertically adjacent cultures in both plates were combined, to give three, larger samples per plate. All of the samples were extracted with 80% EtOAC/hexanes for approximately 1 h with stirring at 25 °C. Each

of these extracts were dried and redissolved in 80% MeOH/H_2O and run over C-18 solid phase extraction cartridges (Varian or Analytichem International, 100 mg) in the same solvent system. All extracts were dried under liquid nitrogen, weighed, and suspended in MeOH at a concentration of approximately 3 mg/mL. The extracts were profiled using a gradient program of 70-100% MeOH/H_2O with an analytical column (Phenomenex Jupiter C-18, 5 μ 300 A) and LC-MS (Thermo-Finnigan Surveyor pump and PDA and LCQ Advantage Max). Retention times for jamaicamide A and B were identified by UV detection, and the total ion counts for all associated adducts (found to repeatedly include $(M+H)^+$, $(M+Na)^+$, $(M+K)^+$ and $(M+45)^+$) of jamaicamide B and A were combined and integrated using Xcalibur software (Thermo Electron) to determine the ratio of jamaicamide A to jamaicamide B in each sample.

3.6 Acknowledgements

We thank Dr. Hyukjae Choi for assistance with analyzing LC-MS data. The chapter author was supported by a graduate fellowship from the Los Angeles Chapter of the Achievement Rewards for College Scientists (ARCS) Foundation, and NOAA Grant NA08OAR4170669, California Sea Grant College Program Project SG-100-TECH-N, through NOAA's National Sea Grant College Program, US Department of Commerce. Chapter 3 of this dissertation contains portions of two manuscripts: 1). Jones, Adam C.; Gerwick, Lena; Gonzalez, David; Dorrestein, Pieter C.; Gerwick, William H. “Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins.” *BMC Microbiology* 2009, 9:247; and 2) Esquenazi, Eduardo; Jones, Adam

C.; Byrum, Tara; Dorrestein, Pieter C.; Gerwick, William H. Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proceedings of the National Academy of Sciences* 2011, 108: 5226-5231. As mentioned above, the dissertation author is the primary investigator and author of the first paper. For the second paper, the dissertation author was involved in several of the MALDI experiments described, as well as organic extractions, LC-MS profiling of abstracts, experimental design, and writing the manuscript.

3.7. References

Bibb MJ (2005) Regulation of secondary metabolism in Streptomycetes. *Curr Op Microbiol* 8:208-215.

Biemann K (1962) Mass spectrometry: Organic Chemical Applications. New York: McGraw-Hill, pp. 223-231.

Blunt JW Copp BR, Hu W-P, Munro MHG, Northcote PT, Prinsep MR (2009) Marine natural products. *Nat Prod Rep* 26:170-244.

Browning DF, Busby SJW (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* 2:57-65.

Cardellina JH II, Marner FJ, Moore RE (1979) Seaweed dermatitis: Structure of lyngbyatoxin A. *Science* 204:193-195.

Castenholz RW (1988) Culturing of cyanobacteria. *Meth Enzymol* 167:68-93.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula***.** *J Nat Prod* 67:1356-1367.

Cragg GM, Newman DJ, Snader KM (1999) Natural products in drug discovery and development: The United States National Cancer Institute Role. In *Phytochemicals in Human Health Protection, Nutrition, and Plant Defense*. Edited by Romeo JT. New York: Kluwer Academic/Plenum Publishers; 1-29.

Cronin G, Hay ME (1996) Induction of seaweed chemical defenses by amphipod grazing. *Ecol* 77:2287-2301.

Cruz-Rivera E, Paul VJ (2007) Chemical deterrence of a cyanobacterial metabolite against generalized and specialized grazers. *J Chem Ecol* 33:213-217.

Dorrestein PC, Blackhall J, Straight PD, Fischbach MA, Garneau-Tsodikova S, Edwards DJ, McLaughlin S, Lin M, Gerwick WH, Kolter R, Walsh CT, Kelleher NL (2006) Activity screening of carrier domains within nonribosomal peptide synthetases using complex substrate mixtures and large molecule mass spectrometry. *Biochem* 45:1537-1546.

Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula***.** *Chem Biol* 11:817-833.

Esquenazi E (2010) Spatial and temporal dynamics of marine natural products biosynthesis. PhD Dissertation, University of California San Diego.

Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226-5231.

Frias-Lopez J, Bonheyo GT, Fouke BW (2004) Identification of differential gene expression in bacteria associated with coral black band disease by using RNA-arbitrarily primed PCR. *Appl Environ Microbiol* 70:3687-3694.

Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blokhin A, Slate DL (1994) Structure of curacin A, a novel antimitotic, antiproliferative, and brine shrimp toxic natural product from the marine cyanbacterium *Lyngbya majuscula*. *J Org Chem* 59:1243-1245.

Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Håkansson K, Wipf P, Smith JL, Gerwick WH, Sherman DH (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature* 459:731-735.

Herms DA, Mattson WJ (1992) The dilemma of plants: To grow or defend. *Quart Rev Biol* 67**:** 283-335.

Hobl B, Mack M (2007) The regulator protein PyrR of *Bacillus subtilis* specifically interacts *in vivo* with three untranslated regions within *pyr* mRNA of pyrimidine biosynthesis. *Microbiol* 153:693-700.

Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.

Jones AC, Monroe EA, Eisman EB, Sherman DH, Gerwick WH (2010) The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* 27:1048-1065.

Kaebernick M, Dittmann E, Börner T, Neilan BA (2002) Multiple alternate transcripts direct the biosynthesis of microcystin, a cyanobacterial nonribosomal peptide. *Appl Environ Microbiol* 68:449-455.

Kaebernick M, Neilan BA, Börner T, Dittmann E (2000) Light and the transcriptional response of the microcystin biosynthesis gene cluster. *Appl Environ Microbiol* 66:3387-3392.

Kahn ML, Parra-Colmenares A, Ford CL, Kaser F, McCaskill D, Ketchum RE (2002) A mass spectrometry method for measuring ^{15}N incorporation into pheophytin. *Anal Biochem* 307:219-225.

Kehoe DM, Gutu A (2006) Responding to color: The regulation of complementary chromatic adaptation. *Ann Rev Plant Biol* 57:127-150.

Li L, Alvey RM, Bezy RP, Kehoe DM (2008) Inverse transcriptional activities during complementary chromatic adaptation are controlled by the response regulator RcaC binding to red and green light-responsive promoters. *Mol Microbiol* 68:286-297.

Li R, Golden SS (1993) Enhancer activity of light-responsive regulatory elements in the untranslated leader regions of cyanobacterial *psbA* genes. *Proc Natl Acad Sci USA*, 90:11678-11682.

Nagle DG, Paul VJ (2002) Production of secondary metabolites by filamentous tropical marine cyanobacteria: Ecological functions of the compounds. *J Phycol* 35:1412-1421.

Noubir S, Luque I Ochoa de Alda JAG, Perewoska I, Tandeau de Marsac N, Cobley JG, Houmard J (2002) Coordinated expressed of phycobiliprotein operons in the chromatically adapting cyanobacterium *Calothrix* PCC 7601: A role for RcaD and RcaG. *Mol Microbiol* 43:749-762.

Palenik B (2001) Chromatic adaptation in marine *Synechococcus* strains. *Appl Environ Microbiol* 67:991-994.

Rachid S, Gerth K, Kochems I, Müller R (2007) Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium *Sorangium cellulosum* So ce56. *Mol Microbiol* 63:1783-1796.

Ramaswamy AV, Sorrels CM, Gerwick WH: Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 2007, 70:1977-1986.

Rossi JV, Roberts MA, Yoo H-D, Gerwick WH (1997) Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically-useful natural metabolite curacin A. *J Appl Phycol* 9:195-204.

Schyns G, Jia L, Coursin T, Tandeau de Marsac N, Houmard J (1998) Promoter recognition by a cyanobacterial RNA polymerase: *In vitro* studies with the *Calothrix* sp. PCC 7601 transcriptional fators RcaA and RcaD. *Plant Mol Biol* 36:649-659.

Sharif DI, Gallon J, Smith CJ, Dudley E (2008) Quorum sensing in cyanobacteria: *N*-octanoyl-homoserine lactone release and response, by the epilithic colonial cyanobacterium *Gloeothece* PCC6909. *ISME J* 2:1171-1182.

Sobczyk A, Bely A, Tandeau de Marsac N, Houmard J (1994) A phosphorylated DNA binding protein is specific for the red-light signal during complementary chromatic adaptation in cyanobacteria. *Mol Microbiol* 13:875-885.

Sorrels CM, Proteau PJ, Gerwick WH (2009) Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl Environ Microbiol* 75:4861-4869.

Stowe-Evans EL, Ford J, Kehoe DM (2004) Genomic DNA microarray analysis: Identification of new genes regulated by light color in the cyanobacterium *Fremyella diplosiphon*. *J Bacteriol* 186:4338-4349.

Tan LT (2007) Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochem* 68:954-979.

Tanner S, Shu H, Frank A, Wang L-C, Zandi E, Mumby M, Pevzner PA, Bafna V (2005) InsPecT: Identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem* 77:4626-4639.

Volk R-B (2007) Studies on culture age versus exometabolite production in batch cultures of the cyanobacterium *Nostoc insulare*. *J Appl Phycol* 19:491-495.

CHAPTER 4

Steps toward the development of a stable heterologous expression system for natural products from filamentous marine cyanobacteria

4.1 Abstract

Filamentous marine cyanobacteria are rich sources of natural products and often employ highly unusual biosynthetic enzymes in secondary metabolite assembly. However, the lack of developed techniques for stable transfer of DNA into these filaments, and the absence of heterologous expression strategies for cyanobacterial gene clusters, limit access to higher yields of cyanobacterial natural products for further development and downstream applications. The work presented in chapter 4 includes attempts to transfer plasmid DNA into *Lyngbya* strains and explore the utility of different bacteria (*E. coli* BL21, *Nostoc* sp. PCC7120, and *Streptomyces coelicolor* M512) to act as expression hosts for *Lyngbya* biosynthetic genes. Although no DNA transfer into *Lyngbya* filaments was achieved, *S. coelicolor* M512 showed promise as a possible heterologous expression platform for *Lyngbya* metabolites. Using an inducible vector, *S. coelicolor* was able to overexpress both LtxB and LtxC from the lyngbyatoxin gene cluster. Functionality of LtxC produced by *S. coelicolor* was demonstrated through *in vitro* assays. Attempts to express the non-ribosomal peptide synthetase (NRPS) LtxA and the entire lyngbyatoxin pathway in *S. coelicolor* were unsuccessful. This may be a result of the difference in the percentage of GC of introduced DNA and/or the high occurrence of the rare leucine codon UUA in LtxA. While codon optimization may be needed to produce *Lyngbya* natural products in *Streptomyces*, the heterologous expression of cyanobacterial enzymes in actinobacteria

may be a valuable complement to actinobacterial natural product biosynthesis and increase molecular diversity in these strains.

4.2 Introduction

The impact of microbial natural products in drug discovery and biotechnology applications has been tremendously enhanced as a result of significant improvements in genetic techniques and genome sequencing over the past decade (Wilkinson and Micklefield 2007; Lane and Moore 2010). Genome sequences for strains of natural product rich model genera such as *Bacillus*, *Myxococcus*, and *Streptomyces* are rapidly becoming available (Joint Genome Institute; jgi.doe.gov), and several milestones have been achieved in these systems for targeted gene replacement (Gust et al. 2003), enhancement of natural product production (Rachid et al. 2007), and heterologous expression of biosynthetic gene clusters using both complex construct assembly (Watanabe et al. 2006) and phage mediated homologous recombination (Eustáquio et al. 2005; Heide 2009).

Despite being prolific producers of secondary metabolites, a complete lack of genetic techniques for marine filamentous cyanobacteria has, in contrast to other bacterial producers, limited access to large amounts of several important natural products. Studies of these natural products have instead resulted from repeated and large scale field collections, long-term culture efforts (Rossi et al. 1997), or sophisticated synthetic methods (Suyama and Gerwick 2008). As discussed in chapters 1 and 3, filamentous cyanobacteria such as *Lyngbya majuscula* grow very slowly, with doubling times of 6 days or more (Esquenazi et al. 2011), and *Lyngbya*

cells are encased in a thick polysaccharide sheath (Gerwick et al. 2008) that inhibits straightforward DNA introduction with conventional laboratory transformation methods. It is also not currently possible to maintain axenic *Lyngbya* strains due to the presence of heterotrophic bacteria that persist on the surface of the sheath (Gerwick et al. 2008; Grindberg et al. 2011). These bacteria may present the cyanobacterial filaments with additional restriction endonucleases that minimizes uptake of intact exogenous DNA into cyanobacterial cells.

Establishment of genetic techniques for *Lyngbya* strains and/or a stable heterologous expression platform for *Lyngbya* natural product gene clusters would greatly enhance the utility of these organisms in drug discovery efforts. Insertion of regulatory proteins such as those described in chapter 3 could augment production levels in cultured strains, while targeted gene addition, replacement, knockouts, or precursor directed mutasynthesis in a different host would add new molecular diversity, as well as permit rapid scale up to generate higher compound yields.

The two aims of the research presented in this chapter are to 1) evaluate the efficacy of different techniques to introduce DNA into *Lyngbya majuscula* cells, and 2) test three different bacteria [*E. coli* BL21, *Nostoc* (*Anabaena*) sp. PCC7120, and *Streptomyces coelicolor* M512] for their potential as heterologous expression hosts for *Lyngbya* secondary metabolites. The majority of this chapter will focus on the second aim, as attempts to transfer DNA into *Lyngbya* strains proved largely unsuccessful. Heterologous expression efforts focused on the gene cluster encoding the lyngbyatoxins (Figure 4.1), protein kinase C activating molecules (Basu et al. 1992) from a Hawaiian strain of *Lyngbya majuscula* (Cardellina II et al. 1979). The

lyngbyatoxins are structurally related to the teleocidin class of natural products from *Streptomyces* (Fujiki et al. 1981) and are the causative agent of a form of blistering dermatitis known as “Swimmer’s itch”.

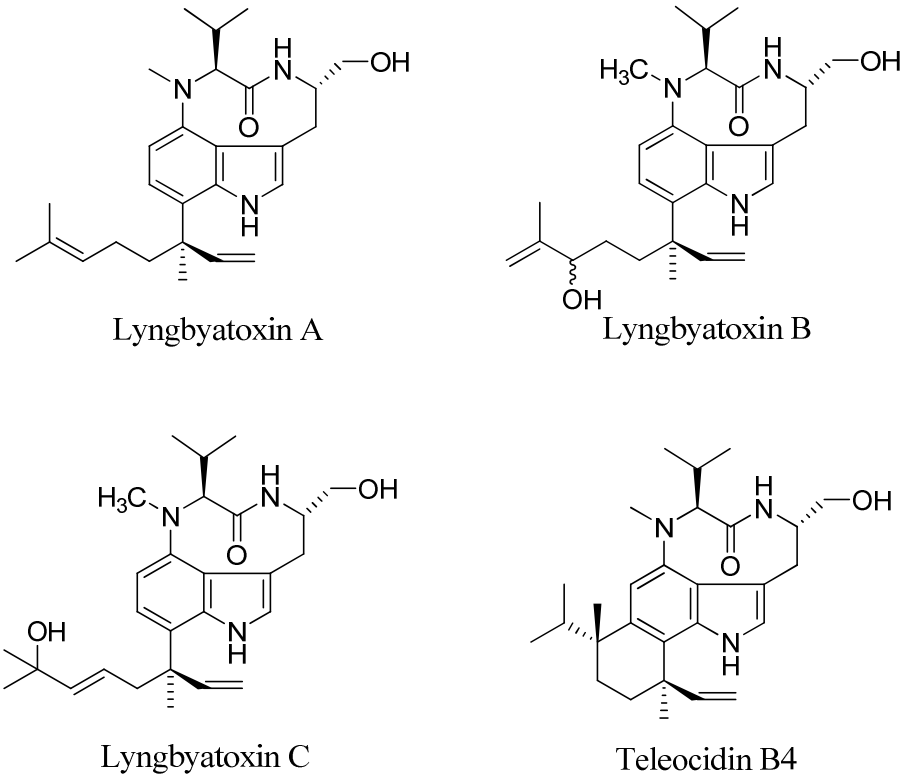


Figure 4.1: Lyngbyatoxins A-C from *Lyngbya majuscula* and teleocidin B4 from *Streptomyces blastmyceticum* (Irie et al. 1998).

The lyngbyatoxin cluster was originally identified through creation of a DNA fosmid library using genomic DNA isolated from *Lyngbya majuscula* collected in Kahala Beach, Oahu (Edwards and Gerwick 2004). This cluster is composed of only 4 genes (Figure 4.2) and is the smallest characterized cluster from any *Lyngbya* strain to date. Lyngbyatoxin biosynthesis begins with the nonribosomal peptide synthetase

(NRPS) LtxA, which contains adenylation domains for L-valine and L-tryptophan, as well as an N-methyl transferase that methylates the nitrogen of the valine residue. Unusual NADPH dependent offloading from this enzyme (Read and Walsh 2007) yields the resulting dipeptide N-methyl-L-valyl-L-tryptophanol, which is subsequently cyclized by the P450 monooxygenase LtxB (Huynh et al. 2010) to give (-) indolactam V (ILV). The protein LtxC is a prenyltransferase that reverse prenylates the tryptophan residue with a geranyl group to produce the lyngbyatoxin A molecule (Edwards and Gerwick 2004). Lyngbyatoxins B and C (Figure 4.1) are thought to be analogs of the original lyngbyatoxin A molecule generated via the putative reductase/oxidase protein LtxD.

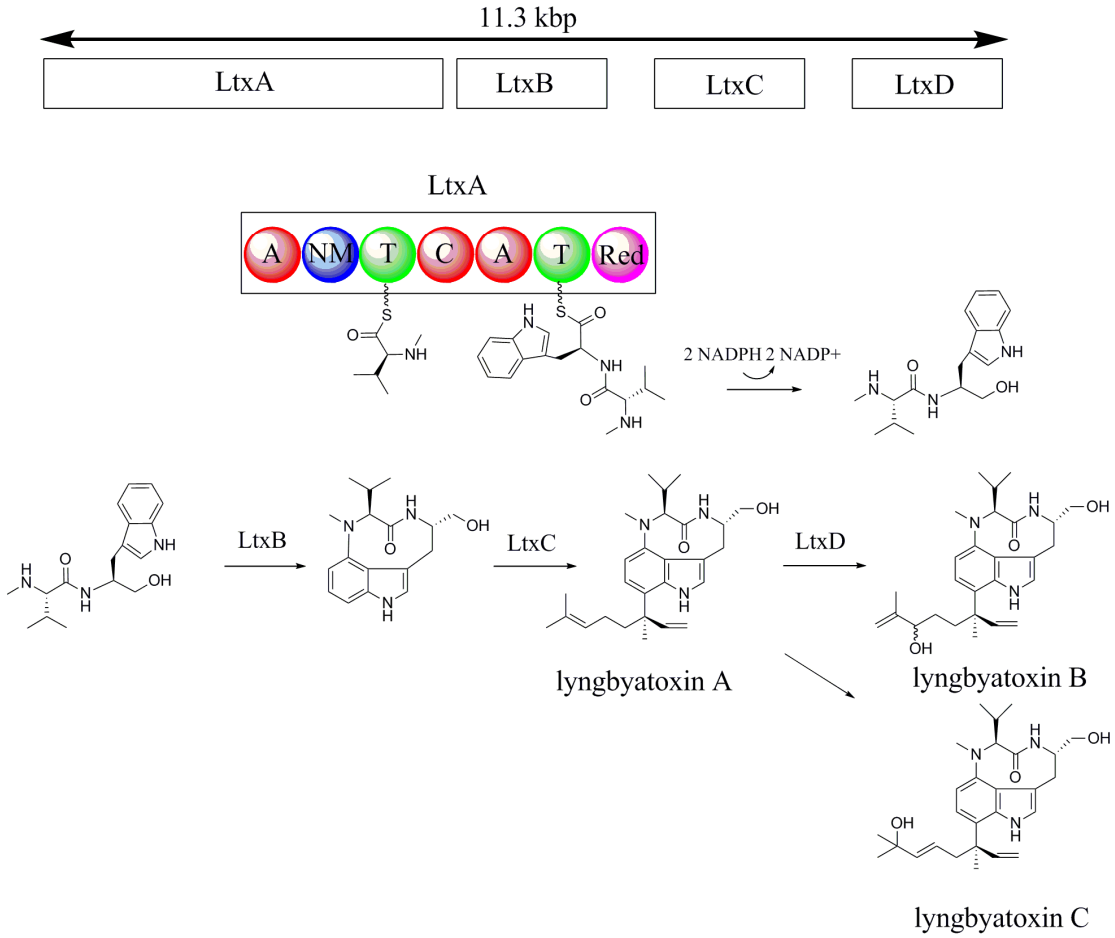


Figure 4.2: Biosynthesis of lyngbyatoxin A and predicted biosynthesis of lyngbyatoxins B and C. Figure adapted from Edwards and Gerwick (2004).

The small size of the lyngbyatoxin gene cluster (approximately 11.3 kb) combined with its availability on a single fosmid from the original pathway characterization efforts (Edwards and Gerwick 2004) enabled the use of this pathway and individual lyngbyatoxin genes in expression experiments with each of the three strains mentioned above. Predicted advantages and disadvantages of each strain in comparison to *Lyngbya* genetic characteristics are provided in Table 4.1. The results of experiments with *S. coelicolor* M512, *N. sp.* PCC7120, and *E. coli* BL21 are provided in detail below. Ultimately, *S. coelicolor* was found to be the most

promising candidate of the bacteria tested based on progress-to-date in both incorporation of genetic material and protein expression.

Table 4.1: Predicted advantages and disadvantages of three different bacterial strains in serving as heterologous hosts for the lyngbyatoxin gene cluster.

Heterologous strain	*Streptomyces coelicolor* M512	*Nostoc (Anabaena) sp.* PCC7120	*Escherichia coli* BL21
Advantages	-Sequenced genome -Natural product (PKS/NRPS) capable -Large suite of genetic techniques -Can handle large DNA transfer -Relatively fast growing	-Sequenced genome -NP capable -Similar GC content to ltx pathway -Some genetic techniques	-Sequenced genome -NP capable -Similar GC content to ltx pathway -Many genetic techniques -Very fast growing
Disadvantages	-High GC content (~70%) -Rare codon issues (TTA)	-Slow growing -Low probability of large DNA transfer -Histidine rich (could interfere with protein isolation)	-No native PPtase -Heterologous promoters likely needed

4.3 Results

4.3.1 Attempts to transfer DNA into *Lyngbya majuscula* filaments

Three strategies to introduce foreign DNA into *Lyngbya* filaments were attempted: 1) Incubation with non-specific, unlabeled or fluorescein labeled morpholinos, 2) conjugation using vectors designed for the filamentous cyanobacterium *Nostoc* (*Anabaena*) sp. PCC7120, and 3) biolistic (gene gun) introduction of DNA using tungsten beads.

Morpholinos are antisense oligomers that knock-down transcription of DNA (Summerton and Weller 1997), and are frequently used in eukaryotic developmental

experiments (Draper et al. 2001). Non-specific morpholinos (Gene-Tools LLC, L. Gerwick laboratory) that were unlabeled or labeled with fluorescein were incubated with *Lyngbya majuscula* JHB filaments to determine whether the morpholinos would be passively taken up by the filaments into the *Lyngbya* cells. After the incubation period, fluorescence microscopy revealed that while the fluorescein label was visible in the media surrounding the *L. majuscula* filaments, no label was detectable within the cells. MALDI analysis of some of these experiments successfully detected the morpholino oligomers when ionized independently, but no morpholinos were detected from filament extracts. The autofluorescence of pigments within the *Lyngbya* cells also obscured detection of the label. Due to this limitation and predicted difficulties with other labels, this method was not pursued further.

In preliminary attempts to introduce plasmid DNA into *L. majuscula*, two conjugation vectors commonly used in *Nostoc* sp. PCC7120 were chosen to determine whether the presence of either plasmid could be confirmed through antibiotic selection or GFP detection. The draft genome of *Lyngbya majuscula* 3L contains a *recA* gene, which indicates *Lyngbya* can undergo homologous recombination (Jones et al. 2011). The *E. coli* strain AM1359 (e.g., Wu et al. 2004) was obtained from the J. Golden laboratory at UCSD. This *E. coli* strain harbors the conjugal bridge plasmid pRL443 and the methylating plasmid pRL623. The cloning vector AM505 and the GFP reporter vector AM1954 (containing a GFP gene under the control of a vegetative promoter) were each separately introduced into AM1359 using electroporation, and colonies harboring either vector combination were grown and incubated with *L. majuscula* 3L or *L. majuscula* JHB filaments in 1.5 mL tubes before recovery and

antibiotic selection. Filaments used in conjugation attempts with AM505 were not able to persist over time in the presence of neomycin, and no GFP production was found in those filaments used in conjugation attempts with AM1954 (examined using specific wavelength parameters on J. Golden laboratory microscope).

Lyngbya filaments were also used in biolistic (gene gun) experiments in collaboration with the M. Hildebrand laboratory (SIO/UCSD) wherein DNA introduction was attempted by shooting DNA coated tungsten particles directly into dispersed filaments on BG-11 agar plates. The broad range cyanobacterial vector RSF1010 (Koksharova and Wolk 2002) was used in transformation attempts with *L. majuscula* JHB filaments, and following the attempted transformation, these filaments were placed in selective (ampicillin) SW BG-11 media after recovery. No significant survival of these filaments was observed after several days. Similar gene gun experiments testing RSF1010 and AM1954 (GFP reporter) vectors with *Nostoc* sp. PCC7120 did not yield any resistant colonies.

4.3.2 Development of heterologous expression strategies for *Lyngbya majuscula* natural products: Work with *Streptomyces coelicolor* M512

In light of the difficulties experienced with direct DNA transformation and conjugation into *Lyngbya* filaments, the focus of the work presented in this chapter shifted completely to experiments attempting to express the lyngbyatoxin gene cluster and portions thereof in other bacterial strains as discussed above. One of the most attractive candidates for this purpose was *Streptomyces coelicolor* M512. Actinobacteria such as *Streptomyces* strains are among the richest source of bacterial

natural products (Baltz 2008) and a wide range of genetic techniques are available for different expression goals (Kieser et al. 2000; Gust et al. 2003). *S. coelicolor* M512 (Floriano and Bibb 1996) is a mutant strain in which the activating proteins for prodiginine (red; *redD*) and actinorhodin (act; *actII-ORF4*) have been deleted and the SCP1 and SCP2 plasmids removed. The two anticipated limitations with using *Streptomyces* as an expression host for the lyngbyatoxin gene cluster (Table 4.1) are the large difference in GC content (*Streptomyces coelicolor* = 72%, Bentley et al. 2002; entire lyngbyatoxin pathway = 48%, Edwards and Gerwick 2004) and problems with codon usage, particularly the TTA codon encoding leucine. TTA codons in *Streptomyces* are rare, absent in any genes involved in vegetative growth (Leskiw et al. 1991), and found only in genes dedicated to regulation, cell differentiation, and secondary metabolism. *Streptomyces coelicolor* contains only one tRNA for the UUA codon (*bldA*), and *bldA* mutants (also referred to as "bald" mutants) are unable to produce antibiotics or form aerial mycelia or spores (Takano et al. 2003; Chater and Chandra 2008). Conversely, UUA is the most common leucine codon in the lyngbyatoxin pathway.

4.3.2.1 Heterologous expression of LtxC in *S. coelicolor* M512

In order to determine whether *S. coelicolor* could recognize individual genes from *Lyngbya*, the prenyltransferase encoding gene *ltxC* was selected for heterologous expression attempts. The first strategy entailed cloning the *ltxC* gene into the conjugal vector pSET152 (Bierman et al. 1992; Wilkinson et al. 2002; Figure 4.3; All primers provided in Table 4.2 in black text in materials and methods). The pSET152

backbone contains an integration cassette for site-directed conjugation into the *S. coelicolor* chromosome. This cassette contains an apramycin resistance gene, a φC31 integrase gene and an attP φC31 recognition site that corresponds to an attB site in *Streptomyces* chromosome during conjugation (Combes et al. 2002). The ermE* promoter involved in regulation of the erythromycin gene cluster (Bibb et al. 1994) was cloned in to precede the *ltxC* gene sequence. In addition, a *Streptomyces* Shine-Dalgarno sequence (Herai et al. 2004) was inserted directly before the *ltxC* start codon. Finally, a C-terminal His tag was engineered onto the 3' end of the *ltxC* gene, as a C-terminal His tag was also used in the original *E. coli* recombinant production of LtxC (Edwards and Gerwick 2004).

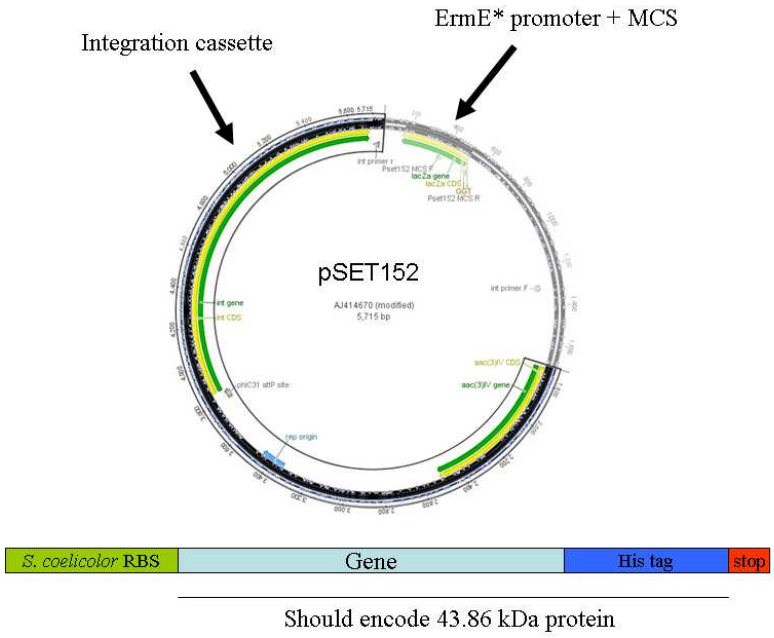


Figure 4.3: pSET152 vector (reference) and C-terminal His tagged *ltxC* construct.

Because *S.coelicolor* restricts methylated DNA from *E. coli*, this pSET152 construct was transformed into the *E. coli* strain ET12567 containing a *dam* mutation and the conjugal bridge plasmid pUZ8002 (Gust et al. 2003). A colony from this transformation was grown and used in a conjugation attempt with *S. coelicolor* spores (materials and methods). Exconjugants from this conjugation were screened with colony PCR to confirm the insertion of the *ltxC* gene before being grown for protein expression.

Preliminary attempts to determine whether *ltxC* was transcribed in *S. coelicolor* M512 were conducted using a separate pSET152 construct containing the ermE* promoter and *ltxC* sequence but without a His tag or *Streptomyces* Shine Dalgarno sequence. RT-PCR of the *ltxC* gene showed detectable transcripts after 4 days of growth in YEME media (materials and methods; Figure 4.4).

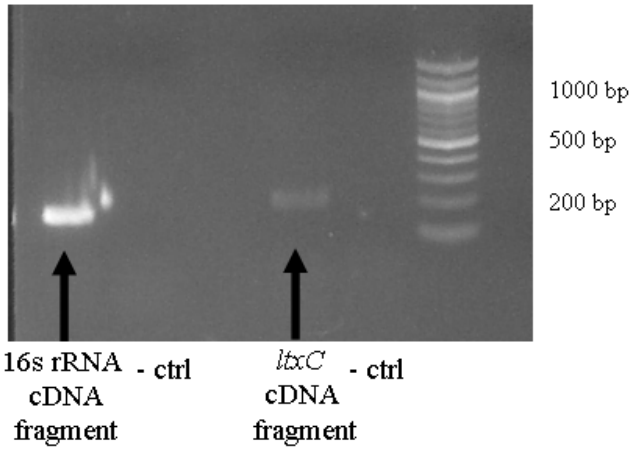


Figure 4.4: Preliminary RT-PCR experiments showing transcription of *ltxC*, as well as 16s rRNA, in *S. coelicolor* M512. Adjacent gel lanes to the right of each PCR product contain parallel reactions in which no Superscript enzyme was added.

Based on these RNA results, a TSB starter culture of the C-terminal His tagged *ltxC* construct was used to inoculate a larger YEME culture and grown for 4 days. Protein isolation efforts using nickel affinity revealed bands in the wash fractions consistent in size with LtxC (44 kDa); Figure 4.5), but not in the elution fractions. However, no bands were observed when screening for LtxC by Western blotting. These results may suggest that the ermE* promoter is not strong enough for LtxC to be produced to detectable levels, so other expression vectors were considered.

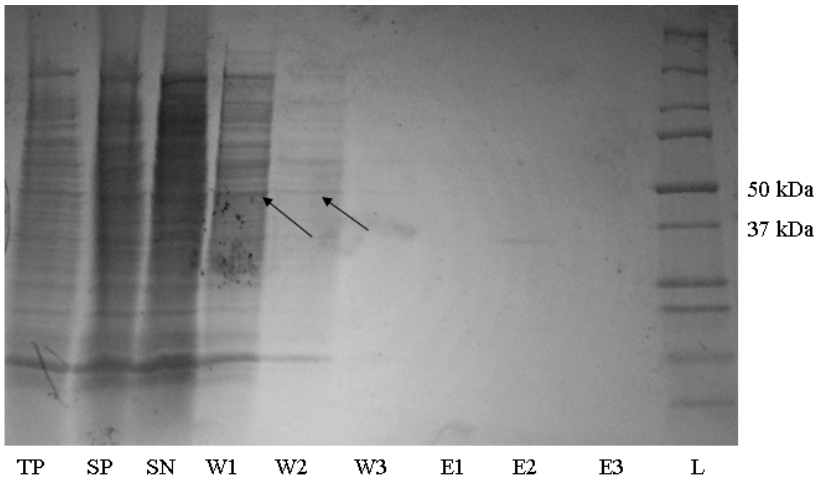


Figure 4.5: Protein bands from ermE* *ltxC* pSET152 construct expression (C-terminal His tag observed in wash fractions. None of these possible proteins were visible by Western blot.

An alternative strategy to overexpress *ltxC* was pursued by using the self replicating, inducible vector pSH19 (Herai et al. 2004). The pSH19 vector contains a regulatory gene used in nitrilase production (NitR) as well as nitrilase specific promoter regions (Figure 4.6).

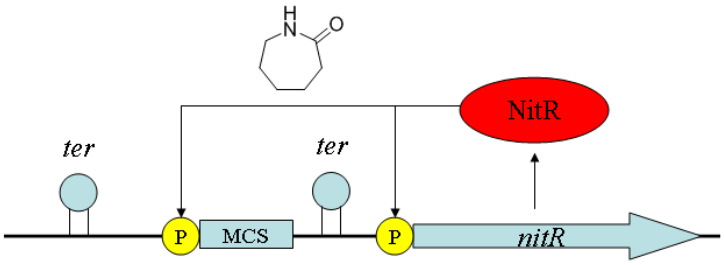
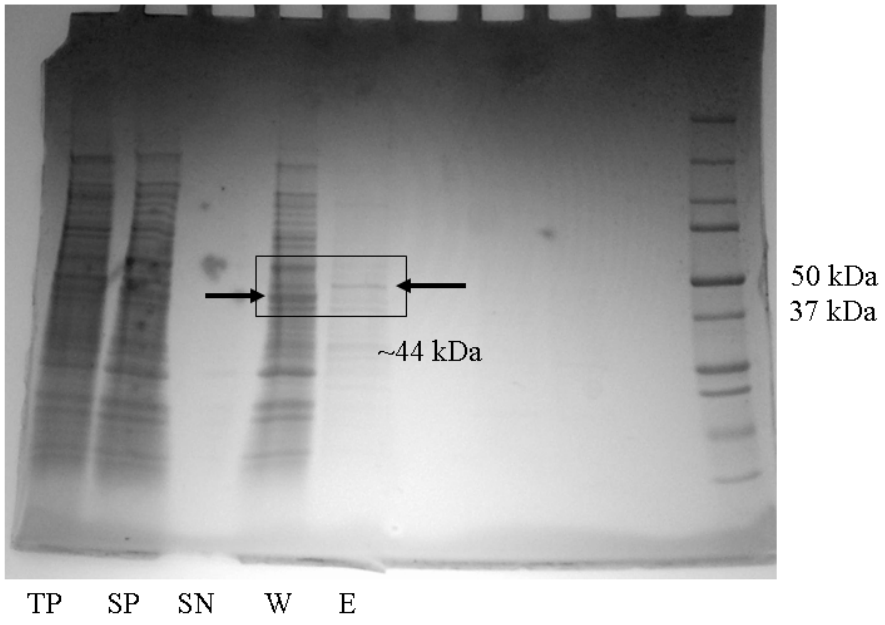


Figure 4.6: Hyper-inducible expression system in *Streptomyces* vector pSH19 (Figure adapted from Herai et al. 2004). The NitR regulatory protein is activated by ε-caprolactam, and activated NitR interacts with promoters in front of the vector multiple cloning site as well as the *nitR* gene, thereby rapidly increasing the number of *nitR* transcripts to produce more NitR protein.

This promoter is strongly upregulated in the presence of ε-caprolactam, and this regulatory system is used industrially in the production of acrylamide (Kobayashi and Shimizu 1998). The *ltxC* gene sequence was cloned into a modified pSH19 vector that contained an *E. coli oriT* (ColE1) inserted into the SpeI site (along with a *bla* gene), enabling *E. coli* cloning (A. Eustaquio, B. Moore Laboratory). The *ltxC* gene was prepared in the same manner as pSET152, with the inclusion of a *Streptomyces* Shine-Dalgarno sequence and a C-terminal His tag. Because the pSH19 vector is self-replicating, this construct was introduced into *Streptomyces* using protoplasts and thiostrepton selection (*via* the *E. coli* strain ET12567 without the pUZ8002 conjugal plasmid). Based on the recommended protocol (Herai et al. 2004), YEME cultures of this pSH19 construct (inoculated with TSB starter cultures) were induced with ε-caprolactam after 96 h, harvested at 120 h, and used in protein purification efforts. Despite protein bands consistent with LtxC in size (Figure 4.7a), once again, no bands appeared by Western blotting.

A



B

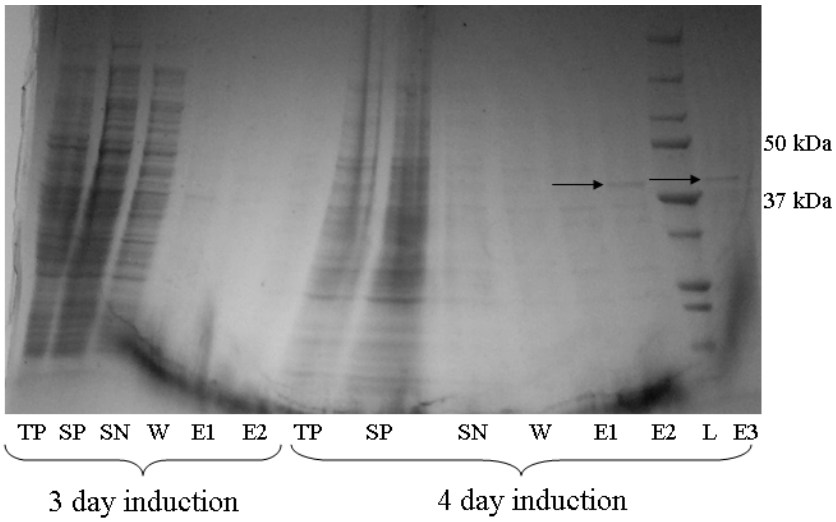


Figure 4.7: LtxC protein expression attempts using the pSH19 vector. A) Bands consistent with C-terminal his tagged LtxC on SDS PAGE 4-20% gel. No corresponding band was observed by Western blotting. B) Protein isolation of N-terminal his tagged LtxC in pSH19 on SDS PAGE 4-20% gel induced on day 3 or day 4 (Total protein, soluble protein, nickel incubation supernatant, wash, and elution fractions shown). Note protein bands (arrows) at approximately 44 kDa.

C

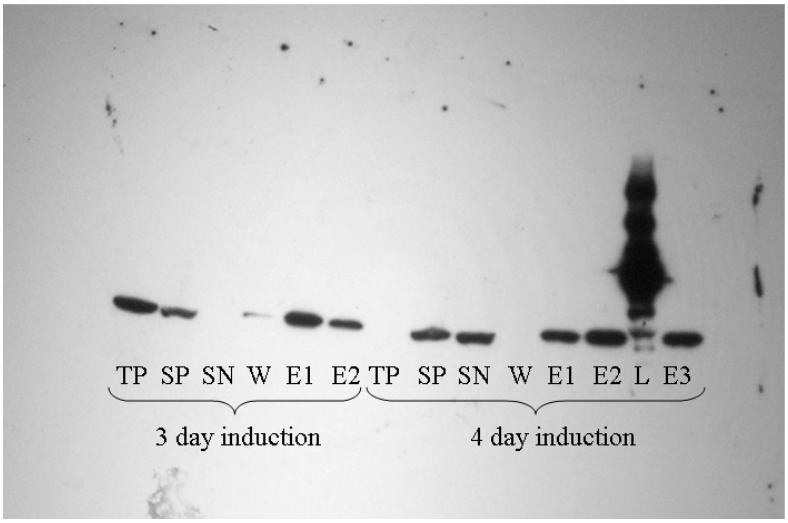


Figure 4.7 (continued): C) Western blot of gel shown in 4.7B. Bands correspond to purified N-His-LtxC.

A separate pSH19 *ltxC* construct was made with an N-terminal His tag immediately following the Shine Dalgarno site (Herai et al. 2004). *S. coelicolor* M512 harboring this construct was grown and induced as above. After protein purification, a band consistent with LtxC was present in the protein extract as shown by colloidal Coomassie staining (Figure 4.7b) that was also present by Western blotting using anti-His-HRP (Figure 4.7c), confirming successful LtxC expression. This experiment was repeated using induced and non-induced cultures, which showed that LtxC expression in pSH19 is dependent on ε-caprolactam being added to the culture medium (Figure 4.8). A separate culture was grown to a larger scale to obtain more soluble protein for purification (Figure 4.8).

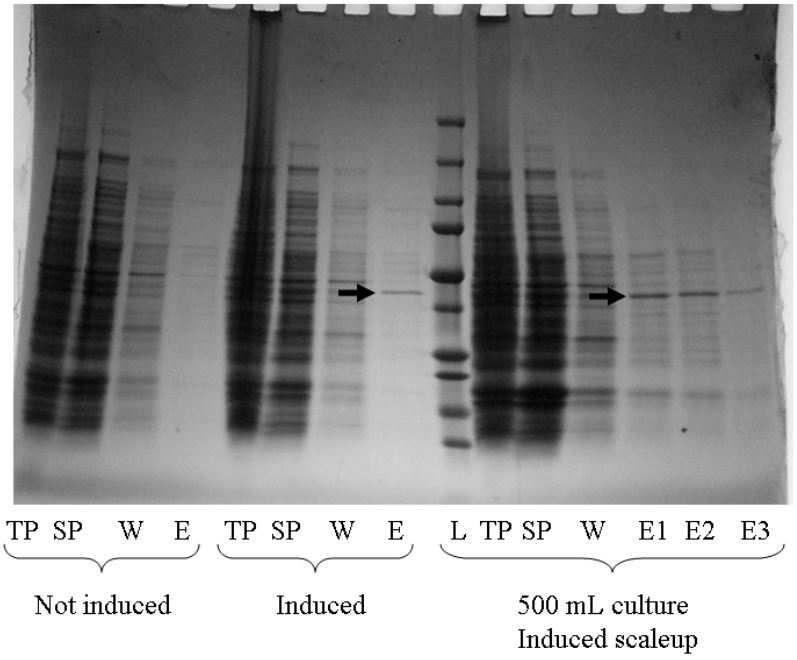
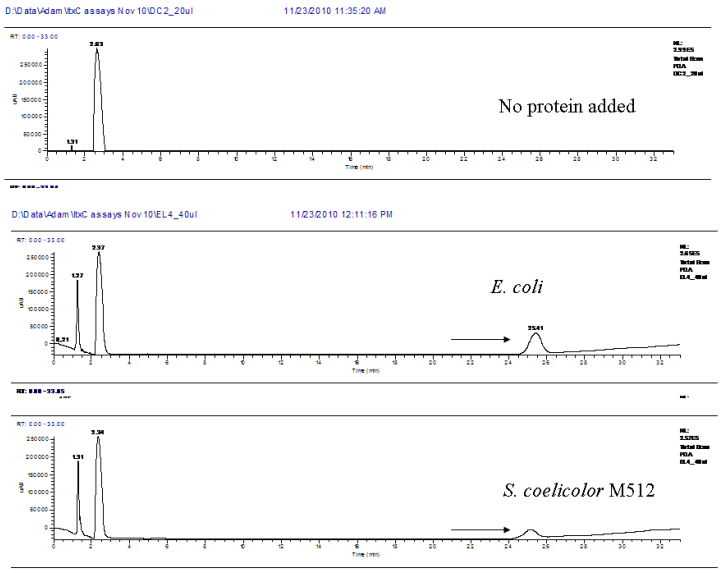


Figure 4.8: Comparison of non-induced and induced pSH19 vectors with the N-terminal his tagged *ltxC* insert (arrows indicate eluted protein), as well as a larger scale (500 mL) purification.

Purified LtxC from *S. coelicolor* M512 was used to repeat the original assays confirming its predicted prenylation activity (Edwards and Gerwick 2004). LtxC was incubated with (-) Indolactam V and geranyl pyrophosphate for 1 h, and ethyl acetate extracts of this reaction were profiled by LC-MS along with parallel reactions conducted with LtxC protein overexpressed in *E. coli* (Edwards and Gerwick 2004). Recombinant LtxC from both hosts was able to catalyze the addition of a geranyl group to ILV to form lyngbyatoxin A (Figure 4.9a). To confirm that LtxC from *S. coelicolor* was responsible for this activity, protein elutions from non-induced pSH19 *ltxC* cultures were also used in this experiment, and no lyngbyatoxin A was detected (Figure 4.9b). Expression of this protein was also attempted using other media (GYM, Shima et al. 1996), but no protein was produced, suggesting the high sucrose component (34%) of the YEME media is important for pSH19 to function correctly. Sucrose is thought to possibly function in osmotic pressure balance between culture media and cytoplasm, or may induce or activate certain enzyme systems (Elibol & Mavituna 1998).

A



B

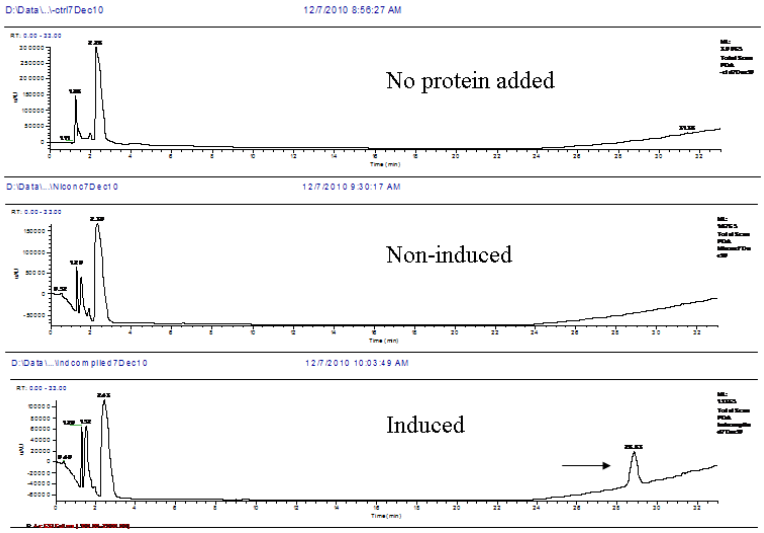


Figure 4.9: LC-MS PDA chromatograms of extracts from LtxC prenylation assays. A) Comparison of reactions with no protein added, recombinant LtxC from *E. coli*, and recombinant LtxC from *S. coelicolor* M512. Production of lyngbyatoxin A is noted with arrow. B) Comparison of reactions with no protein added, protein extract from non-induced pSH19 N-terminal his *ltxC* construct, and induced pSH19 N-terminal his *ltxC* construct. Production of lyngbyatoxin A is noted with arrow.

C

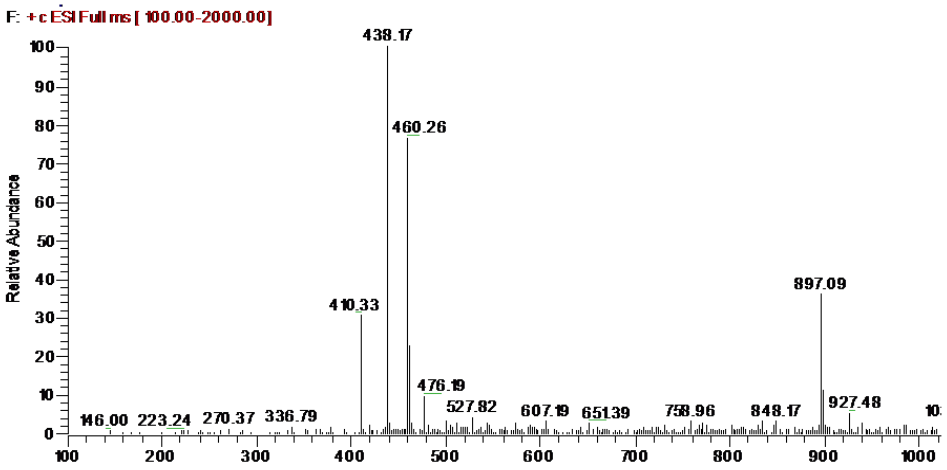
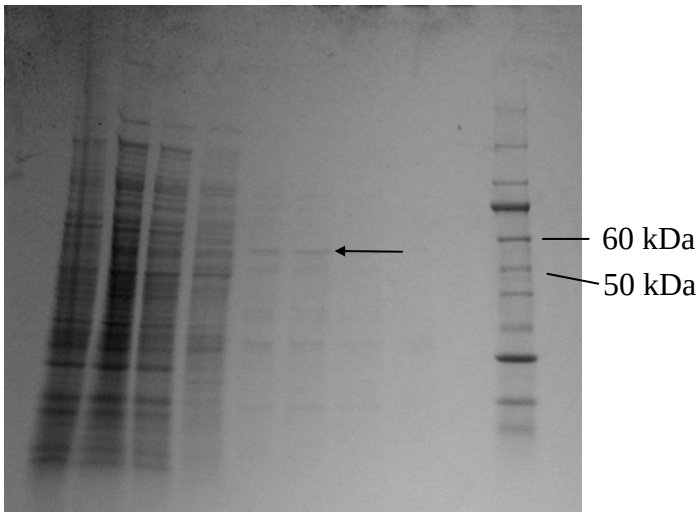


Figure 4.9 (continued): C) Ionization of lyngbyatoxin A (ESI-MS) from ε-caprolactam induced culture shown in Figure 4.8B.

Subsequent expression attempts were performed using the other proteins from the lyngbyatoxin pathway. An N-terminal His tagged *ltxB* pSH19 construct was successfully used in the overexpression of LtxB (Figure 4.10a, b). The large NRPS ORF *ltxA* (7.5 kb) was also cloned into pSH19 with a *S. coelicolor* SD site, both with and without an N-terminal His tag. Protein isolation experiments using the His tagged version were not successful in purifying the entire protein (275 kDa) by nickel affinity methods. A large band that appeared in large scale efforts (Figure 4.11) was submitted for LC-MS/MS analysis, but peptides from trypsin digest were consistent with *Streptomyces* NRPS proteins instead of LtxA. The non-His tagged version of *ltxA* was grown and induced as above, and this culture was extracted with ethyl acetate

and run over C-18 solid phase extraction cartridges. LC-MS profiles of this extract did not reveal any production of the N-methyl-L-valyl-L-tryptophanol normally biosynthesized by LtxA.

A



B

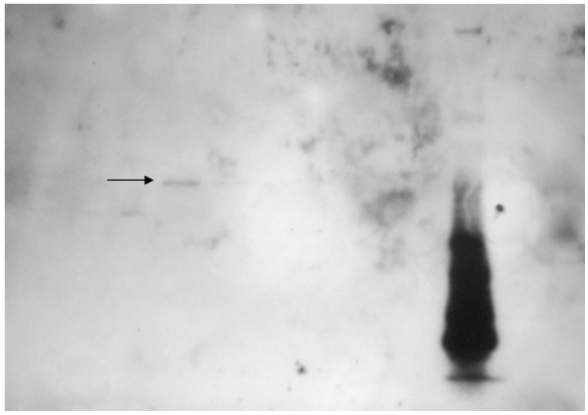


Figure 4.10: Expression of LtxB using the pSH19 vector. A) Expression of N-His-LtxB in pSH19. B) Expression was confirmed with Western blotting.

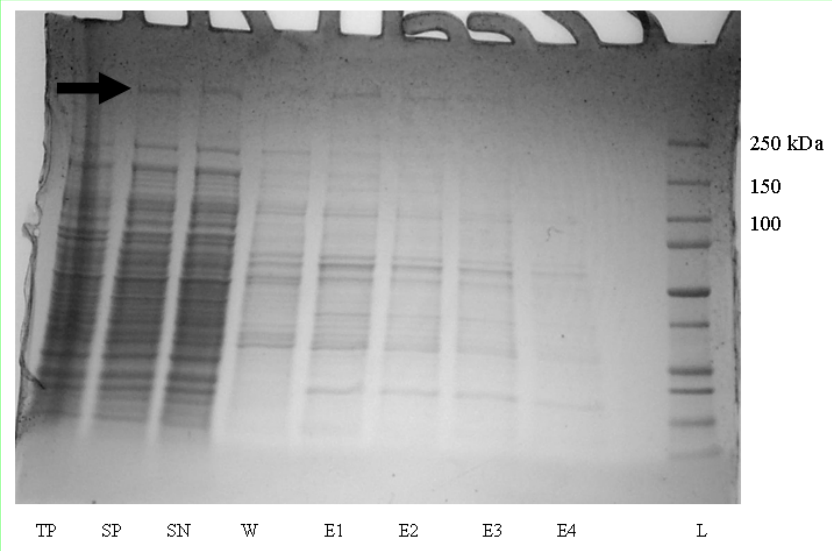


Figure 4.11: Attempt to express N-His-LtxA in pSH19. Large (275 kDa) band (see arrow) was excised and analyzed by LC-MS/MS, but was found to be consistent with a *Streptomyces* NRPS instead of LtxA. No bands were observed by Western blotting.

4.3.2.2. Attempts to heterologously express the lyngbyatoxin gene cluster in *Streptomyces*

Initial attempts to insert the lyngbyatoxin pathway into *S. coelicolor* M512 were attempted by amplifying *ltxA-D* using long-range PCR (Figure 4.12) and cloning this fragment into pSET152 containing the ermE* promoter. Ligation attempts with this DNA fragment were unsuccessful.

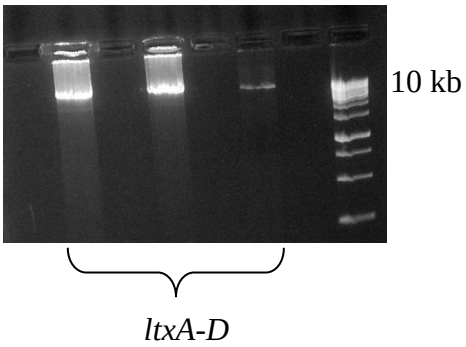
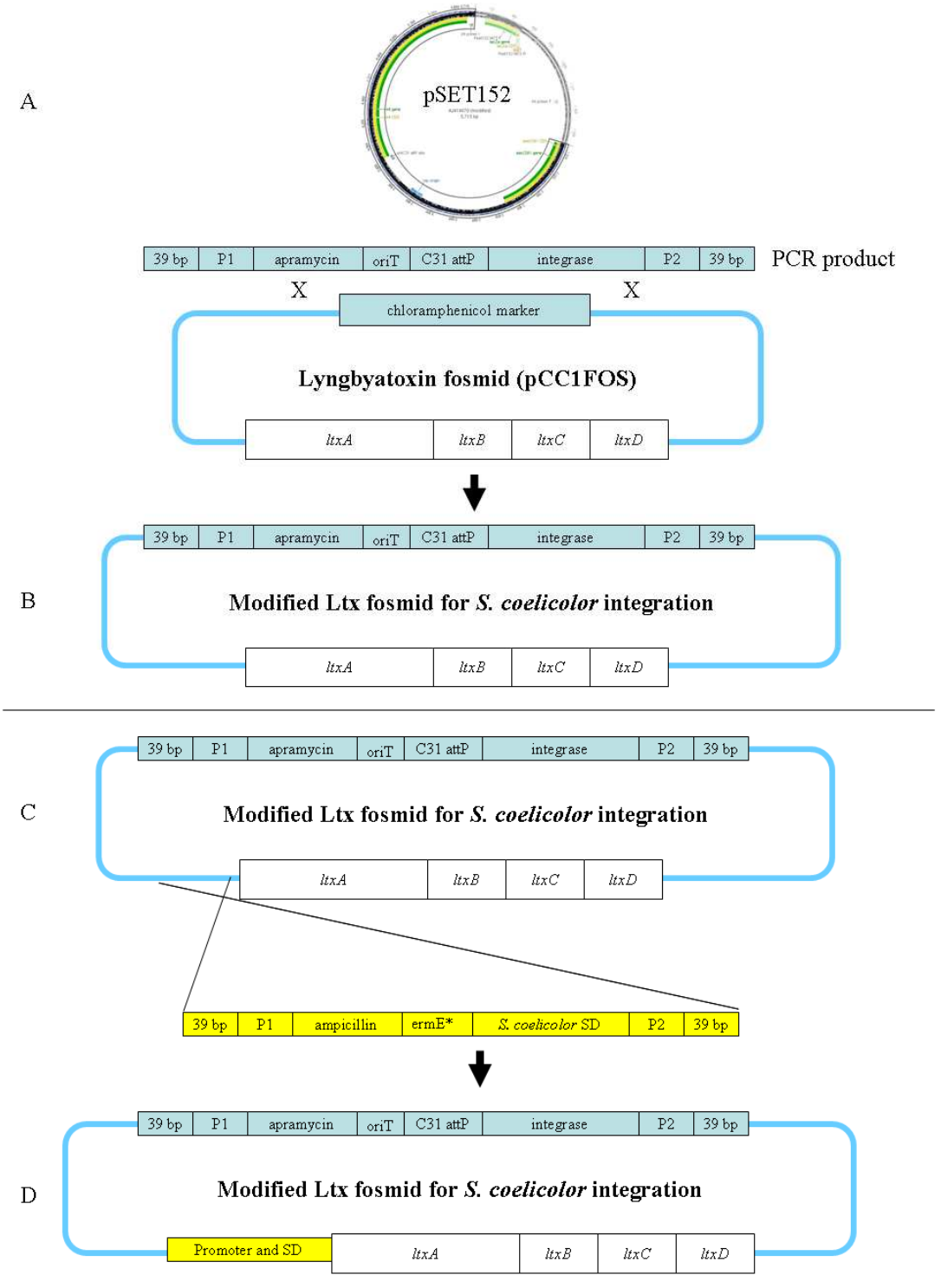


Figure 4.12: Long range PCR amplification of *ltxA-D*. Cloning of this large product was not successful.

Next, in order to utilize the integration cassette portion of the pSET152 backbone in an effort to modify the native (pCC1FOS, Epicentre) fosmid (DE3-86) harboring the lyngbyatoxin pathway (Edwards and Gerwick 2004), this cassette (4.5 kb) was amplified by PCR using primers including 39 bp of homology with the DNA regions surrounding the chloramphenicol resistance gene of DE3-86 and PCR purified (Figure 4.13a, b). The chloramphenicol resistant DE3-86 fosmid was transformed via electroporation into the *E. coli* strain BW25113, which in this case harbored the λ–RED phage vector pKD20 (Gust et al. 2003) for homologous recombination. Colonies from this transformation were grown, made electrocompetent, and subsequently transformed with the purified pSET152 integration cassette. Successful replacement of the chloramphenicol cassette in DE3-86 was confirmed by sequencing. The region immediately upstream of *ltxA* in the DE3-86 fosmid was then replaced in the same manner using the λ–RED phage vector pIJ790 (Gust et al. 2003) with another cassette consisting of an ampicillin marker (pBluescript II SK-), the ermE* promoter, and a *Streptomyces* Shine Dalgarno site (Figure 4.13c, d; Herai et al. 2004). Severe difficulty with these homologous recombination procedures (especially recombination

of the pSET152 integrase cassette) was encountered due to repeated contamination problems with pSET152 and selection of false positive colonies (complications resulted in these procedures taking several months of repeated attempts to accomplish). Upon generating the desired constructs, the modified DE3-86 fosmids (either with or without the ermE* promoter cassette) were used in conjugation efforts with *S. coelicolor* M512 spores, using either the ET12567/pUZ8002 or S17 (Mazodier et al. 1989) *E. coli* conjugation strains. Despite approximately 10 conjugation attempts, no exconjugants were able to grow under apramycin selection, which suggested some inherent incompatibility between the conjugal bridge plasmids of these strains and the DE3-86 fosmid. Separate attempts to introduce these constructs using protoplasts as with pSH19 were also unsuccessful.

Figure 4.13: Homologous recombination strategy for engineering the native lyngbyatoxin fosmid for conjugation and expression in *S. coelicolor* M512 (Gust et al. 2003). A) The integrase cassette from the pSET152 vector is amplified by PCR using 39 bp extension primers homologous to the regions flanking the chloramphenicol resistance gene in the native lyngbyatoxin vector. B) This cassette replaces the chloramphenicol gene via homologous recombination in the *E. coli* strain BW25113/pKD20. C) An additional cassette composed of an ampicillin resistance gene, the ermE* promoter, and a *S. coelicolor* Shine Dalgarno site (Herai et al. 2004) is amplified by PCR using 39 bp extension primers homologous to the upstream region of *ltxA*. D) This cassette replaces the *ltxA* promoter region via homologous recombination in the *E. coli* strain BW25113/pIJ790.


A
pSET152
39 bp
P1
apramycin
oriT
C31 attP
integrase
P2
39 bp
PCR product
X
X
chloramphenicol marker
Lyngbyatoxin fosmid (pCC1FOS)
ltxA
ltxB
ltxC
ltxD
B
Modified Ltx fosmid for S. coelicolor integration
C
Modified Ltx fosmid for S. coelicolor integration
ampicillin
ermE*
S. coelicolor SD
D
Modified Ltx fosmid for S. coelicolor integration
Promoter and SD

Due to lack of success with both cloning the ltx pathway into pSET152 and homologous recombination to equip the DE3-86 fosmid for conjugation, the pSH19 vector was selected for pathway cloning attempts. The genes necessary for lyngbyatoxin A production (*ltxA-C*) were cloned in two steps (Figure 4.14; the entire *ltxA* gene in step 1 and *ltxB-C* in step 2), which included placement of a *Streptomyces* Shine-Dalgarno site in front of *ltxA* and *ltxB*. The native intergenic region upstream of *ltxC* was retained to determine whether a *Lyngbya* ribosome binding site could be recognized in *S. coelicolor* (if this portion of the transcript was not translated it was predicted that ILV could still be produced enzymatically by LtxA and B).

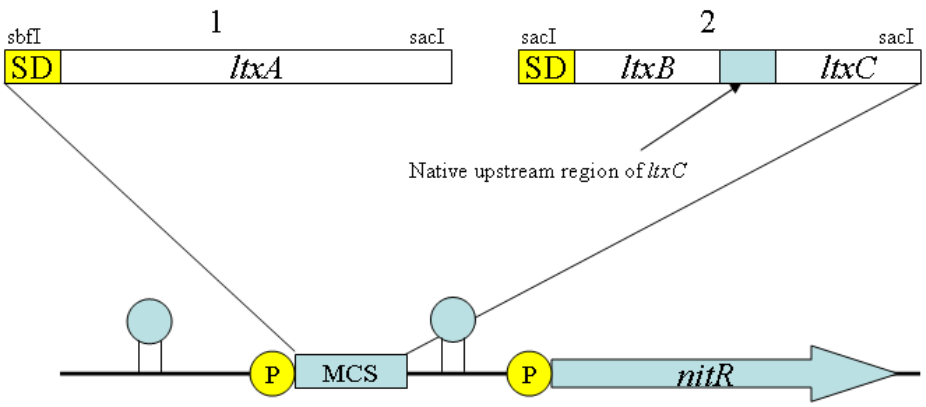


Figure 4.14: First cloning strategy for lyngbyatoxin A gene cluster into pSH19 (Herai et al. 2004). *LtxA* was PCR amplified including a *Streptomyces* Shine Dalgarno sequence and ligated into the vector multiple cloning site at SbfI and SacI sites. A second PCR product including *ltxB* and *ltxC* (adding a *Streptomyces* SD sequence before *ltxB* and retaining the native upstream region of *ltxC*) was ligated into the SacI site downstream of *ltxA* after a second vector digest.

This construct was transformed into *S. coelicolor* protoplasts and grown in starter TSB cultures, which were then used to inoculate a variety of media for lyngbyatoxin A

production. These larger cultures were extracted using either XAD-7 resin (Amberlite) or ethyl acetate, in some cases run over C-18 cartridges, and then profiled by LC-MS. No lyngbyatoxin A was detected in any of these cultures. Followup RT-PCR experiments revealed that transcripts from this construct were detectable using YEME media, but that transcript levels decreased from *ltxA* to *ltxC* (Figure 4.15).

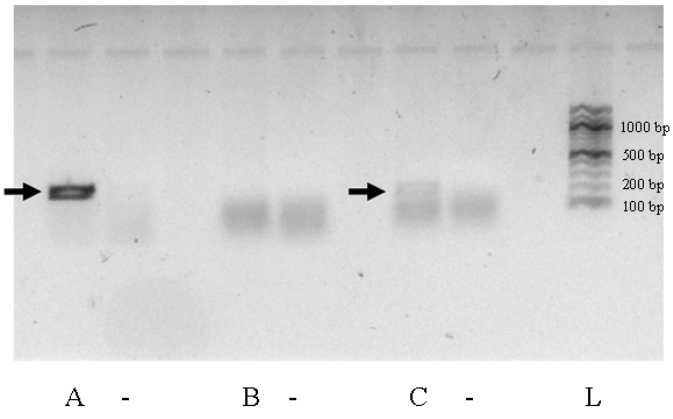


Figure 4.15: RT-PCR of portions of the *ltxA-C* construct in pSH19. Fragments of *ltxA*, *ltxB*, and *ltxC* were amplified from first strand cDNA alongside parallel reactions with no Superscript reverse transcriptase added. Strong transcription is noted for *ltxA* but is not visible for *ltxB* and faint for *ltxC*. No transcripts were found when growing this construct in GYM media.

To verify that lyngbyatoxin A could be detected using these extraction methods, a small culture aliquot (~10 mL) was supplemented with 100 μg of lyngbyatoxin A standard before being extracted. When profiled on LC-MS lyngbyatoxin A was clearly present in this extract (Figure 4.16).

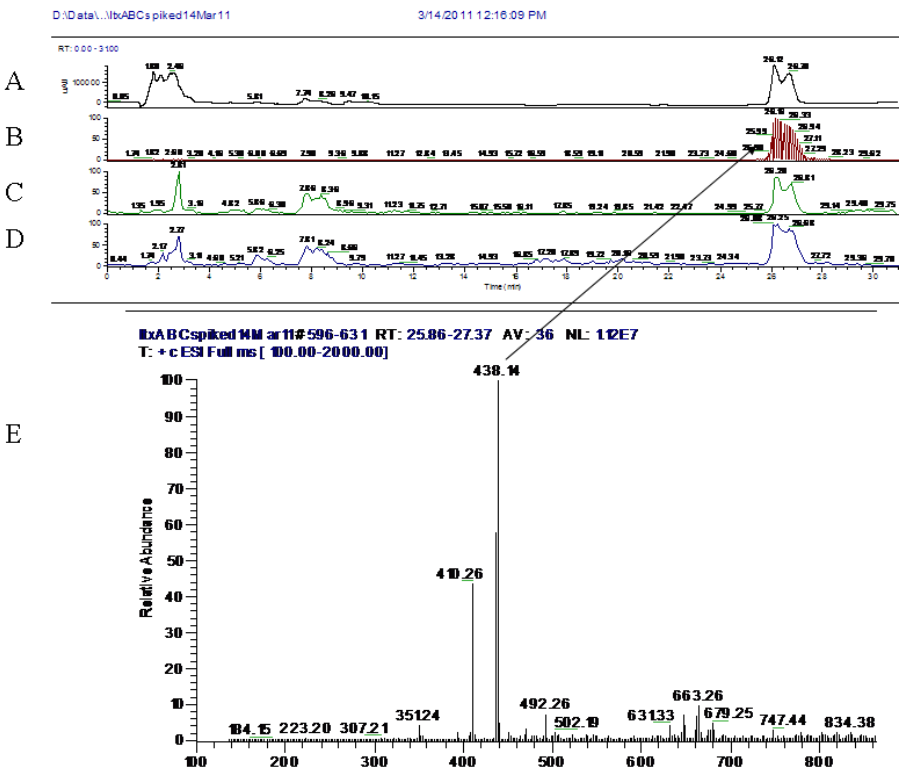


Figure 4.16: Extract of 10 mL YEME *ltxABC* pSH19 culture aliquot (construct shown in Figure 4.13) supplemented with 100 µg lyngbyatoxin A. A) PDA chromatogram from extract. B) MS ion selection (*m/z* 435-440). C) Negative ion selection (*m/z* 100 – 2000). D) Postive ion selection (*m/z* 100-2000). E). Ionization of lyngbyatoxin A. Arrow indicates location of lyngbyatoxin A during LC-MS gradient run.

A combined YEME culture was inoculated with starter TSB cultures of SD-*ltxA*, His-SD-*ltxB*, and His-SD-*ltxC* under thiostrepton selection to determine whether independent expression of each protein would be successful in biosynthetic assembly of lyngbyatoxin A. This culture was induced after 96 h, and was supplemented with approximately 50 mg L-tryptophan, 50 mg L-valine, and 100 mg geraniol. After 144 h this culture was extracted and profiled by LC-MS. No lyngbyatoxin A was observed in this extract (Figure 4.17).

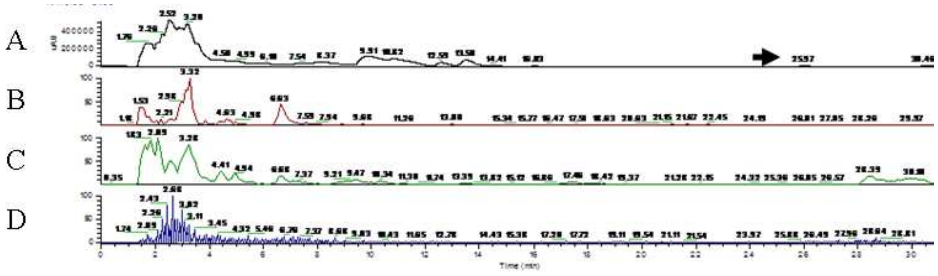


Figure 4.17: LC-MS profile of YEME co-culture inoculated from separate *ltxA*, *ltxB*, and *ltxC* pSH19 starter cultures. This culture was induced at 96 h with ε-caprolactam (0.1% w/v) and supplemented with 50 mg L-valine, 50 mg L-tryptophan, and 100 mg geraniol. The culture was extracted at 144 hr. A) PDA chromatogram from extract. The arrow points to a small peak with a similar retention time to the lyngbyatoxin A standard, but the *m/z* value of this peak is not consistent with lyngbyatoxin A. B) Negative ion selection (*m/z* 100 – 2000). C) Postive ion selection (*m/z* 100-2000). D) MS ion selection (*m/z* 435-440).

4.3.3 Heterologous expression of the lyngbyatoxin gene cluster: Work with *Nostoc* sp. PCC7120

The heterocyst forming, filamentous cyanobacterium *Nostoc* (*Anabaena*) sp. PCC7120 was considered as another viable candidate for heterologous expression of the lyngbyatoxin pathway. As a cyanobacterium, the genetic characteristics of this strain should be closer to *Lyngbya* than *Streptomyces*, and the GC content of *N*. sp. PCC7120 is 41%, which is closer in proximity to the lyngbyatoxin gene cluster (48%). Some genetic techniques for this strain are available (i.e., Yoon and Golden 2001, Koksharova and Wolk 2002) and this approach also facilitated a working collaboration with the J. Golden laboratory (UCSD).

4.3.3.1. Attempts to heterologously express *ltxC* in *Nostoc* sp. PCC7120

In order to insert *ltxC* into the *Nostoc* chromosome, a C-terminal His tag version of *ltxC* gene was cloned into the AM505 vector, a common conjugation vector used by the Golden laboratory. The majority of this vector backbone (Figure 4.18) is composed of sequence homologous to a region of the *Nostoc* genome. This vector also contained an RNA polymerase promoter (RNAP; Figure 4.18).

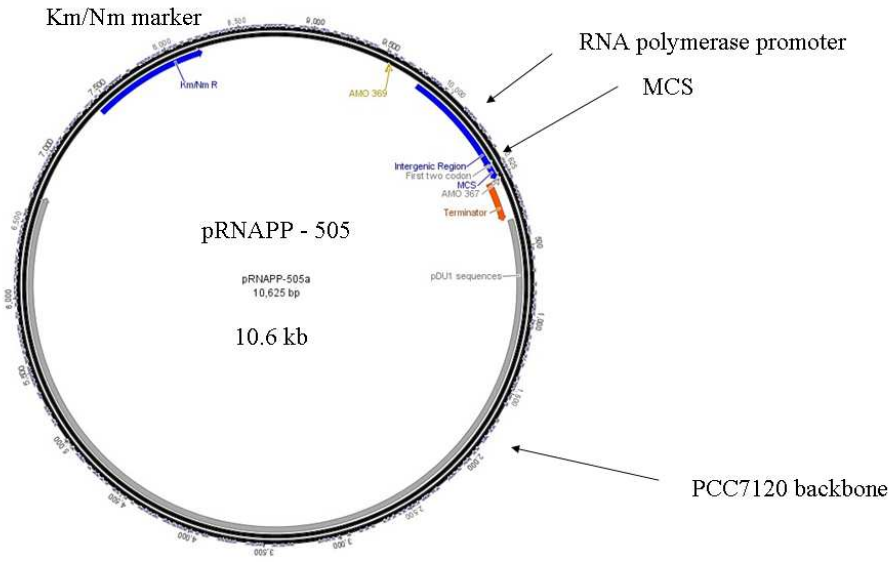


Figure 4.18: AM505 *Nostoc* sp. PCC7120 conjugation vector including the RNA polymerase promoter.

Each construct was transformed into the AM1359 *E. coli* strain, and colonies from this transformation were grown and used in conjugation attempts with *Nostoc* cultures. Exconjugants were found after approximately 10 days, but these exconjugants were unable to grow in liquid FW BG-11 media, suggesting they were false positives. These conjugation efforts were repeated several times without success.

4.3.3.2 Attempts to heterologously express the lyngbyatoxin gene cluster in *Nostoc* sp. PCC7120

Parallel homologous recombination strategies to those used in modifying the DE3-86 fosmid as described for *S. coelicolor* were used to prepare this fosmid for conjugation attempts with *Nostoc* sp. PCC7120. A gene cassette composed of a kanamycin resistance gene and a portion of a *Nostoc* gas vesicle protein gene (sufficient for recombination into the *Nostoc* genome; S. Saha, personal communication) was generated with PCR using primers with 39 bp extensions homologous to the region surrounding the chloramphenicol gene in DE3-86 (Figure 4.13) and PCR purified. This cassette was transformed into electrocompetent BW25113 cells containing the DE3-86 fosmid and λ–RED phage vector pKD20 as described above (Gust et al. 2003) and colonies from this transformation were screened for correct replacement of the chloramphenicol gene in the DE3-86 fosmid backbone using PCR (attempts to sequence this modified fosmid were unsuccessful).

Several attempts were made to introduce this modified fosmid into *Nostoc* via conjugation. Exconjugants were observed in each instance but none of these were able to grow in liquid BG-11 media to pursue further. Discussions with the Golden laboratory (S. Saha) led to the conclusion that this construct may be too large for successful recombination into the *Nostoc* chromosome. Because of these limitations, additional work with *Nostoc* sp. PCC7120 was not pursued.

4.3.4 Attempts to heterologously express the lyngbyatoxin gene cluster in *E. coli* BL21

A number of NRPS based natural products have been successfully expressed in *E. coli* (Pfeifer et al. 2003; Watanabe et al. 2006). Initial attempts were made by Dan Edwards to express the lyngbyatoxin pathway using the original DE3-86 fosmid and a *Bacillus* Sfp (phosphopantetheinyl transferase; Quadri et al. 1998) construct under control of a T7 promoter, but this effort was unsuccessful. In order to further modify the DE3-86 cassette for *E. coli* expression, a gene cassette consisting of an ampicillin resistance gene and a T7 promoter (Figure 4.19) was generated by PCR with 39 bp extension primers homologous to the region immediately upstream of the *ltxA* gene. This cassette was introduced into the DE3-86 fosmid containing the apramycin gene cassette from earlier attempts with *S. coelicolor* due to limitations in available antibiotic selection choices (enabling the use of the chloramphenicol resistant pIJ790 λ–RED phage vector) using the same homologous recombination methods as above. Successful completion of this construct was confirmed by sequencing.

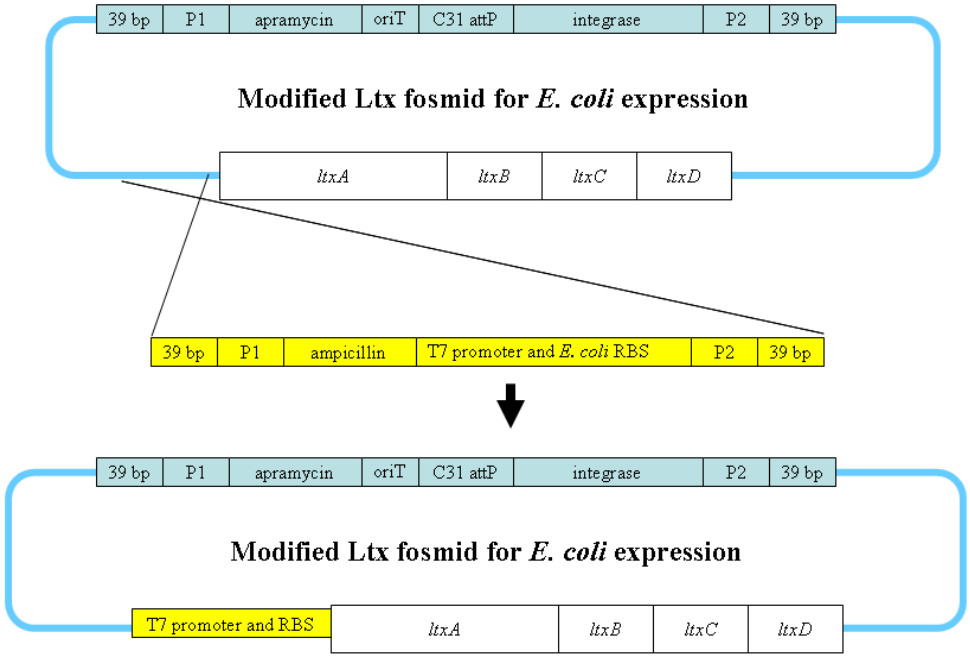


Figure 4.19: Insertion of T7 promoter and *E. coli* ribosome binding site into modified lyngbyatoxin fosmid by homologous recombination in *E. coli* BW25113/pIJ790.

This modified DE3-86 construct was transformed into the *E. coli* expression strain BL21 along with a Pet28 vector containing the *Bacillus* gene encoding Sfp. A colony from this transformation was grown and used to innoculate LB media for expression attempts. After IPTG induction of this larger culture and profiling using LC-MS and MALDI analysis, no lyngbyatoxin A was detected.

4.4 Discussion

The absence of genetic techniques and lack of a heterologous expression system for filamentous marine cyanobacteria are becoming more conspicuous in light of substantial advances made in these areas over the last 10 years with other natural product producing bacteria and fungi (Galm and Shen 2006; Wilkinson and

Micklefield 2007; Gressler et al. 2011). The work presented in this chapter was aimed at method development to ultimately improve genetic access to *Lyngbya* natural products and enable the large scale production of compounds of interest to drug discovery and biotechnology efforts. Due to inherent limitations in *Lyngbya* morphology and physiology and based on the problems encountered while attempting to introduce foreign DNA into *Lyngbya* cells, it is evident that shuttling *Lyngbya* DNA to a more amenable heterologous host appears to be a more tractable approach to achieving these goals. Genetic mutations of *Lyngbya* strains may be possible in the future, but will probably require specific attention to DNA stability upon introduction because of the slow doubling rate of *Lyngbya* filaments (Esquenazi et al. 2011). Sequencing of additional *Lyngbya* strains and completion of *Lyngbya* genomes (Jones et al. 2011) will undoubtedly reveal more genetic information that can be used in optimizing strategies appropriate for these cyanobacteria and potentially avoiding problems with associated heterotrophic bacteria.

Heterologous expression is one of the most promising approaches to access large amounts of desired natural products and to generate novel analogs through gene knockouts, replacement, or precursor directed mutasynthesis (Wilkinson and Micklefield 2007; Heide 2009). Use of *Escherichia coli* as a heterologous host for natural product pathways is becoming more common (Pfeifer et al. 2003; Watanabe et al. 2006) and may be an attractive option for *Lyngbya* compounds derived from larger gene clusters. Attempts to express the lyngbyatoxin pathway using the T7 promoter in *E. coli* BL21 while also expressing a recombinant Sfp enzyme were not successful, but more fine tuning of culture conditions may be needed to enable proper translation and

function of this gene cluster. *Streptomyces coelicolor* and *Nostoc* (*Anabaena*) sp. PCC7120 were pursued in this chapter to a greater extent primarily because each strain is already "natural product capable," and *Lyngbya* DNA may require less manipulation before introduction into either of these strains. Unfortunately, neither *ltxC* nor the entire lyngbyatoxin pathway was accommodated by *Nostoc* sp. PCC7120 in conjugation attempts, despite the use of different vectors (materials and methods). Another cyanobacterial strain that could serve in heterologous expression of *Lyngbya* metabolites is *Synechocystis* sp. PCC6803 (Roberts et al. 2009). *Synechocystis* does not produce any NRPS – or PKS-type natural products, despite having a native phosphopantetheinylase (PPTase). This PPTase was found to activate the *Synechocystis* native cognate fatty acid synthesis protein but not other cyanobacterial carrier proteins. However, introduction of the *Nodularia spumigena* NSOR10 PPTase into *Synechocystis* conferred the ability to activate a non-cognate carrier protein (Roberts et al. 2009). The ability to naturally transform this *Synechocystis* strain and its fast doubling times may allow for the difficulties experienced with *Nostoc* sp. PCC7120 to be avoided. This approach may be pursued in the future in collaboration with the Brett Neilan laboratory (University of New South Wales, Sydney, Australia).

Genetic manipulations and heterologous expression in *Streptomyces* strains are reliable and well documented (Kieser et al. 2000; Eustaquio et al. 2005; Flinspach et al. 2010) and, if the GC% content and related codon recognition issues are surmountable, actinobacteria such as *Streptomyces coelicolor* may be the best candidates for heterologous expression of *Lyngbya* natural product proteins and gene clusters. At the same time, introduction of *Lyngbya* genes into *Streptomyces* strains

may also allow for novel compounds to be produced if *Streptomyces* are able to utilize the encoded *Lyngbya* enzymes in natural product biosynthesis. Successful expression of both LtxB and LtxC in *S. coelicolor* M512 using the pSH19 vector and *in vitro* validation of the functionality of *Streptomyces* produced LtxC demonstrated that individual, nonmodular proteins from *Lyngbya* may in fact be suitable for this latter approach.

All of the natural product gene clusters characterized to date from *Lyngbya* strains include at least one modular polyketide synthase (PKS) and/or non-ribosomal peptide synthetase (NRPS) encoding ORF (Gu et al. 2009; Jones et al. 2010). These ORFs are typically very large in size (5 kbp or more), making them difficult to clone by traditional PCR techniques, and they usually contain a high number of TTA codons encoding leucine [e.g., *ltxA* has 64; the mixed PKS-NRPS gene *curF* from the curacin A pathway (9588 bp; Chang et al. 2004) has 112]. The rarity of the TTA codon in *S. coelicolor* and its use in regulating secondary metabolite production and morphological changes (Chater and Chandra 2008) suggests that ORFs rich in TTA codons may not be properly translated in expression attempts. The lack of success with overexpressing LtxA when cloned into the pSH19 vector, either using a His tag for nickel affinity or looking at organic *S. coelicolor* extracts for the production of the expected dipeptide, may mean that adequate production of these large proteins will require codon optimization for complete translation to occur. Further modification of *S. coelicolor* could also be attempted to improve the ability of this strain to translate TTA codons. One approach could be to overexpress the *bldA* gene encoding the tRNA for leucine.

Homologous recombination to modify the original lyngbyatoxin fosmid circumvented the need for cloning *ltxA-D*, and also provided for upstream promoter and ribosome binding sites to be exchanged. Although no exconjugants or transformants were obtained when using these constructs, they may be viable candidates for conjugation into other actinobacterial strains in the future. Cloning of *ltxA-C* was also achieved in the pSH19 vector. Repeated attempts to express these genes and isolate lyngbyatoxin A were unsuccessful; however, RT-PCR experiments found that YEME was the only medium examined where transcripts of these genes were detected (RT-PCR also conducted on *ltxA-C* in GYM media). YEME media was used in following the original protocol for this vector (Herai et al. 2004), but it was also chosen because *S. coelicolor* mycelia grow in a more dispersed form than in other media, such as TSB or GYM (Gomez-Escribano and Bibb 2010), possibly enabling easier extraction of proteins and organic compounds (A. Eustáquio, personal communication). The high sugar content (34%) may be necessary for the pSH19 vector to function correctly, although the nitrilase promoter has been used successfully in other media (Nagasawa et al. 1990). If produced, it is also unclear whether any biosynthetic intermediates or the final lyngbyatoxin molecule would be secreted by *S. coelicolor*. Lack of any detectable amount of lyngbyatoxin A by MALDI analysis in spent culture broth and no production of any new compounds during co-culture of 3 separate pSH19 vectors expressing each gene independently may support the absence of compound secretion. Disc diffusion assays did not reveal any toxicity of N-methyl-L-valyl-L-tryptophanol, ILV, or lyngbyatoxin A to *S. coelicolor*, indicating that gene

translation limitations may be the only obstacle to *S. coelicolor* lyngbyatoxin expression.

Future attempts to express the lyngbyatoxin gene cluster or other natural product pathways from *Lyngbya* strains in actinobacteria will probably need to employ homologous recombination strategies instead of cloning entire gene clusters into inducible vectors, as larger cluster sizes will either inhibit cloning or require multiple vectors to be used (Watanabe et al. 2006). If genomic DNA libraries are created, it may be useful to use different cosmid or fosmid types beyond the pCC1Fos fosmid (Epicentre) that has been used several times in cloning *Lyngbya* pathways (Edwards et al. 2004; Edwards and Gerwick 2004), instead using others that have been compatible with the conjugal bridge vector pUZ8002 (Eustáquio et al. 2005).

The work presented in this chapter represents a first attempt to develop genetic techniques for *Lyngbya* strains and *Lyngbya* natural product gene clusters in *Streptomyces*. Although complications with several approaches prevented the expression of lyngbyatoxin A and the entire lyngbyatoxin gene cluster, the successful production of LtxB and LtxC in *Streptomyces coelicolor* M512 using the pSH19 vector represents the first recombinant expression of any *Lyngbya* protein outside of an *E. coli* expression system, and demonstrates that individual enzymes from *Lyngbya* strains can be used in efforts to complement biosynthetic pathways in actinobacteria. Use of homologous recombination or replicating vector strategies to express *Lyngbya* natural products in *Streptomyces* or other hosts will likely find success in the near future as these genetic methods continue to be improved and codon optimization options become more cost-effective and readily available.

4.5 Materials and Methods

4.5.1 *E. coli* strains and culture conditions

E. coli DH5a, DH10b, or TOP10 (Invitrogen) were used for routine cloning (DH5a or DH10b were either made electrocompetent with glycerol washes or purchased from the manufacturer (Allele Extreme competent cells or Genessee GC5 or GC10) as chemically competent cell aliquots. When necessary, *E. coli* strains were grown under antibiotic selection [ampicillin (100 μg mL^{-1}), kanamycin (50 μg mL^{-1}), chloramphenicol (12.5 or 25 μg mL^{-1}), apramycin (50 μg mL^{-1}), or different combinations thereof when appropriate]. *E. coli* BW25113 (pKD20 and pIJ790; Gust et al. 2003) and *E. coli* 12567 were obtained from the Moore laboratory (SIO/UCSD). *E. coli* AM1359 and AM1954 were obtained from the J. Golden laboratory (UCSD).

4.5.2 Plasmids and fosmids used for cloning and conjugation

The lyngbyatoxin fosmid (fos-DE3-86; pCC1Fos Copycontrol vector, Epicentre, Madison, WI) originally used in sequencing of the lyngbyatoxin gene cluster was used for all PCR amplifications of lyngbyatoxin genes or the entire pathway. The chloramphenicol resistance gene in this fosmid was replaced with the apramycin resistant integrase cassette from pSET152 (Floriano and Bibb 1994) or a cassette consisting of a kanamycin/neomycin resistance gene and fragment of a *Nostoc* gas vesicle protein encoding gene by homologous recombination using the *E. coli* strain BW25113 and the λ–RED phage vector pKD20 (ampicillin resistance) under arabinose induction. The region immediately upstream of the *ltxA* gene was replaced with cassettes consisting of a *bla* gene, ermE* promoter and a *Streptomyces* Shine

Dalgarno site or a *bla* gene, T7 promoter and *E. coli* ribosome binding site as described above using *E. coli* BW25113 and the λ–RED phage vector pIJ790 (chloramphenicol resistance) under arabinose induction. The pSH19 vector was obtained from the M. Kobayashi laboratory (University of Tsukuba, Japan) via a Material Transfer Agreement with the Moore laboratory. The *E. coli oriT* enabling *E. coli* cloning was inserted into the SpeI vector site by A. Eustáquio (SIO/UCSD). The AM505 and AM505 + pRNAPP promoter vectors were obtained via a Material Transfer Agreement with the J. Golden laboratory. pSH19 and AM505 constructs containing *ltxA*, *ltxB*, *ltxC* or combinations thereof were created using purified PCR products amplified from the DE3-86 fosmid.

4.5.3 Growth media

E. coli were grown using LB, SOB, or on occasion Terrific Broth (TB) media. *Nostoc* sp. PCC7120 was grown in freshwater BG-11 (FW BG-11 media), while *Lyngbya* strains were grown in saltwater BG-11 (SW BG-11 media; Castenholz 1988). *Streptomyces coelicolor* M512 was grown in TSB and YEME media for most purposes, and plated on MS, R5, or R2YE media (Kieser et al. 2000). For lyngbyatoxin A production attempts, *S. coelicolor* was also grown in GYM (Shima et al. 1996), SMM, NMMP (Kieser et al. 2000), and GPS media (Strohl et al. 1999: glucose 22.5 g L^{-1}, cottonseed flour (Sigma) 10 g L^{-1}, NaCl 3 g L^{-1}, $CaCO_3$ 3 g L^{-1}, trace elements 10 mL L^{-1}; Dekleva et al.1985).

4.5.4 PCR, digestion, ligation enzymes, protein reagents

Normal PCR screening of vectors and ligation products was performed with PCR Mastermix (Promega). Vector inserts and homologous recombination cassettes were amplified with the proofreading Taq polymerases Pfx50 or Platinum Taq High Fidelity (Invitrogen). Digestion reactions were performed with enzymes from New England Biolabs. Ligations were performed using T4 DNA ligase (New England Biolabs or Fermentas). Proteins were visualized using SDS-PAGE 4-20% Tris-HCl gels (Bio-Rad), and dialysis was conducted using Slide-A-Lyzer Dialysis Cassettes (Thermo Scientific).

4.5.5 Primers

A list of primers used in this study are provided in Table 4.2.

Table 4.2: Primers used in Chapter 4. Those in black were used in cloning individual genes for protein expression. Those in green were used for RNA experiments and general sequencing. Those in blue were used in the creation of gene cassettes for homologous recombination.

Primer name	Primer sequence
LtxA SD SbfI F	GGGGCCTGCAGGTTAGCAACGGAGGTAC GGACATGATTATGAATCAACCTTGG
LtxA SD SbfI His F	GGGG CCTGCAGG AGCAACGGAGGTACGGAC ATG CACCACCACCACCACCAC ATGATTATGAATCAACCT
LtxB SD SbfI His F	GGGGCCTGCAGGAGCAACGGAGGTACGG ACATGCAC CACCACCACCACCACATGACAAATCCTTT TGCA
LtxB sacI SD F	GGGG GAGCTCAGCAACGGAGGTACGGAC ATGACAAATCCTTTTGCA

Table 4.2 (continued): Primers used in Chapter 4. Those in black were used in cloning individual genes for protein expression. Those in green were used for RNA experiments and general sequencing. Those in blue were used in the creation of gene cassettes for homologous recombination.

LtxC SD SbfI His F	GGGG CCTGCAGG AGCAACGGAGGTACGGAC ATG CACCACCACCACCACCAC ATGAATTCAAAGATCGCT
LtxA SacI R stop	CCCCGAGCTCTTACTTTCCTGTATAAGT
LtxB SacI R stop	CCCCGAGCTCTTACCACTCAGCAGGTAAC T
LtxC SacI F	GGGGGAGCTCATGAATTCAAAGATCGCTG T
LtxC SacI R stop	CCCC GAGCTC TTA TGA CCG TTT ATA GAC TCC
LtxC His SacI R stop	CCCCGAGCTCTTAGTGGTGGTGGTGGTGG TG CCCCCCTGACCGTTTATAGACTCC
LtxD SacI R stop	CCCC GAGCTC TTAGATCAACCCTCCTGTAACT
LtxC SalI R His stop	CCCCGTCGACTTAGTGGTGATGGTGGTGG TGTGACCGTTTATAGACTCCTG
LtxC NotI SD F	GGGGGCGGCCGCAGCAACGGAGGTACGG ACATGAATTCAAAGATCGCTG
LtxC His SmaI R stop	CCCCCCCGGGTTAGTGGTGGTGGTGGTGG TG CCCCCCTGA CCGTTTATAGACTCCTGG
LtxA SacI F	ATGATTATGAATCAACCTTGGA
LtxD SalI R Stop	CCCC GTCGAC TTAGATCAACCCTCCTGTAACT
LtxD SacI R stop	CCCC GAGCTC TTAGATCAACCCTCCTGTAACT
LtxA RNA F	ATGATTATGAATCAACCTTG
LtxA RNA R	CAGCAACCACTAACTCAATC
LtxB RNA F	ATGACAAATCCTTTTGCA
LtxB RNA R	CCATGTTAAGGGCAACATAT
LtxC RNA F	ATGAATTCAAAGATCGCTG
LtxC RNA R	TACAGAGTTTACGAAGGAAGG

Table 4.2 (continued): Primers used in Chapter 4. Those in black were used in cloning individual genes for protein expression. Those in green were used for RNA experiments and general sequencing. Those in blue were used in the creation of gene cassettes for homologous recombination.

Pset cat F	GAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGTGC TACAGAGTTCTTGAAGTG
Pset cat R	AGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTATCGATCAGAAACTTCTCGAC
P1 bla F	GCACTTTTCGGGGAAATGTG
P2 ermE bla R	GGCGCAAGCCGCCACTCGAACGGACACTCGCTGAGTAAACTTGGTCTGACAG
P3 ermE F	GCGAGTGTCCGTTCGAGTG
P4 ermE R	GGATCCTACCAACCGGCAC
P5 ltxA bla erm swap F	ACACTTTTGCCAACTTGTATAGTAGGAAACTTCTACCTGGCACTTTTCGGGGAAATGTG
P6 ltxA SD erm swap R	CGCCCGCCTACTTCCACTCCAAGGTTGATTCATAATCATGTCCGTACCTCCGTTGCTGGATCCTACCAACCGGCAC
P2 T7 R primer bla	CACCTGTGGCGCCGGTGATGCCGGCCACGAAAACTTGGTCTGACAG
P3 T7 S F	CACCTGTGGCGCCGG
P4 T7 S R	GGTATATCTCCTTCTTAAAGTTAA
P6 LtxA T7 R	CGCCCGCCTACTTCCACTCCAAGGTTGATTCATAATCATGGTATATCTCCTTCTTAAAG
Km F	GATCAAGAGACAGGATGAGG
Km R	TCTACCTTTTCTAAATTTCTTGAATT
Ana GVP F	AATTCAAGAAATTTAGAAAAGGTAG
Ana GVP R	CCCAAGGGATCAAAAAT
KM GVP swap F	GAGTTATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAT CAAGAGACAGGATGAGG
KM GVP swap R	AGGCGTTTAAGGGCACCAATAACTGCCTTAAAAAAATTACCCAAGGGATCAAAAAT
LtxC KpnI SD F	GGTACCAGCAACGGAGGTACGGACATGAATTCAAAGATCGCT
LtxB stitch R	ACTTAAGTACAGGCATGGAGGCAACGACCATGGTTACCACTCAGCAGGTAA

4.5.6 Morpholino incubations with *Lyngbya majuscula* filaments

Two morpholino experiments were conducted with *Lyngbya majuscula* JHB filaments to determine whether passive uptake into the cells was possible. Nonspecific morpholinos that were either unlabeled or tagged with fluorescein (Gene-Tools LLC; L. Gerwick laboratory) at 1 mM concentration were added to wells of a 24 well plate containing 100 μL of SW-BG-11 media in combination with small (~1 mg) amounts of *L. majuscula* JHB filaments for several days. Following this incubation period MALDI analysis of filament samples were performed (in collaboration with E. Esquenazi) and filaments were examined using fluorescence microscopy.

4.5.7 Gene gun attempts to introduce DNA into *Lyngbya majuscula* and *Nostoc* sp. PCC 7120

Biolistic (gene gun) transformations were attempted using the Bio Rad Gene Gun in the M. Hildebrand laboratory (SIO/UCSD). For this procedure, tungsten particles (10 mg) were first washed in ethanol and sterile water, and divided into aliquots. Each aliquot received at least 1 μg of plasmid DNA, 50 μL of sterile 2.5M $CaCl_2$, and 20 μL sterile 0.1M spermidine. These aliquots were vortexed, centrifuged, and washed and resuspended in ethanol. *Lyngbya* filaments or *Nostoc* cultures were dispersed or spread onto BG-11 agar plates, and after gene gun assembly these cultures were shot with the tungsten particles. *Nostoc* cultures were allowed to grow on the BG-11 plates before antibiotic was added, while *Lyngbya* filaments were allowed to recover in liquid media before the appropriate antibiotic was added the next day.

4.5.8 RNA extraction and cDNA synthesis from *Streptomyces coelicolor* M512

RNA from *Streptomyces coelicolor* was isolated from YEME culture pellets that were frozen in liquid nitrogen and homogenized using a mortar and pestle. RNA was obtained using Trizol (Invitrogen) and procedures based on those of the manufacturer with minor modifications. The isolated RNA was treated with DNAse I (Ambion) for 2 h at 38°C, and was stored at -80°C after being treated with DTT and RNAseOUT (Invitrogen). For RT-PCR experiments and generation of first strand cDNA, *S. coelicolor* RNA was used in reactions with Superscript III Reverse Transcriptase (Invitrogen) and primers shown in Table 4.2 (blue). For each primer set, parallel reactions were performed as negative controls in which Superscript enzyme was omitted. First strand cDNA was used as template for second strand cDNA reactions.

4.5.9 PCR-targeted gene replacement in fosmids for subsequent use in *Streptomyces coelicolor* M512, *Nostoc* sp. PCC7120 and *E. coli* BL21

As described above, gene cassettes for homologous recombination were amplified with PCR primers containing 39 bp DNA fragments homologous to the regions surrounding the gene to be replaced. In each case, the DE3-86 fosmid was transformed into electrocompetent *E. coli* BW25113 cells containing a λ–RED phage vector and plated at 30°C. A colony from this transformation was grown in SOB media at 30°C under arabinose induction (10 mM) to an OD_{600} of approximately 0.6, at which time it was induced again with 10 mM arabinose and grown for an additional

30 min. This culture was made electrocompetent by washing 2 times in 10% glycerol and the PCR cassette was transformed into these cells by electroporation and plated only using the antibiotic selection of the cassette. Colonies from this transformation were grown, and fosmid DNA isolated from the colonies was transformed into Epi300 copy control cells (Epicentre) by electroporation. These cells were grown and induced to produce the fosmid at a higher copy number before the fosmid was reisolated and analyzed by PCR, restriction digest, and/or sequencing to confirm successful gene replacement had occurred.

4.5.10 *Streptomyces* conjugation

Conjugation was performed in *S. coelicolor* M512 by first transforming the desired construct into *E. coli* 12567 containing the conjugal bridge plasmid pUZ8002 (Gust et al. 2003). Colonies from this transformation were grown to an OD_{600} of between 0.6 and 1.0 and these cultures were pelleted and washed to remove antibiotic. After resuspension in 1 mL LB media, these pellets were mixed with approximately 1 x 10^8 spores of *S. coelicolor* M512 and plated on MS agar + 10 mM $MgCl_2$. After 16-20 h these plates were overlaid with 0.5 mg nalidixic acid (20 μL of 25 mg mL^{-1} stock) and 1.25 mg apramycin (25 μL of 50 mg mL^{-1} stock), and the plates were grown for several days before the appearance of exconjugants. When present, these exconjugants were screened for the insertion of the construct by colony PCR.

4.5.11 *Nostoc* sp. PCC7120 conjugation

Constructs for conjugation into *Nostoc* sp. PCC7120 were first transformed into electrocompetent AM1359 *E. coli*, which contained the conjugal bridge plasmid pRL443 and the methylating plasmid pRL623 (e.g., Wu et al. 2004). A colony from this transformation was grown to an OD600 of between 0.6 and 1.0, and 1.5 mL of this culture was pelleted and washed with LB media to remove any antibiotics. This pellet was resuspended in residual LB media and mixed with a pellet from 1.5 mL of a *Nostoc* sp. PCC7120 culture that had been grown for 2-4 days. This mixture was incubated at room temperature for up to 30 minutes before plating on FW BG-11 agar plates and incubated at approximately 30 μE m^{-2} s^{-1} at 28°C for 16 to 20 h. After this incubation period, the plates were underlaid with 25 μg mL^{-1} neomycin (100X stock in 400 μL of BG-11 medium) and moved to approximately 50 – 70 μE m^{-2} s^{-1} and allowed to grow for 7-10 days before the appearance of resistant colonies. None of the colonies obtained were able to grow in liquid FW BG-11 media in plastic or glass test tubes.

4.5.12 Protoplast preparation of *Streptomyces coelicolor* spores and induction of pSH19 vector (protocol from Kieser et al. 2000)

Streptomyces spores were grown for 36-40 h in YEME media supplemented with glycine (0.5%). This culture was pelleted and washed twice in 10.3% sucrose. The pellet was resuspended in 4 mL of 1 mg/mL lysozyme solution in P buffer (Kieser et al. 2000) and incubated for 1 h at 30°C. The pellet was drawn in and out of a pipet 3 times, incubated for another 15 minutes at 30°C, diluted with another 5 mL of P

buffer and filtered through cotton wool. These protoplasts were collected by centrifugation and dispensed in 50 μL aliquots. Transformation into these protoplasts by mixing them with up to 5 μL of construct DNA isolated from *E. coli* 12567 cells (with or without pUZ8002) in 200 μL T buffer {per 10.33 mL: 25% sucrose, 1 mL; 50% PEG1000, 5 mL; 140 mM K_2SO_4, 100 μL; 1M $MgCl_2$, 100 μL; 40 mM KH_2PO_4, 100 μL; 2.5M $CaCl_2$, 400 μL; 500 mM Tris Maleate pH 8, 1 mL; trace elements [Dekleva et al. (1985) with minor modification, 30 μL)]} before plating on R2YE or R5 plates (Kieser et al. 2000). After 16-20 h these plates were overlaid with 50 μg mL^{-1} thiostrepton in 1 mL sterile H_2O and incubated at 28°C for 2-4 days until transformants appeared. These colonies were grown for 2-3 days in TSB media under thiostrepton selection and used to inoculate larger cultures in YEME or other media (see below). Induction with ε-caprolactam (0.1% w/v) was performed after 96 h of growth.

4.5.13 Protein purification from pSH19 cultures

After induction at 96 h, *S. coelicolor* M512 was harvested at 120 h. These cultures were centrifuged and resuspended in potassium phosphate buffer (pH 7.5) and 1mM DTT, along with a broad range protease inhibitor (Complete, Roche). The cell pellets were sonicated 5-6 times with 10 second pulses on ice, and soluble protein was obtained by centrifugation at 14,000 rpm. This soluble protein was incubated with Nickel Agarose or Nickel Superflow resin (Qiagen) for 2 h on an end-over-end rotator at 4°C. After the incubation the nickel resin was collected on a 5 mL polypropylene column (Qiagen) and washed with wash buffer (20 mM Tris, pH 8, 500 mM NaCl, 40

mM imidazole). The protein was then eluted from the resin in 3 x 1 mL aliquots in elution buffer (20 mM Tris, pH8, 500 mM NaCl, 750 mM imidazole) and visualized on SDS-PAGE gels with colloidal Coomassie staining. When needed, protein elutions were dialyzed against 50 mM PBS pH 7.5, 200 mM NaCl and 10% glycerol before being separated into aliquots, flash frozen in liquid N_2 and stored at -80°C.

4.5.14 Protein assays with LtxC

Protein assays with LtxC purified from *S. coelicolor* M512 were conducted as described by Edwards and Gerwick (2004). LtxC (1 μM) was incubated with (-) ILV and geranyl pyrophosphate (0.1 mM) in the presence of 2 mM $MgCl_2$ for 1 h at room temperature (200 μL scale). Other parallel reactions were conducted with either no protein added or purified protein or protein extracts added to approximately 1 μM concentrations. Each reaction was extracted twice with ethyl acetate, and the ethyl acetate extract was dried under nitrogen. These extracts were resuspended in 40 μL of 80% MeOH in H2O for LC-MS analysis. Each extract was profiled using gradient conditions (75% MeOH for 20 minutes, followed by an increase to 100% MeOH by 30 minutes) on a Phenomenex Prodigy 3 uL ODS (3) 100 A, 100 X 4.6 mm 3 μ HPLC column.

4.5.15 *Streptomyces coelicolor* M512 extract profiling for lyngbyatoxin production

Cultures being examined for lyngbyatoxin production were extracted using either XAD-7 (Amberlite) resin or ethyl acetate. For resin extracts, approximately 20 g/L of resin was added to each culture and incubated for 2 h with shaking at 220 RPM

at 28°C. This resin was filtered through cheesecloth and extracted with acetone for 2 h, while the culture broth was extracted with ethyl acetate. The acetone extract was evaporated to the remaining aqueous material, and this material was back extracted with ethyl acetate. The ethyl acetate extracts from both culture portions were evaporated and in some instances run over solid phase C_{18} cartridges (either with an acetonitrile gradient or with 100% acetonitrile). A portion of each extract was filtered (0.2 μm) and profiled by LC-MS (either the Phenomenex Prodigy column mentioned above or a Phenomenex Luna 5 u C18 (2) 100A 100 X 4.6 mM 5 micron column) to look for the presence of lyngbyatoxin A. For whole culture extracts, ethyl acetate was added in a 1:1 ratio with each culture and shaken at 150 – 220 RPM for 2 h. The ethyl acetate layer was obtained using a separatory funnel, evaporated to dryness, and treated as described in the resin extract procedure. The small culture supplemented with 100 μg lyngbyatoxin A was extracted with ethyl acetate, filtered, and profiled directly by LCMS.

4.6 Acknowledgements

The chapter author acknowledges Dr. Sabine Ottilie, Dr. Alessandra Eustáquio and Dr. Sushanta Saha for helpful discussion and experimental design assistance through all phases of this project. The author also wishes to thank the B. Moore laboratory (SIO/UCSD) for general assistance and support. The Kobayashi laboratory (University of Tsukuba, Japan) is acknowledged for provision of the pSH19 vector. The M. Hildebrand Laboratory (SIO) and Dr. Roshan Shrestha are acknowledged for gene gun assistance. Support to the chapter author was provided by a graduate

fellowship from the Los Angeles Chapter of the Achievement Rewards for College Scientists (ARCS) Foundation, and NOAA Grant NA08OAR4170669, California Sea Grant College Program Project SG-100-TECH-N, through NOAA's National Sea Grant College Program, US Department of Commerce. Chapter 4 of this dissertation includes a portion that is being prepared for submission in 2011. Jones, Adam C; Ottilie, Sabine; Eustáquio, Alessandra S.; Edwards, Daniel J; Gerwick, Lena; Moore, Bradley S; Gerwick, William H. Evaluation of *Streptomyces coelicolor* as a heterologous expression host for natural products from the filamentous cyanobacterium *Lyngbya majuscula*. The dissertation author was the primary investigator and author of this paper.

4.7 References

Baltz RH (2008) Renaissance in antibacterial discovery from actinomycetes. *Curr Op Pharmacol* 8:557-563.

Basu A, Kozikowski AP, Lazo JS (1992) Structural requirements of lyngbyatoxin A for activation and downregulation of protein kinase C. *Biochemistry* 31:3824-3830.

Bibb MJ, White J, Ward JM, Janssen GR (1994) The mRNA for the 23S rRNA methylase encoded by the *ermE* gene of *Saccharopolyspora erythraea* is translated in the absence of a conventional ribosome-binding site. *Mol Microbiol* 14:533-545.

Bierman M, Logan R, O'Brien K, Seno ET, Rao RN, Schoner BE (1992) Plasmid cloning vectors for the conjugal transfer of DNA from *Escherichia coli* to *Streptomyces* spp. *Gene* 116:43-49.

Cardellina II JH, Marner FJ, Moore RE (1979) Seaweed dermatitis: Structure of lyngbyatoxin A. *Science* 204:193-195.

Chang Z, Sitachitta N, Rossi JV, Roberts MA, Flatt PM, Jia J, Sherman DH, Gerwick WH (2004) Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 67:1356-1367.

Chater KF, Chandra G (2008) The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in *Streptomyces*. *J Microbiol* 46:1-11.

Combes P, Till R, Bee S, Smith MCM (2002) The *Streptomyces* genome contains multiple pseudo-*attB* sites for the φC31-encoded site-specific recognition system. *J Bacteriol* 184:5746-5752.

Dekleva ML, Titus JA, Strohl WR (1985) Nutrient effects on anthracycline production by *Streptomyces peucetius* in a defined medium. *Can J Microbiol* 31:287-294.

Draper BW, Morcos PA, Kimmel CB (2001) Inhibition of zebrafish *fgf8* pre-mRNA splicing with morpholino oligos: A quantifiable method for gene knockdown. *Genesis* 30:154-156.

Edwards DJ, Gerwick WH (2004) Lyngbyatoxin biosynthesis: Sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *J Am Chem Soc* 126: 11432-11433.

Elibol M, Mavituna F (1998) Effect of sucrose on actinorhodin production by *Streptomyces coelicolor* A3(2). *Process Biochem* 33:307-311.

Eustáquio AS, Gust B, Galm U, Li S-M, Chater KF, Heide L (2005) Heterologous expression of novobiocin and chlorobiocin biosynthetic gene clusters. *Appl Environ Microbiol* 71:2452-2459.

Flinspach K, Westrich L, Kaysser L, Siebenberg S, Gomez-Escribano JP, Bibb MJ, Gust B, Heide L (2010) Heterologous expression of the biosynthetic gene clusters of coumermycin A1, chlorobiocin and caprazamycins in genetically modified *Streptomyces coelicolor* strains. *Biopolymers* 93:823-832.

Fujiki H, Mori M, Nakayasu M, Terada M, Sugimura T, Moore RE (1981) Indole alkaloids: Dihydroteleocidin B, teleocidin, and lyngbyatoxin A as members of a new class of tumor promoters. *Proc Natl Acad Sci USA* 78: 3872-3876.

Gerwick WH, Coates RC, Engene N, Gerwick L, Grindberg RV, Jones AC, Sorrels CM (2008) Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3:277-284.

Gressler M, Zaehle C, Scherlach K, Hertweck C, Brock M (2011) Multifactorial induction of an orphan PKS-NRPS gene cluster in *Aspergillus terreus*. *Chem Biol* 18:198-209.

Grindberg RV, Ishoey T, Brinza D, Esquenazi E, Coates RC, Liu W-T, Gerwick L, Dorrestein PC, Pevzner P, Lasken R, Gerwick WH (2011) Single cell genome amplification accelerates identification of the apratoxin biosynthetic pathway from a complex microbial assemblage. *PLoS ONE* (in press).

Gust B, Challis GL, Fowler K, Kieser T, Chater KF (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA* 100:1541-1546.

Heide L (2009) Aminocoumarins: Mutasynthesis, chemoenzymatic synthesis, and metabolic engineering. *Meth Enzymol* 459:437-455.

Herai S, Hashimoto Y, Higashibata H, Maseda H, Ikeda H, Ōmura S, Kobayashi M (2004) Hyper-inducible expression system for streptomycetes. *Proc Natl Acad Sci USA* 101:14031-14035.

Huynh MU, Elston MC, Hernandez NM, Ball DB, Kajiyama S-i, Irie K, Gerwick WH, Edwards DJ (2010) Enzymatic production of (-)-Indolactam V by LtxB, a cytochrome P450 monooxygenase. *J Nat Prod* 73:71-74.

Irie K, Nakagawa Y, Tomimatsu S, Ohigashi H (1998) Biosynthesis of the monoterpenoid moiety of teleocidins *via* the non-mevalonate pathway in *Streptomyces*. *Tet Lett* 39:7929-7930.

Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) Practical *Streptomyces* Genetics, 2nd ed. Norwich, UK: John Innes Foundation.

Lane AL, Moore BS (2010) A sea of biosynthesis: Marine natural products meet the molecular age. *Nat Prod Rep* 28:411-428.

Leskiw BK, Bibb MJ, Chater KF (1991) The use of a rare codon specifically during development? *Mol Microbiol* 5:2861-2867.

Mazodier P, Petter R, Thompson C (1989) Intergeneric conjugation between *Escherichia coli* and *Streptomyces* species. *J Bacteriol* 171:3583-3585.

Nagasawa T, Nakamura T, Yamada H (1990) ε-caprolactam, a new powerful inducer for the formation of *Rhodococcus rhodochrous* J1 nitrilase. *Arch Microbiol* 155:13-17.

Pfeifer BA, Wang CCC, Walsh CT, Khosla C (2003) Biosynthesis of yersiniabactin, a complex polyketide-nonribosomal peptide, using *Escherichia coli* as a heterologous host. *Appl Environ Microbiol* 69:6698-6702.

Quadri LE, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* 37:1585-1595.

Rachid S, Gerth K, Kochems I, Müller R (2007) Deciphering regulatory mechanisms for secondary metabolite production in the myxobacterium *Sorangium cellulosum* So ce56. *Mol Microbiol* 63:1783-1796.

Read JA, Walsh CT (2007) The lyngbyatoxin biosynthetic assembly line: Chain release by four-electron reduction of a dipeptidyl thioester to the corresponding alcohol. *J Am Chem Soc* 129:15762-15763.

Roberts AA, Copp JN, Marahiel MA, Neilan BA (2009) The *Synechocystis* sp. PCC6803 Sfp-type phosphopantetheinyl transferase does not possess characteristic broad-range activity. *ChemBioChem* 10:1869-1877.

Rossi JV, Roberts MA, Yoo H-D, Gerwick WH (1997) Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically-useful natural metabolite curacin A. *J Appl Phycol* 9:195-204.

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

Strohl WR, Dickens ML, Desanti CL (1999) "Methods of Producing Doxorubicin." US Patent No. 5,976,830.

Summerton J, Weller D (1997) Morpholino antisense oligomers: Design, preparation, and properties. *Antisense Nucleic Acid Drug Dev* 7:187-195.

Suyama TL, Gerwick WH (2008) Stereospecific total synthesis of somocystinamide A. *Org Lett* 10:4449-4452.

Takano E, Tao M, Long F, Bibb MJ, Wang L, Li W, Buttner MJ, Bibb MJ, Deng ZX, Chater KF (2003) A rare leucine codon in *adpA* is implicated in the morphological defect of *bldA* mutants of *Streptomyces coelicolor*. *Mol Microbiol* 50:475-486.

Watanabe K, Hotta K, Praseuth AP, Koketsu K, Migita A, Boddy CN, Wang CCC, Oguri H, Oikawa H (2006) Total biosynthesis of antitumor nonribosomal peptides in *Escherichia coli*. *Nat Chem Biol* 2:423-428.

Wilkinson B, Micklefield J (2007) Mining and engineering natural-product biosynthetic pathways. *Nat Chem Biol* 3:379-386.

Wilkinson CJ, Hughes-Thomas ZA, Martin CJ, Bohm I, Mironenko T, Deacon M, Wheatcroft M, Wirtz G, Staunton J, Leadlay PF (2002) Increasing the efficiency of heterologous promoters in actinomycetes. *J Mol Microbiol Biotechnol* 4:417-426.

Wu X, Liu D, Lee MH, Golden JW (2004) *patS* minigenes inhibit heterocyst development of *Anabaena* sp. strain PCC 7120. *J Bacteriol* 186:6422-6429.

CHAPTER 5

Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*

5.1 Abstract

Filamentous cyanobacteria of the genus *Lyngbya* are important contributors to coral reef ecosystems, occasionally forming dominant cover and impacting the health of many other co-occurring organisms. Moreover, they are extraordinarily rich sources of bioactive secondary metabolites, with 35% of all reported cyanobacterial natural products deriving from this single pantropical genus. However, the true natural product potential and life strategies of *Lyngbya* strains are poorly understood because of phylogenetic ambiguity, lack of genomic information, and their close associations with heterotrophic bacteria and other cyanobacteria. To gauge the natural product potential of *Lyngbya* and gain insights into potential microbial interactions, the genome of *Lyngbya majuscula* 3L, a Caribbean strain that produces the tubulin polymerization inhibitor curacin A and the molluscicide barbamide, was sequenced using a combination of Sanger and 454 approaches. While approximately 293,000 nucleotides of the draft genome are putatively dedicated to secondary metabolism, this is far too few to encode a large suite of *Lyngbya* metabolites, suggesting *Lyngbya* metabolites are strain specific and may be useful in species delineation. Genome analysis revealed a complex gene regulatory network, including a large number of sigma factors and other regulatory proteins, indicating an enhanced ability for environmental adaptation or microbial associations. Although *Lyngbya* species are reported to fix nitrogen, nitrogenase genes were not found in the genome or by PCR of

genomic DNA. Subsequent growth experiments confirmed that *L. majuscula* 3L is unable to fix atmospheric nitrogen. These unanticipated life history characteristics challenge current views of the genus *Lyngbya*.

5.2 Introduction

Among the oldest life forms on Earth, cyanobacteria are well recognized for their global ecological importance and ubiquitous distribution across virtually all ecosystems (Zehr et al. 2001). In the marine realm, some species of cyanobacteria contribute significantly to nitrogen fixation and global carbon flux (Berman-Frank et al. 2001), while others are prevalent as benthic constituents of tropical coral reefs (Thacker et al. 2001). Over the past several decades, cyanobacteria have become recognized as an extremely rich source of novel, bioactive secondary metabolites (= natural products), with approximately 700 different compounds having been isolated and characterized (Tidgewell et al. 2010). These compounds have gained considerable attention due to their pharmaceutical and biotechnology potential (Tan 2007), but also notoriety for their environmental toxicity and threats to humans, wildlife, and livestock (Carmichael 2001).

Marine strains of the genus *Lyngbya* are one of the most prolific producers of natural products. Nearly 240 compounds are reported from this genus, and 76% of these are attributed to a single species, *Lyngbya majuscula* (Harvey ex Gomont), which is found globally in shallow tropical and subtropical environments (Tidgewell et al. 2010). *Lyngbya* bloom events pose a significant challenge to coral reefs, as *Lyngbya* can negatively impact coral larval recruitment (Kuffner and Paul 2004),

quickly colonize available substrate, and persist in the presence of herbivores because of their chemical defenses (Paul et al. 2005). In the last ten years, focused investigations into the biosynthesis of *L. majuscula* natural products have revealed gene clusters that encode the molecular assembly of several of these compounds, including the anticancer agent curacin A (Chang et al. 2004; Gu et al. 2007; Gu et al. 2009), the neurotoxic jamaicamides (Edwards et al. 2004), the UV-sunscreen pigment scytonemin (Sorrels et al. 2009), and the lyngbyatoxins (Edwards and Gerwick 2004), dermatotoxic agents responsible for "swimmer's itch". Most of the gene clusters encode modular, mixed polyketide synthase (PKS)/non-ribosomal peptide synthetase (NRPS) assembly lines, with several employing highly unusual mechanisms to incorporate other functional groups into the resultant molecules (Jones et al. 2010).

Despite these advances in compound identification and biosynthesis, comparatively little is known about *Lyngbya* evolution or the full potential of specific *L. majuscula* strains to produce the natural products attributed to this species. Recent reassessment of phylogenetic diversity in the genus *Lyngbya* using the 16S rRNA gene has shown that *Lyngbya* appears to occupy three distinct clades: a halophilic/brackish/freshwater lineage, a lineage more closely related to the genus *Oscillatoria*, and a marine lineage (Engene et al. 2010). Moreover, metabolites attributed to *L. majuscula* have typically been isolated from field collections, which poses two problems: most taxonomic classifications have been based on morphological characteristics and not genetic evidence, and this cyanobacterium typically grows in close association with other microorganisms. Thus, it is possible

that the total number of natural products associated with the species *L. majuscula* has been overestimated.

To determine the capacity for natural products biosynthesis in a specific strain of *L. majuscula*, genome sequencing was performed using *L. majuscula* 3L, a strain that falls within the marine lineage described earlier, and has also recently been referred to as *Lyngbya sordida* 3L (Engene et al. 2010). *Lyngyba majuscula* 3L was originally isolated in Curaçao, Netherlands Antilles and has been maintained in stable culture for approximately 15 years (Rossi et al. 1997). This strain produces the anti-cancer agent curacin A (Gerwick et al. 1994), the molluscicidal compound barbamide (Orjala and Gerwick 1996) and the lipopeptide carmabin B (Hooper et al. 1998).

A draft genome was obtained from an integrated strategy involving Sanger sequencing of DNA from cultured filaments in combination with 454 sequencing of DNA generated via multiple displacement amplification (MDA) from single *L. majuscula* 3L cells; the latter approach was necessary to overcome the inability to create or maintain axenic cultures of *L. majuscula*. This sequencing effort aimed to confirm the presence of the gene clusters encoding each of these molecules, search for other unknown (orphan) natural product biosynthetic pathways, and gain insights into the physiological ecology of *L. majuscula* in tropical environments, including possible interactions with other microorganisms and the ability of *L. majuscula* to fix atmospheric nitrogen (Jones et al. 1990; Lundgren et al. 2003).

5.3 Results and Discussion

5.3.1 Genome assembly and annotation

L. majuscula 3L sequence reads were obtained from two independent, non-axenic cultures, using two different DNA isolation procedures and two different sequencing technologies (Sanger and 454 approaches). The reads from both the Sanger and 454 libraries were pooled and treated as a single metagenomic data set in order to identify core sequences of the *Lyngbya* genome that were common to both data sets. This strategy enabled evaluation of whether scaffolds assembled from sequences in both data sets contained constituent reads from one or both library sources in order to assist in separating consensus *Lyngbya* sequences from non-*Lyngbya* contaminants. Co-assembly of 712,948 Sanger and 454 reads produced 6,217 scaffolds, ranging in size from 1,000 to 59,782 nucleotides, G+C content between 25 and 76%, and coverage depth from 1 to 62 fold. General genome features and predicted COG categories are listed in Tables 5.S1 and 5.S2. Classification of 16S genes in the combined assembly, taxonomic heterogeneity of the scaffolds, and details regarding the binning procedure are provided in Table 5.S3, Figure 5.S1, and Jones et al. (2011). A total of 161 scaffolds were identified as likely originating from *L. majuscula*, based on a combination of 16S rRNA genes, predicted protein matches to Genbank nr sequences, inclusion of reads from both Sanger and 454 libraries, percent G+C nucleotide composition, and assembly coverage depth. Detailed properties for all *Lyngbya*-associated scaffolds are provided in Jones et al. (2011). The combined scaffolds total approximately 8.5 Mb, a total genome size consistent with other

filamentous cyanobacteria, such as *Nostoc punctiforme* (8.2 Mb) and *Trichodesmium erythraeum* (7.8 Mb). It is uncertain if this draft assembly represents the entire *Lyngbya majuscula* genome, but several lines of evidence suggest it is nearly complete. A survey of 102 housekeeping genes identified as nearly universal in bacteria (Puigbo et al. 2009) indicates that 101 of these are present in the *Lyngbya* draft genome. Copy numbers for these housekeeping genes correlate well with other sequenced cyanobacterial genomes, including those expected to have single copies (Jones et al. 2011). In addition, previously known, independently sequenced *L. majuscula* genes for the curacin A (Chang et al. 2004) and barbamide (Chang et al. 2002) pathways are present and complete, despite not being used to guide any aspect of the assembly. The *L. majuscula* 3L draft genome was submitted to the Joint Genome Institute Integrated Microbial Genomes (IMG) Expert Review for automated annotation of putative Open Reading Frames (ORFs). Within the 8.5 Mb genome (44% GC content), 56 tRNAs, 2 rRNA operons, and 7479 protein encoding genes were identified, with 54% of these protein encoding genes having predicted functions. This number is higher than for *N. punctiforme* (6086 genes) and *T. erythraeum* (4451 genes). The largest percentage of annotated genes (based on COG categories) appear to be involved in replication, recombination and DNA repair (9%), cell wall biogenesis (8%), and signal transduction mechanisms (7%). Despite previous reports that *L. majuscula* strains are diazotrophic, no nitrogenase genes were found in this draft genome.

5.3.2 Secondary metabolism genes in *L. majuscula* 3L draft genome

Despite the large number of natural products attributed to *L. majuscula*, only 126 genes (3%, 293 kb) of the *L. majuscula* 3L draft genome are predicted to be involved in secondary metabolite biosynthesis, transport and catabolism. The largest number of these are modular non-ribosomal peptide synthetase (NRPS) and/or polyketide synthase (PKS) related genes (44%, 199 kb). Eight biosynthetic gene clusters were identified that likely encode natural products (Figures 5.1 and 5.2). The two most apparent clusters were those of the previously characterized natural products curacin A (Accession # HQ696500) and barbamide (HQ696501) (Figure 5.1a). The sequences for both pathways were complete and consistent with the sequences previously reported (Chang et al. 2002; Chang et al. 2004). Two separate scaffolds contain genes putatively involved in carmabin biosynthesis (Figure 5.1b) based on predictions of adenylation domain substrate specificity (see methods) from the NRPS ORFs in each partial gene cluster.

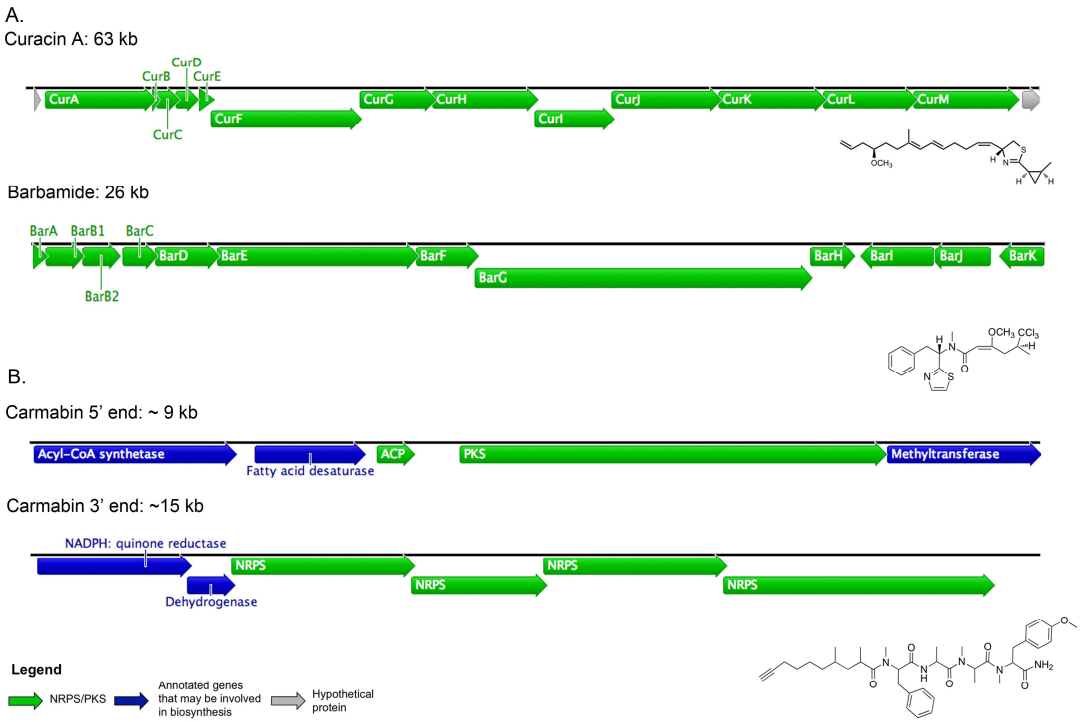


Figure 5.1: Secondary metabolite gene clusters for known compounds in the *L. majuscula* 3L draft genome. A) Intact gene clusters for curacin A and barbamide, both of which are entirely consistent with previous descriptions. B) Two scaffolds (51865 and 52117) containing probable portions of the carmabin biosynthetic gene cluster, based on adenylation domain substrate specificity. Figures constructed using Geneious 5.1 (Drummond et al. 2010).

Five additional biosynthetic gene clusters were found in the *L. majuscula* 3L genome; however, they do not appear to encode natural products previously detected from this species (Figure 5.2).

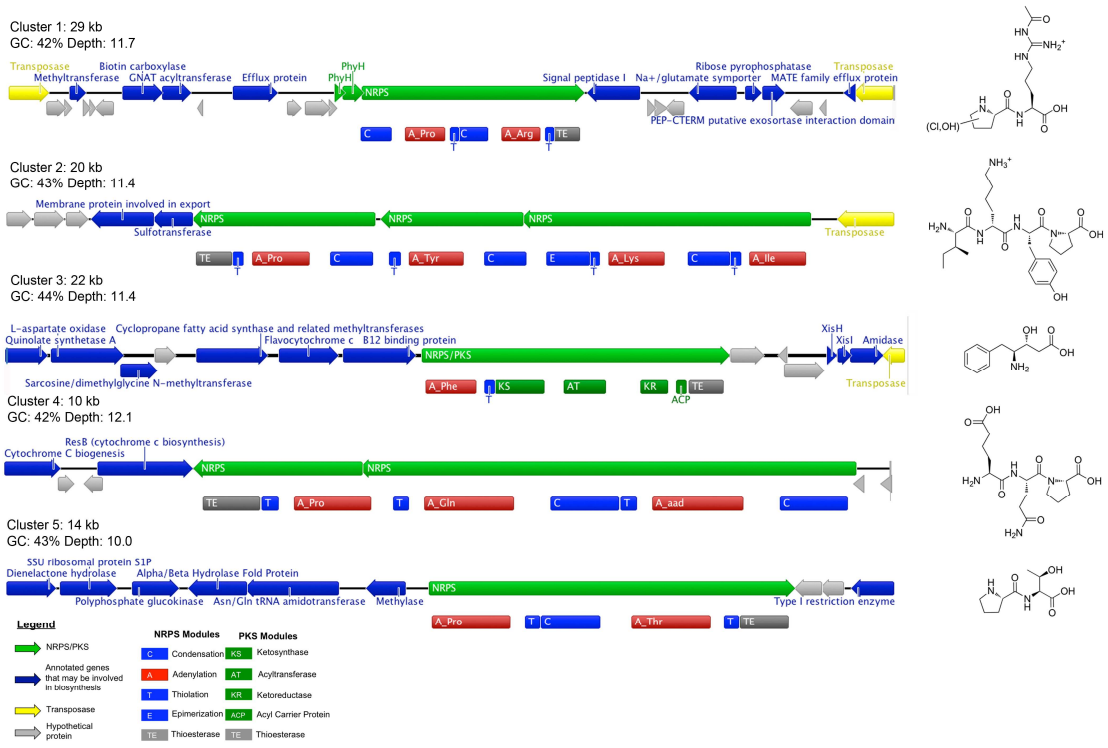


Figure 5.2: Orphan gene clusters in the *L. majuscula* 3L genome with predicted natural product structures for each pathway. Figures constructed using Geneious 5.1 (Drummond et al. 2010).

The largest of these is an apparently intact 29 kb NRPS-dominated gene cluster on scaffold 52116 that is flanked by transposase genes on both sides (HQ696495). The adenylation domain active sites of the bi-modular NRPS protein are predicted to activate and incorporate proline and arginine. Surrounding the NRPS are genes for an arginosuccinate lyase, which may provide arginine for the NRPS adenylation domain, and a GCN5-related N-acetyltransferase (GNAT), which may acetylate arginine similarly to what has been observed with lysine acetylation in histones (Yang and Seto 2008). Although a GNAT motif was described as a component of a novel PKS chain initiation mechanism for the curacin A gene cluster (Gu et al. 2007), the gene context of this motif in the current cluster appears to be different and is thus more likely to be

involved in acetylation of a basic amino acid. Additionally, two adjacent phytanoyl-CoA dioxygenase (phyH)/L-proline 4-hydroxylase genes immediately proceed the NRPS gene, and possibly are involved in hydroxylation or halogenation of the proline residue (one hydroxyl group shown; Figure 5.2).

A second, potentially complete orphan gene cluster in the *L. majuscula* 3L genome, located on scaffold 52118, is approximately 20 kb in size and is flanked on the 5'-side by a transposase gene (HQ696496) (Figure 5.2). Three NRPS ORFs, one of which is bimodular, were predicted to encode isoleucine, lysine, tyrosine, and proline. The cluster also contains a sulfotransferase, suggesting the amino acid chain could be sulfated. An epimerase domain is present in the module incorporating lysine, and thus, as in almost all known cyanobacterial metabolites containing this basic amino acid, it is likely of D-configuration (Tidgewell et al. 2010). Another separate, mixed NRPS/PKS ORF is located 13 kb downstream of this latter NRPS cluster on the same scaffold, and appears to have an adenylation domain specific for either phenylalanine or tyrosine (HQ696497) (Figure 5.2). From catalytic activities predicted in the PKS portion, the amino acid is likely extended with acetate, the intermediate ketone reduced to an alcohol, and then released from the enzyme by a thioesterase. The majority of the genes surrounding this standalone NRPS/PKS gene appear to be involved in primary metabolism, and it is unclear whether they are involved in modifications of the NRPS/PKS product.

The remaining orphan clusters are on scaffolds 52120 and 52117 (Figure 5.2). Scaffold 52120 has bimodular and single module NRPS genes that have predicted adenylation specificities for alpha-aminoadipic acid, glutamine and proline,

respectively (HQ696498). A predicted thioesterase is present at the terminus of the second NRPS gene. These are flanked by genes encoding hypothetical proteins and proteins predicted to be involved in cytochrome c biosynthesis. The single NRPS ORF on scaffold 52117 encodes two modules (proline and threonine adenylation specificity), and as with the NRPS/PKS on scaffold 52118, the surrounding genes appear to be related to primary metabolism (HQ696499).

To determine if any of the above predicted 'cryptic metabolites' were expressed in cultures of *L. majuscula* 3L, water soluble and organic extracts were profiled by LC-MS (Figure 5.3), Matrix Assisted Laser Desorption Ionization (MALDI)-MS, and Fourier Transform (FT)-MS. Curacin A and carmabin were readily observed using all three techniques, and barbamide was detected using FT-MS. No mass/charge values ascribable to the unknown metabolites predicted above were detected.

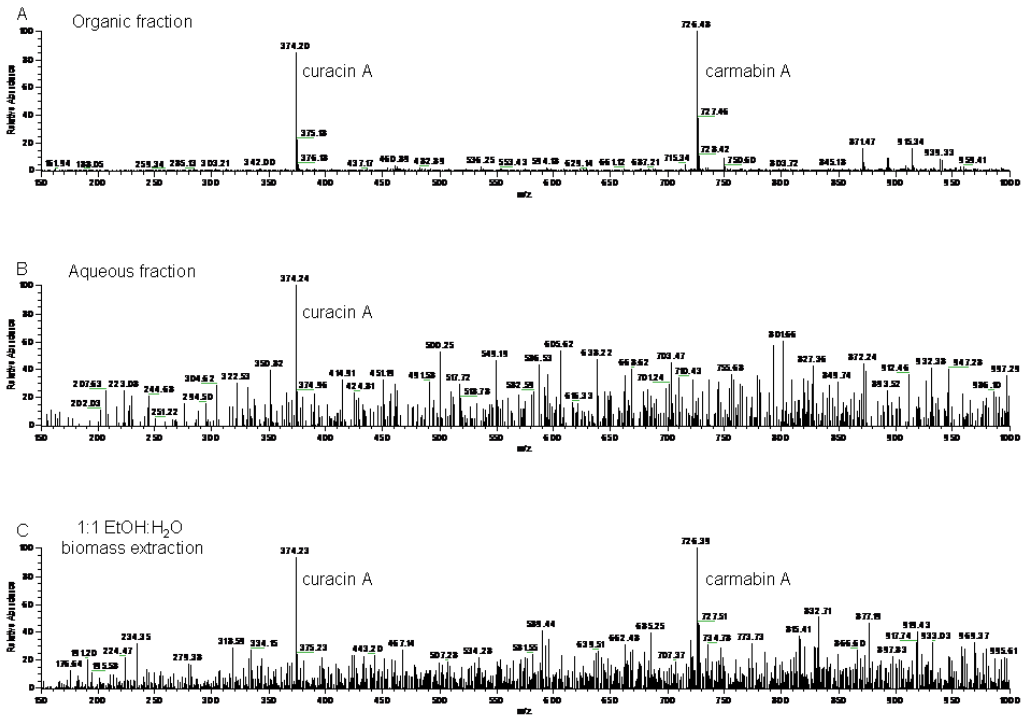


Figure 5.3: Organic extraction of *L. majuscula* 3L filaments for secondary metabolite profiling. Mass spectrometry analysis **(***m/z*) values from direct injection of *L. majuscula* 3L A) organic extract, B) aqueous extract, and C) 1:1 EtOH:H_2O biomass extract. Curacin A (*m/z* 347) was present in all three extracts, while carmabin A (*m/z* 726) was present in the organic and biomass extracts. Predicted *m/z* values of the molecules depicted in Figure 2 [gene cluster 1 (*m/z* 329.4), cluster 2 (*m/z* 533.7), cluster 3 (*m/z* 209.2), cluster 4 (*m/z* 358.4), cluster 5 (*m/z* 216.2)] were not observed.

Using *L. majuscula* 3L soluble protein extracted from cultured biomass, a proteomic analysis was also performed to determine relative expression levels of secondary metabolite biosynthetic proteins under normal culture conditions. Multidimensional Protein Identification Analysis (MudPIT; Washburn et al. 2001) yielded spectral counts from at least 2 of 4 technical replicates for 1043 proteins (Jones et al. 2011), which represented approximately 14% of the encoded proteins annotated in the *L. majuscula* 3L genome. The most readily detected proteins using MudPIT were

pigment-associated proteins, including phycocyanin subunits and phycobilisome proteins (approximately 2000 spectral counts per protein, Table 5.S4a). Spectral counts for nearly all of the proteins in the curacin A and barbamide pathways were quantified, as were several proteins predicted to be involved in carmabin biosynthesis (Table 5.S4b). Only one probable secondary metabolite protein from an orphan pathway was detected in at least two technical replicates (the free standing proline/threonine NRPS from scaffold 52117). None of the remaining orphan pathway proteins discovered during the genome annotation process were expressed to a measurable level by MudPIT analyses. Together with the absence of any of the predicted compounds in either the water soluble or the organic extracts from *L. majuscula* 3L, these data suggest that these orphan gene clusters are either expressed at low, nearly undetectable levels or not expressed at all under normal culture conditions.

5.3.3 Complex regulatory gene network of *L. majuscula* 3L

In light of the considerable number of natural products attributed to various strains of *Lyngbya majuscula* (nearly 200 reported metabolites), it was unexpected that *L. majuscula* 3L dedicates only 3% of its genome to secondary metabolism, which is significantly lower than that observed in marine actinobacteria such as *Salinispora* (9.9%; Udwary et al. 2007), and that only three NRPS/PKS type biosynthetic pathways larger than 30 kb were present. A large component of the *L. majuscula* 3L genome is devoted to regulatory genes involved in transcription and signal transduction. Marine *Lyngbya* strains grow in shallow tropical areas with frequent

exposure to diverse environmental stress factors such as desiccation during low tide or exposure to high fluxes of UV light. As noted previously, *Lyngbya* can usually be found living in close association with other cyanobacteria and heterotrophic bacteria. Even when growing separately from macroscopic assemblages, *L. majuscula* filaments retain a large number of associated bacterial cells on their polysaccharide sheath that are visible using DAPI staining (Gerwick et al. 2008). Therefore, a more careful evaluation of the *L. majuscula* 3L transcription and transduction genes was performed to evaluate the capacity of this organism for environmental adaptation and microbial communication.

Lyngbya majuscula 3L contains an unusual assortment of regulatory genes when compared to other cyanobacteria. Comparison of the 15 sigma factor genes annotated in *L. majuscula* 3L against the 9 well-characterized type I, II and III factors of *Synechocystis* sp. PCC 6803 (Imamura and Asayama 2009) revealed that *L. majuscula* has precisely 1 matching sigma factor for each of the 5 type I and II σ^{70} factors SigA-SigE. In addition, it possesses another 5 factors belonging to the type III class (Figure 5.4).

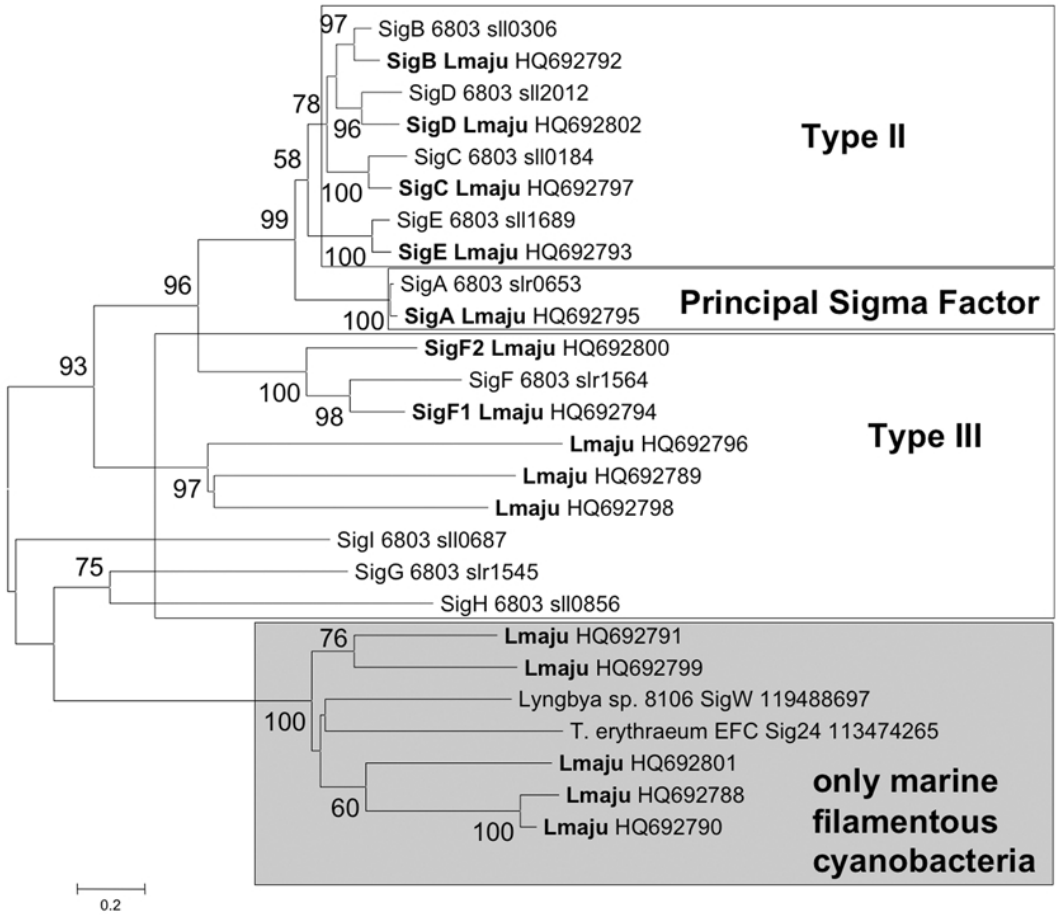


Figure 5.4: Phylogenetic relationships among *L. majuscula* sigma factors (Accession numbers provided for *L. mauscula* sigma factors). 15 *L. majuscula* sigma factors were compared to the 9 sigma factors from *Synechocystis* PCC 6803. The 3 major groups of cyanobacterial sigma factors are annotated and gene identifications are provided. Clade 4 occurs only in marine filamentous cyanobacteria and consists of 5 factors from *L. majuscula* and single representatives from *T. erythraeum* and *Lyngbya* sp. PCC8106 (Minimum evolution method). The optimal tree with the sum of branch length = 14.01963947 is shown. The percentage of replicate trees in which the associated taxa clustered together (1000 bootstrap replicates) are shown next to branches when ≥60 (tree to scale; branch lengths provided in same units as inferred evolutionary distances) (Tamura et al. 2007).

Of this latter group, 2 most closely resemble SigF of *Synechocystis* sp. PCC 6803, while the remaining 3 are distinct from all other known type III factors.

However, the most striking observation is the presence of the 5 additional sigma

factors which have no close homologue in any of the previously sequenced model cyanobacteria *Synechocystis* sp. PCC 6803, *Anabaena* sp. PCC 7120 or *Synechococcus* sp. PCC 7942. These sigma factors are between 257 and 563 residues in length and have an unusual domain structure. A domain with pronounced similarity to σ^{24} - type factors of the extracytoplasmic function (ECF) subfamily is located in the N-terminal half of the proteins. This is intriguing because the large and diverse group of ECF sigma factors plays a key role in adaptation to environmental conditions (Helmann 2002). The only other 2 related proteins occur as single-copy genes in the marine filamentous cyanobacteria *Trichodesmium* and *Lyngbya* sp. PCC8106 (Figure 5.4) and are annotated as SigW and Sig24 ECF-type sigma factors. However, the fact that *L. majuscula* 3L possesses 5 such factors suggests that a multitude of regulatory mechanisms could exist in this organism for potential interaction with the marine environment or associated microorganisms.

Moreover, the numbers and diversity of sigma factors which are global regulators of gene expression in bacteria appear higher in *L. majuscula* 3L than in most other cyanobacteria (i.e., *Anabaena* PCC 7120 has 11 sigma factors, while *Anabaena variabilis* ATCC 29413 and *Trichodesmium erythraeum* each have 7). In the longest open reading frame (HQ692799), this domain is preceded by and partially overlaps a SpvB domain (closest homolog: *Salmonella* virulence plasmid 65 kDa B protein, pfam03534). The C-terminal halves of these proteins have no close homologs in the NCBI or pfam databases, and they contain two 48 residue long repeats. The similarity among 4 of these 5 proteins in their C-terminal component suggests the presence of a novel protein domain that is presently uncharacterized. ECF sigma

factors are frequently co-transcribed with one or more downstream negative regulators which function as anti-sigma factors that bind and inhibit the cognate sigma factor (Helmann 2002). The *L. majuscula* 3L ECF-type sigma factors appear to belong to a class of sigma factors in which a regulatory domain has been fused to the protein. The recently identified sigma factor PhyR in *Methylobacterium extorquens* provides a possible paradigm for such a possibility (Francez-Charlot et al. 2009). In PhyR, an amino terminal ECF sigma factor-like domain is fused to a carboxyterminal receiver domain of a response regulator, suggesting PhyR can respond by sensing changes in the environment directly. A number of predicted short microRNA sequences were also found throughout the draft genome, including a cluster of mir-569 microRNA genes, suggesting that *L. majuscula* 3L use some level of posttranscriptional regulation. Although regulation of secondary metabolism in filamentous cyanobacteria has not been extensively evaluated, homologs of 2 proteins to those possibly involved in jamaicamide biosynthetic regulation (Chapter 2; Jones et al. 2009) were found to be expressed to detectable levels according to MudPIT analysis. Whether these latter proteins are involved in the regulation of curacin A or other secondary metabolite pathways in *L. majuscula* 3L remains to be determined.

5.3.4 Absence of nitrogen fixation in *L. majuscula* 3L

Perhaps the most unexpected finding in the *L. majuscula* 3L genome analysis was the lack of any genes involved in nitrogen fixation. Nitrogen availability is thought to be a major factor regulating primary production in shallow marine environments, and fixation of atmospheric nitrogen (N_2) by some prokaryotes,

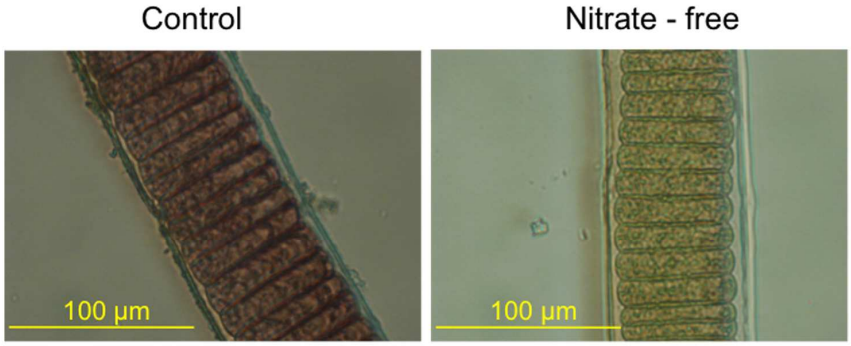
including cyanobacteria, is a critical source of bioavailable nitrogen for marine ecosystems worldwide (Berman-Frank et al. 2001). Several genera of cyanobacteria have been shown to fix nitrogen, including *Lyngbya* species (Jones 1990). *L. majuscula* nitrogen fixation has been detected previously by acetylene reduction assays (Jones 1990; Lundgren et al. 2003), and a dinitrogenase reductase (*nifH)* has been characterized from *L. majuscula* collected near Zanzibar in the Indian Ocean (Lundgren et al. 2003).

To independently investigate the capacity of *L. majuscula* 3L to fix nitrogen, the presence of nitrogenase genes was evaluated by PCR approaches as well as in several growth experiments performed in the absence of nitrate in the culture media. Using primers previously published to amplify *nifH* from *L. majuscula* (Lundgren et al. 2003), a PCR product was successfully amplified from genomic DNA isolated from *Oscillatoria nigro-viridis* 3LOSC, a cyanobacterial strain found growing in association with *L. majuscula* 3L in the field, but failed to amplify a product from *L. majuscula* 3L genomic DNA (Figure 5.S1). Additional experiments were performed to determine if *L. majuscula* 3L could grow and survive in the absence of a fixed nitrogen source. Single filaments grown in nitrate-free (98%) media were comparable in length to filaments grown in normal SW BG-11 after one week of growth; however, cell morphology and pigmentation were significantly altered in the nitrate-free samples. The length of cells grown in the nitrate free media visibly increased and the filaments changed from dark red to light green, and became colorless upon extended culture (Figure 5.5a). Similar phenotypes were observed when this experiment was repeated using larger scale 50 mL batch cultures. To assess nitrogen accumulation by

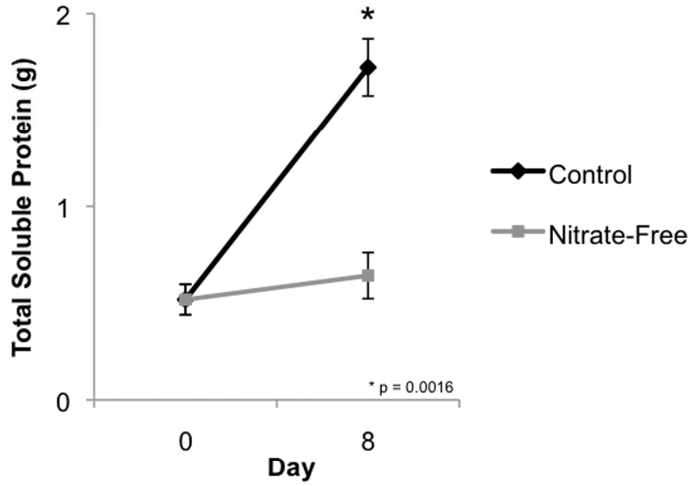
L. majuscula 3L in nitrate-free media, soluble protein was isolated from 50 mL nitrate-free batch cultures after 1 week of growth and compared to control cultures grown in control media for the same duration. During the course of the two independent experiments performed, control cultures significantly increased in protein content ($p = 0.0077$) whereas the nitrate free cultures showed no increase in protein content, indicating that *L. majuscula* 3L was unable to actively assimilate nitrogen from atmospheric dinitrogen (Figure 5.5b).

Figure 5.5: Absence of nitrogen fixation in *L. majuscula* 3L. A) Phenotypic changes of *L. majuscula* 3L when grown in nitrate-free media. Microscopic images are 400x. B) Total soluble protein in *L. majuscula* 3L grown with (control, diamonds) and without nitrate (squares) for eight days compared to day 0. Error bars represent SEM between replicate experiments (n=6 per treatment). C) ^{14}N incorporation into pheophytin *a* assayed by MALDI-TOF. *L. majuscula* 3L was grown in ^{15}N-labeled nitrate until nitrogen-containing molecules were fully labeled. Filaments were then grown in SW BG-11 media with ^{15}N-labeled nitrate, nitrate-free SW BG-11 media, and control SW BG-11 media with ^{14}N labeled nitrate (n=8) for 10 days and incorporation of ^{14}N into pheophytin *a* was measured by MALDI-TOF. Error bars are SEM.

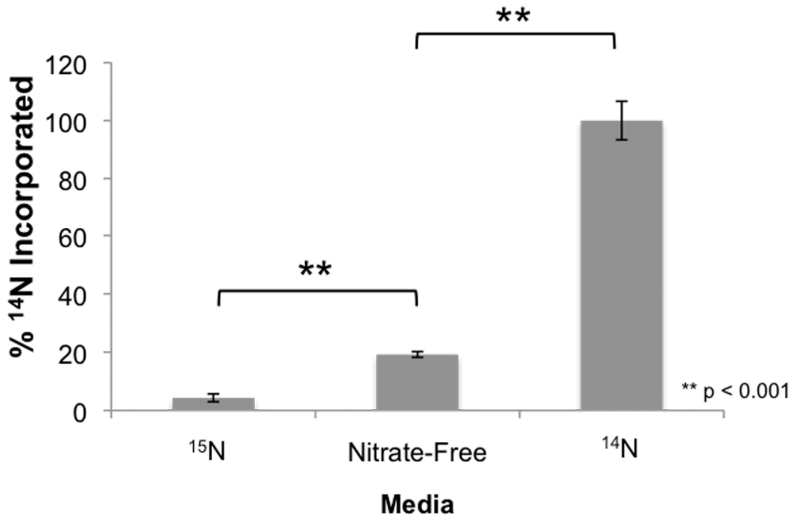
A. Phenotypic changes in Nitrate-Free Media



B. Total Soluble Protein



C. % ^{14}N Incorporation into Pheophytin *a*



The ability of *L. majuscula* 3L to assimilate atmospheric nitrogen was also explored through ^{15}N isotope feeding experiments. *L. majuscula* 3L filaments were grown in media containing ^{15}N-labeled sodium nitrate for approximately 21 days until nitrogen-containing compounds were fully labeled with this heavy isotope, as assessed by MALDI-TOF mass spectrometry of the metabolome (see methods). The fully ^{15}N-labeled filaments were then grown in nitrate-free media for 10 days, and the incorporation of the prevailing natural ^{14}N isotope from atmospheric N_2 into nitrogen-containing compounds was evaluated using MALDI-MS. The shift from ^{15}N to ^{14}N was calculated for pheophytin *a*, a chlorophyll breakdown product that readily ionizes during MALDI analysis of *Lyngbya* filaments (Figure 5.5c; Esquenazi et al. 2011). In the absence of a nitrogen source, 19% (± 0.98%) of the pheophytin *a* shifted to lighter mass by incorporation of ^{14}N. Controls remaining in ^{15}N-labeled nitrate media showed a shift of 4% (± 1.35%) to lighter mass, and controls grown in regular ^{14}N nitrate media shifted to lighter mass by 99% (± 6.65%).

The 19% shift observed in the nitrate-free media may be due to other trace amounts of nitrogen in the media, or may represent recycling of internal nitrogen stores that were not labeled during the incubation with ^{15}N-nitrate. A recent study examining proteomic changes in the cyanobacterium *Synechocystis sp.* 6803 in response to various environmental stresses, including low nitrogen, found that in addition to switching to alternative carbon and nitrogen assimilation pathways, *Synechocystis* can access internal carbon and nitrogen stores based on upregulation of proteins associated with cyanophycin breakdown and downstream arginine catabolism (Wegener et al. 2010). To provide nitrogen and carbon to the cell, cyanophycin, a

storage polymer of L-aspartic acid and L-arginine found in most cyanobacteria, is broken down into arginine and aspartic acid by cyanophycinase. Arginine and aspartic acid can be subsequently broken down by arginine decarboxylases and agmantinase and/or arginase (Wegener et al. 2010). The *L. majuscula* 3L genome contains genes for a cyanophycin synthetase (HQ692807), cyanophycinase (HQ692806), two arginine decarboxylases (HQ692803 and HQ692804), and one agmantinase (HQ692805). The second arginine decarboxylase is in close proximity to the agmantinase on scaffold 52022, supporting their suggested role in cyanophycin recycling. The presence of these genes in the *L. majuscula* 3L genome provides evidence that *L. majuscula* is capable of obtaining nitrogen from cellular storage, and this capacity to use internal nitrogen stores in low nitrogen environments could explain the 19% ^{14}N shift observed in the MALDI growth experiments. The loss of pigmentation observed under nitrate-free conditions is also consistent with the observations of down regulation of photosystem proteins in response to environmental stresses (Wegener et al. 2010). Collectively, these phenotypic and growth assessments strongly suggest that *L. majuscula* is unable to fix atmospheric nitrogen, and that under nitrate-free growth conditions, recycles nitrogen from storage proteins such as cyanophycin.

The apparent discrepancy between these investigations with *L. majuscula* 3L and past demonstrations of nitrogen fixation in *Lyngbya majuscula* may reflect strain differences, or possibly, the different criteria used in assigning the taxonomy of these cyanobacteria (Jones 1990; Lundgren et al. 2003). A 16S rRNA phylogenetic assessment of *L. majuscula* 3L places this strain in the marine *Lyngbya* lineage, as recently reported (Engene et al. 2010), which is phylogenetically distinct from

freshwater *Lyngbya* strains. It is also conceivable that previous nitrogen fixation experiments with *L. majuscula* wherein the organism was identified solely by morphology may have actually investigated other morphologically similar but unrelated genera. For example, *Oscillatoria*, another cyanobacterial genus reported to fix nitrogen (Stal and Krumbein 1981), is morphologically very similar to *Lyngbya*, and can be easily misidentified without the taxonomic support provided by phylogenetics.

5.4 Conclusions

Since the first evaluation of their natural products 40 years ago, tropical filamentous marine cyanobacteria are now established as rich sources of novel bioactive molecules. *L. majuscula* 3L was selected for genome sequencing because it is a strain that has been successfully cultivated in the Gerwick laboratory for 15 years and has been studied extensively for its natural products and biosynthetic pathways. The genome sequence contained intact gene clusters for curacin A and barbamide, consistent with previous reports (Chang et al. 2002; Chang et al. 2004), as well as genes in good agreement with carmabin biosynthesis. However, no other gene clusters above 30 kb were evident in the draft genome, and only 5 other PKS and/or NRPS pathways were detected. Organic extracts from *L. majuscula* 3L and expression analysis of the soluble proteome revealed that these unknown pathways are either expressed at low or undetectable levels under typical culture conditions. Thus, the more than 200 metabolites reported from this species are likely due to a very large number of different chemical strains or chemotypes. Moreover, the processes of

horizontal gene transfer or evolutionary pathway divergence, as suggested for the curacin A and the jamaicamide pathways from *L. majuscula* (Gu et al. 2009), or the apratoxins from strains of *Lyngbya bouillonii* (Tidgewell et al. 2010), are likely responsible for this impressive molecular diversity.

The discovery of regulatory genes conferring an enhanced ability for microbial interactions and/or environmental adaptation and lack of traditional nitrogen fixation pathways were additional unexpected findings in the *L. majuscula* 3L genome. *Lyngbya* is almost always found growing in close association with other cyanobacteria, diverse microorganisms, and invertebrates in the field. A wide variety of heterotrophic bacteria remain on the surface of the polysaccharide sheath even after extensive purification of field isolates. The relationship between *Lyngbya* and these associated organisms remains unclear, but the possibility of complex interactions taking place among them is a fascinating focus for future research and may also explain the large number and variety of ECF sigma factors and other regulatory genes described above. The finding that nitrogen fixation does not occur in this *L. majuscula* strain is in direct contrast to previous reports, but may be another indication that finer scale phylogenetic relationships of marine filamentous cyanobacteria need to be better defined and used in taxonomic classification. Among nitrogen fixing organisms, cyanobacteria form a monophyletic group; however, within cyanobacteria, the capacity to fix nitrogen appears to be polyphyletic, suggesting multiple gene losses of nitrogen fixation genes over time, horizontal gene transfer causing independent introduction of nitrogen fixation genes, or a combination of both (Swingley et al. 2008; Bolhuis et al. 2010). Recent examination of nitrogen fixation gene evolution

implies a complex history of both gene loss and HGT events (Swingley et al. 2008; Bolhuis et al. 2010). The current *L. majuscula* 3L genomic data does not reveal gene loss events (via the presence of pseudogenes or regions in the genome where there was possible loss of the entire *nif* gene cluster), but as additional sequence data for diazotrophic and non-diazotrophic marine filamentous cyanobacteria becomes available, the evolutionary history of nitrogen fixation in this group can be better understood. Since *L. majuscula* strains previously found to fix nitrogen were identified using morphological techniques, it is difficult to determine how closely related these may be to *L. majuscula* 3L, and whether the ability to fix nitrogen is more of an exception than a rule for this genus.

The first sequencing of a marine *Lyngbya* species presented here clearly accentuates the need for genomic study of additional *Lyngbya* strains. Recent phylogenetic assessment has revealed a significant degree of ambiguity in *Lyngbya* taxonomy (Engene et al. 2010). At least three separate lineages have been described from different environments. Specific natural products isolated from *Lyngbya* may be a more effective means of delineating these cyanobacterial strains, as has been proposed for marine actinomycetes (Jensen et al. 2007). Additional genome sequencing of other *Lyngbya* collections will be required to better understand how the traits described here compare between species, strains and across geographic locations.

5.5 Materials and Methods

5.5.1 Genome accession information

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AEPQ00000000. The version described in this paper is the first version, AEPQ01000000.

5.5.2 Culturing Techniques

The original collection of *Lyngbya majuscula* 3L in 1993 was conducted near the CARMABI Research Station on the island of Curaçao, Netherlands Antilles. Pan (10 L) and Erlenmeyer flask (1 L) cultures have since been maintained (Gerwick et al. 1994; Rossi et al. 1997) at Oregon State University and Scripps Institution of Oceanography, UCSD in SW BG-11 (Castenholz 1988) at 28°C, under 16 h light / 8 h dark cycles at approximately 5 μE m^{-2} s^{-1}.

5.5.3 Preparation of *Lyngbya majuscula* 3L DNA for sequencing

For Sanger sequencing, DNA from laboratory cultured *L. majuscula* 3L was extracted and used in creation of two separate fosmid libraries (CopyControl Fosmid Library Production Kit, Epicentre) using standard protocols, and these libraries were sequenced at the Max – Planck Institute for Molecular Genetics. For 454 sequencing, single cells from a *L. majuscula* 3L were liberated from filaments using a razor blade and captured with a capillary tube using a microscope. The cells were lysed and their DNA amplified using the Repli-g minikit (Qiagen). Random primers were added (Frias-Lopez et al. 2004) and MDA of genomic DNA was performed. The amplified

DNA was screened for purity by PCR using cyanobacterial 16S rRNA gene primers (Nübel et al. 1997) and cloning (TOPO-TA, Invitrogen). and submitted for 454 sequencing (J.C. Venter Institute, La Jolla, CA). A total of 223 Mb were sequenced and the average length of the sequence reads was 387 bp.

5.5.4 Genome assembly, binning, and annotation

136,560 Sanger sequences and 576,388 pyrosequencing (FLX 454) reads were combined into a single hybrid assembly using Celera Assembler software, version 5.4 (Myers et al. 2000). The mer overlapper program setting was used to achieve optimal integration of Sanger and 454 data sets (Miller et al. 2008). Relative numbers of Sanger and 454 reads incorporated into each scaffold produced by the combined assembly were parsed from Celera Assembler output using a custom perl script. Open reading frames and amino acid sequences were predicted from all scaffolds using the gene finding program Metagene (Noguchi et al. 2006). Predicted proteins were evaluated for phylogenetic relatedness to known sequences in NCBI Genbank nr using the DarkHorse program, version 1.3, with a threshold filter setting of 0.1 (Podell et al. 2008). Only matches with alignments covering at least 70% of total query length, and BLASTP e-value scores of $1e^{-5}$ or better were included in the DarkHorse analysis.

Assembled scaffolds were placed in one of three phylogenetic categories (*Lyngbya*, non-*Lyngbya*, or ambiguous) based on manual assessment of several different, independent parameters. For each scaffold, number and alignment quality of predicted protein matches to Genbank sequences associated with phylum Cyanobacteria were compared to proteins from other taxonomic groups based on

DarkHorse analysis of the Genbank nr BLASTP search. 16S rRNA matches to known microbial sequences were identified by BLASTN search against the GreenGenes reference database, requiring a minimum alignment length of 200 nucleotides, and e-value of 1^{e-7} or better (DeSantis et al. 2006). For scaffolds longer than 5,000 nucleotides, inclusion of reads from both Sanger and 454 libraries, percent G+C nucleotide composition, and coverage depth were also considered in determining scaffold origin. Scaffolds classified as most likely belonging to *Lyngbya* were annotated using the Integrated Microbial Genome Expert Review (IMG-ER) service of the Joint Genome Institute (Markowitz et al. 2009). Further annotation was performed using IMG-ER database tools.

5.5.4.1 Annotation of secondary metabolite pathways

To identify secondary metabolite pathways, the genome was queried for genes annotated as polyketide synthase or nonribosomal peptide synthetase-related genes in addition to BLAST queries using previously sequenced secondary metabolite pathways. Molecule predictions for the orphan pathways were based on adenylation domain specificities predicted using NRPSpredictor (Rausch et al. 2005) and other domains present in the pathways. Secondary metabolite pathway figures were generated using Geneious bioinformatics software (Drummond et al. 2010).

5.5.5 Organic extraction of *L. majuscula* 3L filaments for secondary metabolite profiling

Lyngbya majuscula 3L filaments (140 mg dry weight) were obtained from a

pan culture that had been growing under standard culture conditions for approximately 4 months. The filaments were extracted twice in 2:1 dichloromethane (DCM):methanol (MeOH), and these extractions were filtered through cheesecloth and Whatman filter paper and combined. This crude extract was then extracted with MilliQ water in a separatory funnel to generate organic and aqueous fractions. The *L. majuscula* filaments were extracted again in 1:1 ethanol (EtOH):water with stirring, and this extract was also filtered through cheesecloth and filter paper. Portions of each of these three extracts ("organic", "aqueous", and "1:1 EtOH:H_2O biomass extraction") were evaporated to dryness before being resuspended in MeOH and passed through a 0.2 um syringe filter. The extracts were profiled using direct injection (positive ion mode) on a Thermo-Finnigan LCQ Advantage Max mass spectrometer at 1 or 10 mg/ml, as well as by MALDI-MS as previously described (Esquenazi et al. 2011). Separate organic *L. majuscula* 3L extracts were profiled using Fourier-Transform (FT)-MS (Brown et al. 2005).

5.5.6 Proteomic analysis

5.5.6.1 Preparation of soluble protein

Lyngbya majuscula 3L tissue was harvested from a pan culture using forceps and measured by displacement in a 50 ml Falcon tube (4 ml of cultured biomass) containing lysis buffer (10 mM Tris HCl, pH 7.5, 150 mM NaCl). Prior to cell lysis, the buffer was treated with a broad range protease inhibitor (Complete, EDTA Free, Roche). The *L. majuscula* filaments were sonicated repeatedly on ice, and the resulting material was centrifuged at 4°C at 14,000 rpm to pellet cellular debris and insoluble

protein. The supernatant containing soluble protein was collected after centrifugation, and flash frozen in liquid N_2 before use in proteomic analysis.

5.5.6.2 Liquid chromatography – mass spectrometry (MudPIT) analysis

100 µgs of protein from the preparation described above were TCA precipitated. Precipitated proteins were resuspended in 8M urea 50 mM Tris pH 8.0 and digested in the presence of ProteasMAX using the suggested protocol (Promega). Peptides were acidified to a final concentration of 5% formic acid and 25 µgs of peptides were bombed loaded onto a biphasic (strong cation exchange/reverse phase) capillary column for multidimensional protein identification analysis (MudPIT). Peptides were separated and analyzed by 2D-LC separation in combination with tandem MS as previously described (Washburn et al. 2001). Peptides were eluted in an 11-step salt gradient and data were collected in an ion trap mass spectrometer (ThermoFisher LTQ) set in a data-dependent acquisition mode with dynamic exclusion turned on (90 s). Each full MS survey scan was followed by 7 MS/MS scans. Spray voltage was set to 2.75 kV and the flow rate through the column was 0.20 µL/min.

5.5.6.3 Analysis of MS data

RAW files were generated from mass spectra using XCalibur version 1.4, and ms^2 spectra data extracted using RAW Xtractor (version 1.9.1) which is publicly available (http://fields.scripps.edu/?q=content/download). Ms^2 spectral data were searched using the SEQUEST algorithm (Version 3.0) against a custom made database

containing 30,414 sequences (7479 *Lyngbya majuscula* 3L and 22935 human IPI protein sequences) that were concatenated to a decoy database in which the sequences for each entry in the original database was reversed (Eng et al. 1994). In total the search database contained 60,828 protein sequence entries (30,414 real sequences and 30,414 decoy sequences). SEQUEST searches allowed for oxidation of methionine residues (16.0 Da), static modification of cysteine residues (57.0 Da-due to alkylation), no enzyme specificity, and a mass tolerance set to ±1.5 Da for precursor mass and ±0.5 Da for product ion masses. The resulting ms^2 spectra matches were assembled and filtered using DTASelect2 (version 2.0.27). For this analysis, tryptic, half-tryptic and fully-tryptic peptides were each individually evaluated using the DTASelect2 software (Tabb et al. 2002). In each of these sub-groups the distribution of Xcorr and DeltaCN values for a direct (to the direct database) and decoy (reversed database) were separated by quadratic discriminate analysis. Outlier hits in the two distributions were removed. Spectral matches were retained with Xcorr and deltaCN values that produced a maximum peptide false positive rate of 1% that was derived from the frequency of matches to the decoy reverse database (number of decoy database hits/number of filtered peptides identified× 100). This value is calculated by the DTASelect2 software. In addition, a minimum peptide length of seven amino acids residues was imposed and protein identification required the matching of at least two peptides per protein. Such criteria resulted in the elimination of most decoy database hits. In this dataset the identification of non-tryptic peptides included half-tryptic peptides from the N- and C- terminus of the identified proteins. Other non-tryptic peptides that were identified may represent endogenous activities of cellular

proteases or peptides generated by in-source fragmentation.

5.5.7 Phylogenetic analysis of regulatory proteins

Fifteen *Lyngbya majuscula* sigma factors were compared to the 9 sigma factors from *Synechocystis* PCC 6803. The phylogenetic relationships were inferred using the Minimum Evolution method and 1000 bootstrap replicates. The evolutionary distances were computed using the Poisson correction method and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the Close-Neighbor-Interchange (CNI) algorithm at a search level of 1. The Neighbor-joining algorithm was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 126 positions in the final dataset. Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007).

5.5.8 Nitrogen fixation methods

5.5.8.1 Amplification and sequencing of *nifH*

Genomic DNA was isolated from *L. majuscula* 3L and *Oscillatoria nigro-viridis* 3L-OSC using the Promega Wizard genomic DNA purification kit (Madison, WI) and quantified using an ND-1000 NanoDrop spectrophotometer (Thermo Scientific, Wilmington, DE). Primers previously used to identify and sequence *nifH* from *L. majuscula* (PN1: 5’ –CGTCACGGTCAAAGAATCAT – 3’; PN2: 5’-ACACCACCAGCATGAGCATA-3’; Lundgren et al. 2003) were used in 25 μL PCR reactions containing 1X Promega PCR Master Mix, 0.4 μM of each primer, and 1 μL

of genomic DNA template. The following cycling parameters were used for PCR: Initial denaturation at 95°C for 1 min, 30 cycles of 95°C for 30s, 48°C for 30s and 72°C for 1.5 min, and a final extension at 72°C for 5 min. General 16S primers for cyanobacteria (Nübel et al. 1997) were used as a positive control. PCR products from *O. nigro-viridis* 3L-OSC were purified using Qiagen Minelute PCR purification kit, cloned into a TOPO TA cloning vector (Invitrogen) for sequencing, and sequenced by Seqxcel (San Diego, CA).

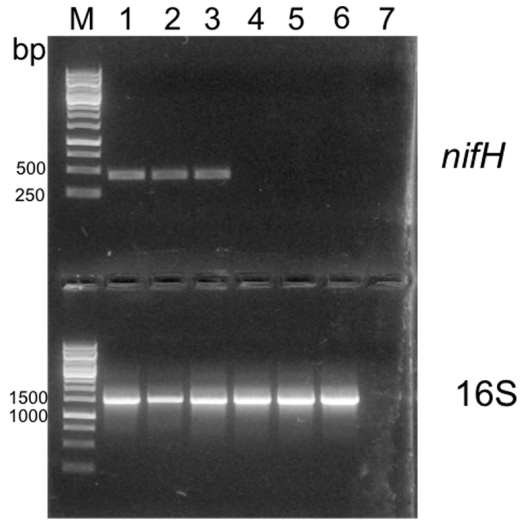


Figure 5.S1: Independent investigation of *nifH* in *L. majuscula* 3L. To independently assess the presence of *nifH* genes in *L. majuscula* 3L, primers previously published for *L. majuscula nifH* genes were used in PCR to amplify *nifH* from genomic *L. majuscula* 3L DNA. *O. nigo-viridis* 3L-OSC was used as a positive control for the *nifH* PCR reaction, and 16S primers were used as an overall positive control. Lanes: M: 1 Kb marker; 1-3: *O. nigo-viridis* 3L-OSC; 4-6: *L. majuscula* 3L; 7: no template control.

5.5.8.2 Growth of *L. majuscula* in nitrate-free media

5.5.8.2.1 Growth of single filaments

Single *L. majuscula* 3L filaments were grown in 24-well plates in SW BG-11 media at 28°C on a 16:8 h light:dark cycle. Twelve wells contained regular SW BG-11 media, and twelve wells contained nitrate-free SW BG-11 media. Filaments were measured using ImageJ (Abramoff et al. 2004) at Day 0 and Day 7. Filament lengths at Day 7 were compared to their respective lengths at Day 0 to approximate growth and cell size.

5.5.8.2.2 Growth of batch cultures

A total of sixteen 50 mL batch cultures were set up in either control SW BG-11 or nitrate-free SW BG-11 media (8 per set) with equal amounts of starting material (approximately 0.05 g wet weight) under the same culturing conditions as above. Triplicate samples of the starting material were sampled to represent Day 0. To assess protein content as a metric for growth, samples were placed in 2 mL of protein lysis buffer (10 mM Tris pH 7.5, 150 mM NaCl, and Roche Complete EDTA-free protease inhibitors) and homogenized using a probe sonicator. Samples were centrifuged at 13,000 x *g* for 10 min and soluble protein was recovered. Total soluble protein was calculated using the Pierce BCA assay according to manufacturer's protocols and using BSA as protein standard (Rockford, Il). At Day 8, total biomass was harvested from triplicate control and nitrate-free cultures, and protein measurements were done in the same manner. Statistical analyses were performed on the two replicate experiments for a total of 6 biological replicates (3 per experiment). Student's t-test

was used to analyze significant differences between the experimental treatments using JMP software Version 9, SAS Institute Inc., Cary, NC. The standard error between experimental replicates was calculated by the following equation: $\sqrt{[(sd_1^2/n_1) + (sd_2^2/n_2)]}$. Microscopy was performed using an Olympus 1X51 microscope.

5.5.8.3 ^{15}N assimilation studies

L. majuscula 3L filaments were grown in SW BG-11 media containing ^{15}N-labeled nitrate for approximately 21 days until the majority of nitrogen-containing molecules contained labeled nitrogen as assessed by MALDI-TOF (Esquenazi et al. 2011). Filaments were then grown in 24-well plates in either SW BG-11 media containing ^{15}N-labeled nitrate (negative control), regular SW BG-11 media (positive control), or nitrate-free SW BG-11 media (8 wells per treatment). Filaments were sampled at Day 0 and Day 10 and frozen until subsequent MALDI-TOF analysis. Approximately 1 μL of MALDI matrix solution (Per 1 mL: 35 mg α-cyano-4-hydroxycinnamic acid (CHCA), 35 mg 2,5-Dihydroxybenzoic acid (DHB) (Universal MALDI matrix, Sigma Aldrich), 750 μL acetonitrile, 248 μL milliQ H_20, 2 μL TFA) per 0.1 μg of biomass was mixed in a tube or well. After 20-30 seconds, 1 μL of this crude matrix solution was deposited on a well (spot) of the Bruker Microflex MSP 96 Stainless Steel Target Plate. After each spot had dried at room temperature, the plate was analyzed using a Bruker Microflex MALDI-TOF mass spectrometer equipped with flexControl 3.0 as described (Esquenazi et al. 2011).

5.6 Supporting Text

5.6.1 16S rRNA genes in the assembly and taxonomic heterogeneity of scaffolds

16S rRNA sequences identified in the assembled scaffolds include several representatives from Alphaproteobacteria, Gammaproteobacteria, and Bacteriodetes, but only *Lyngbya* from phylum Cyanobacteria (Table 5.S3). Most of the non-*Lyngbya* matches are comprised exclusively of reads from only one of the two libraries (100% or 0% from 454), suggesting they were present in only one of the two cultures used to obtain sequencing data. The exception to this pattern was scaffold 46429, which most closely matched an uncultured Alpha-proteobacterium isolated from a diseased Caribbean coral. (Sunagawa et al. 2009).

After the binning procedure that identified 161 scaffolds being of *Lyngbya* sources, 2,161 scaffolds were categorized as most likely derived from non-*Lyngbya* sources. These scaffolds contain multiple predicted proteins matching non-cyanobacteria Genbank entries. Most of these scaffolds are comprised exclusively of reads from either Sanger or 454 libraries, but not both types of reads co-assembled in the same scaffold. A total of 3896 scaffolds were left unclassified. Most of these sequences are very short, with coverage too low to determine whether unusual G+C composition or absence of reads from one of the two sequencing libraries might be due to random statistical variation. Few of the unclassified scaffolds are large enough to provide even one open reading frame with reliable database matches. Some contain a single predicted peptide or peptide fragment closely matching a single database protein, but this evidence alone was not considered sufficient to determine scaffold origin, given the strong possibility of horizontal gene transfer between unrelated

bacteria in mixed cultures.

Overall taxonomic heterogeneity of scaffolds generated in the combined assembly was assessed by binning scaffolds according to G+C content in 1% increments, then adding together the lengths of all scaffolds in each bin to obtain total assembled nucleotides (Figure 5.S2).

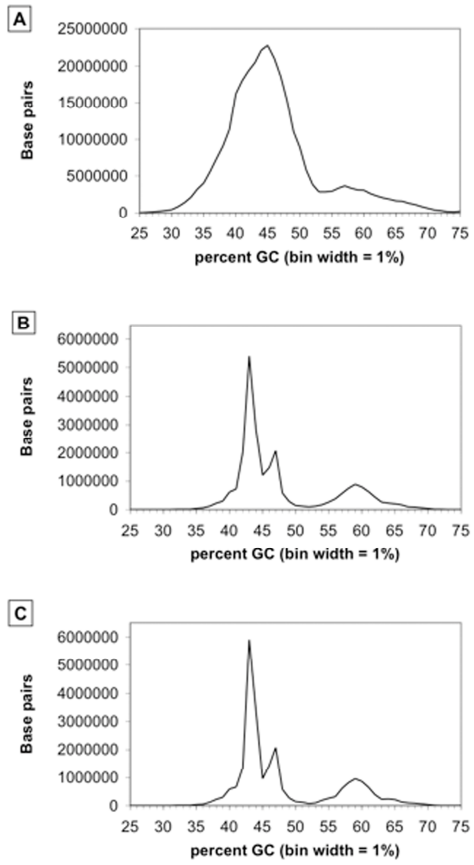


Figure 5.S2: Percent G+C content of combined Sanger-454 assembly for A) raw reads, B) assembled contigs, and C) scaffolds created by joining contigs with gaps.

Longer sequences originating from the same organism are expected to have a fairly uniform G+C percentage, consistent with species-specific nucleotide

composition generally characteristic of genomic DNA (Karlin et al. 1995). This pattern is observed in Figure 5S2, as raw reads (Figure 5.S2a) are first recruited to contigs (Figure 5.S2b), then to slightly longer scaffolds by joining adjacent contigs with gaps (Figure 5.S2c). The most dominant peak in Figure 5.S2c, at 43% G+C, is associated with *Lyngbya*, containing co-assembled reads from both Sanger and 454 libraries. Most scaffolds in the 47% G+C peak are comprised almost exclusively of 454 reads, in many cases without any contribution from Sanger reads. On the other hand, most scaffolds in the broad 59% G+C peak contain exclusively Sanger reads, and lack 454 reads. These results, in combination with 16S rRNA data in Table 5.S3, support the premise that the two cultures used to obtain these sequencing libraries contained different types of impurities, and that a core *Lyngbya* genome can be recovered from a metagenomic environment by combining sequence data from biological replicates.

5.7 Acknowledgements

This work was supported by the Network of Excellence in Marine Genomics Europe, the NIH (CA108874), and California Sea Grant (R/NMP-103EPD). Another author on this manuscript (EAM) is a San Diego IRACDA Postdoctoral Fellow supported by NIH Grant GM06852. Dr. Hyukjae Choi (assistance with LC-MS analyses) and Roland Kersten (assistance with FT-MS) are acknowledged for their support. Chapter 5, in full, is reprinted as it appears in Jones, Adam C.; Monroe, Emily A.; Podell, Sheila; Hess, Wolfgang R.; Klages, Sven; Esquenazi, Eduardo; Niessen, Sherry; Hoover, Heather; Rothmann, Michael; Kube, Michael; Reinhardt,

Richard; Lasken, Roger S.; Yates III, John R., Burkart, Michael D.; Allen, Eric E.; Dorrestein, Pieter C.; Gerwick, William H.; Gerwick, Lena. Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proceedings of the National Academy of Sciences* 2011, doi:10.1073/pnas.1101137108. The dissertation author is one of two primary investigators and authors of this paper (along with EA Monroe). For this project, the author was involved in genome annotation, natural product gene cluster elucidation, organic extractions, proteomic analysis, experimental design, data analysis, and writing the manuscript.

5.8 References

Abramoff MD, Magelhaes PJ, Ram SJ (2004) Image Processing with ImageJ. *Biophotonics Int* 11:36-42.

Berman-Frank I, Lundgren P, Chen Y-B, Küpper H, Kolber Z, Bergman B, Falkowski P (2001) Segregation of nitrogen fixation and oxygenic photosynthesis in the marine cyanobacterium *Trichodesmium*. *Science* 294:1534-1537.

Bolhuis H, Severin I, Confurious-Guns, V, Wollensien UIA, Stal LJ (2010) Horizontal gene transfer of the nitrogen fixation gene cluster in the cyanobacterium *Microcoleus chthonoplastes*. *ISME J* 4:121-130.

Brown SC, Kruppa G, Dasseux L (2005) Metabolomics applications of FT-ICR mass spectrometry. *Mass Spec Rev* 24:223-231.

Carmichael WW (2001) Health effects of toxin-producing cyanobacteria: "The cyanoHABs." *Human Ecol Risk Assess* 7:1393-1407.

Castenholz RW (1988) Culturing of cyanobacteria. *Meth Enzymol* 167**:** 68-93.
Chang Z, 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* 67:1356-1367.

Chang Z, Flatt P, Gerwick WH, Nguyen V-A, Willis CL, Sherman DH (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* 296:235-247.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Anderson GL (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 72:5069-5072.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.1, Available from http://www.geneious.com.

Edwards DJ, Gerwick WH (2004) Lyngbyatoxin biosynthesis: Sequence of biosynthetic gene cluster and identification of a novel aromatic prenytransferase. *JACS* 126:11432-11433.

Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH (2004) Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem Biol* 11:817-833.

Eng JK, McCormack AL,Yates III JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database, *J Am Soc Mass Spectrom* 5: 976-989.

Engene N, Coates RC, Gerwick WH (2010) 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46:591-601.

Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226-5231.

Francez-Charlot A, Frunzke J, Reichen C, Ebneter JZ, Gourion B, Vorholt JA (2009) Sigma factor mimicry involved in regulation of general stress response. *Proc Nat Acad Sci USA* 106:3467-3472.

Frias-Lopez J, Klaus JS, Bonheyo GT, Fouke BW (2004) Bacterial community associated with black band disease in corals. *Appl Environ Microbiol* 70:5955-5962.

Gerwick WH, Coates RC, Engene N, Gerwick L, Grindberg RV, Jones AC, Sorrels CM (2008) Giant marine cyanobacteria produce exciting potential pharmaceuticals. *Microbe* 3:277-284.

Gerwick WH, Proteau PJ, Nagle DG, Hamel E, Blokhin A, Slate DL (1994) Structure of curacin A, a novel antimitotic, antiproliferative, and brine shrimp toxic natural product from the marine cyanobacterium *Lyngbya majuscula*. *J Org Chem* 59:1243-1245.

Gu L, Geders TW, Wang B, Gerwick WH, Håkansson K, Smith JL, Sherman DH (2007) GNAT-like strategy for polyketide chain initiation. *Science* 318:970-974.

Gu L, Wang B, Kulkarni A, Geders TW, Grindberg RV, Gerwick L, Håkansson K, Wipf P, Smith JL, Gerwick WH, Sherman DH (2009) Metamorphic enzyme assembly in polyketide diversification. *Nature* 459:731-735.

Helmann JD (2002) The extracytoplasmic function (ECF) sigma factors. *Adv Microb Physiol* 46:47-110.

Hooper GJ, Orjala J, Schatzman RC, Gerwick WH (1998) Carmabins A and B, new lipopeptides from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 61: 529-533.

Imamura S, Asayama M (2009) Sigma factors for cyanobacterial transcription. *Gene Regul Syst Bio* 3:65-87.

Jensen PR, Williams PG, Oh D-C, Zeigler L, Fenical W (2007) Species – specific secondary metabolite production in marine actinomycetes of the genus *Salinispora*. *Appl Environ Microbiol* 73:1146-1152.

Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247.

Jones AC, Monroe EA, Eisman EB, Sherman DH, Gerwick WH (2010) The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* 27:1048-1065.

Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S, Hoover H, Rothmann M, Kube M, Reinhardt R, Lasken RS, Yates III JR, Burkart MD, Allen EE, Dorrestein PC, Gerwick WH, Gerwick L (2011). Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci USA* doi:10.1073/pnas.1101137108.

Jones K (1990) Aerobic nitrogen fixation by *Lyngbya* sp., a marine tropical cyanobacterium. *Eur J Phycol* 25:287-289.

Karlin S, Burge C (1995) Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet* 11:283-290.

Kuffner IB, Paul VJ (2004) Effects of the benthic cyanobacterium *Lyngbya majuscula* on larval recruitment of the reef corals *Acropora surculosa* and *Pocillopora damicornis*. *Coral Reefs* 23:455-458.

Lundgren P, Bauer K, Lugomela C, Söderbäck E, Bergman B (2003) Reevaluation of the nitrogen fixation behavior in the marine non-heterocystous cyanobacterium *Lyngbya majuscula*. *J Phycol* 39:310-314.

Markowitz VM, Mavromatis K, Ivanova NN, Chen I-MA, Chu K, Kyrpides NC (2009) IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 25:2271-2278.

Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, Johnson J, Li K, Mobarry C, Sutton G (2008) Aggressive assembly of pyrosequencing reads with mates. *Bioinformatics* 24:2818-2824.

Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KHJ, Remington KA, Anson EL, Bolanos RA, Chou H-H, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng X, Rubin GM, Adams MD, Venter JC (2000) A whole-genome assembly of *Drosophila*. *Science* 287:2196-2204.

Noguchi H, Park J, Takagi T (2006) MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* 34:5623-5630.

Nübel U, Garcia-Pichel F, Muyzer G (1997) PCR primers to amplify 16s rRNA genes from cyanobacteria. *Appl Environ Microbiol* 63:3327-3332.

Orjala J, Gerwick WH (1996) Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 59:427-430.

Paul VJ, Thacker RW, Banks K, Golubic S (2005) Benthic cyanobacterial bloom impacts the reefs of South Florida (Broward County, USA). *Coral Reefs* 24:693-697.

Podell S, Gaasterland T, Allen EE (2008) A database of phylogenetically atypical genes in archaeal and bacterial genomes, identified using the DarkHorse algorithm. *BMC Bioinformatics* 9:419.

Puigbo P, Wolf YI, Koonin EV (2009) Search for a 'Tree of Life' in the thicket of the phylogenetic forest. *J Biol* 8:59.

Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH (2005) Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nuc Acid Res* 33:5799-5808.

Rossi JV, Roberts MA, Yoo H-D, Gerwick WH (1997) Pilot scale culture of the marine cyanobacterium *Lyngbya majuscula* for its pharmaceutically-useful natural metabolite curacin A. *J Appl Phycol* 9:195-204.

Sorrels CM, Proteau PJ, Gerwick WH (2009) Organization, evolution, and expression analysis of the biosynthetic gene cluster for scytonemin, a cyanobacterial UV-absorbing pigment. *Appl Environ Microbiol* 75:4861-4869.

Stal LJ, Krumbein WE (1981) Aerobic nitrogen fixation in pure cultures of a benthic marine *Oscillatoria* (cyanobacteria). *FEMS Microbiol Lett* 11:295-298.

Sunagawa S, DeSantis TZ, Piceno YM, Brodie EL, DeSalvo MK, Voolstra CR, Weil E, Andersen GL, Medina M (2009) Bacterial diversity and White Plague Disease associated community changes in the Caribbean coral *Montastraea faveolata*. *ISME J* 3:512-521.

Swingley WD, Blankenship RE, Raymond J (2008) Integrating Markov Clustering and Molecular Phylogenetics to Reconstruct the Cyanobacterial Species Tree from Conserved Protein Families. *Mol Biol Evol* 25:643-654.

Tabb DL, McDonald WH, Yates III JR (2002) DTASelect and Contrast: Tools for Assembling and Comparing Protein Identifications from Shotgun Proteomics. *J Proteome Res* 1: 21-26.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molec Biol Evol* 24:1596-1599.

Tan LT (2007) Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochem* 68:954-979.

Thacker R, Ginsburg D, Paul VJ (2001) Effects of herbivore exclusion and nutrient enrichment on coral reef macroalgae and cyanobacteria. *Coral Reefs* 19:318-329.

Tidgewell K, Clark BR, Gerwick WH (2010) The Natural Products Chemistry of Cyanobacteria. In *Comprehensive Natural Products II Chemistry and Biology*; Mander L, Lui H-W, Eds.; Elsevier: Oxford: 2, 141-188.

Tidgewell K, Engene N, Byrum T, Media J, Doi T, Valeriote FA, Gerwick WH (2010) Evolved diversification of a modular natural product pathway: Apratoxins F and G,

two cytotoxic cyclic depsipeptides from a Palmyra collection of *Lyngbya bouillonii*. *Chembiochem* 11:1458-1466.

Udwary DW, Zeigler L, Asolkar RN, Singan V, Lapidus A, Fenical W, Jensen PR, Moore BS (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete *Salinispora tropica*. *Proc Nat Acad Sci USA* 104:10376-10381.

Washburn MP, Wolters D, Yates III JR (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat Biotech* 19:242-247.

Wegener KM, Singh AK, Jacobs JM, Elvitigala TR, Welsh EA, Keren N, Gritsenko MA, Ghosh BK, Camp II DG, Smith RD, Pakrasi HB (2010) Global proteomics reveal an atypical strategy for carbon/nitrogen assimilation by a cyanobacterium under diverse environmental perturbations. *Molec Cell Proteom* 9:2678-2689.

Yang X-J, Seto E (2008) Lysine acetylation: Codified crosstalk with other posttranslational modifications. *Mol Cell* 31:449-461.

Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, Hansen A, Karl DM (2001) Unicellular cyanobacteria fix N_2 in the subtropical North Pacific Ocean. *Nature* 412:635-638.

5.9 Supplemental information

Table 5.S1: Genome features of *Lyngbya majuscula* 3L draft assembly.

	Lyngbya majuscula 3L
Bases	8,498,160
GC%	0.44
CDS	7,479
% coding	76%
tRNA	56
rRNA operons	2
Genes	7,565
% with function prediction	54%
Transposases	252
CRISPR-associated proteins	0
Status	Draft
Scaffolds	161

Table 5.S2: COG categories in the *L. majuscula* 3L genome.

COG Categories	Gene Count	% of Annotated Genes
Replication, recombination, and repair	366	9
Cell wall/membrane/envelope biogenesis	353	8
Signal transduction mechanisms	311	7
Amino acid transport and metabolism	281	7
Coenzyme transport and metabolism	213	5
Energy production and conversion	203	5
Transcription	182	4
Translation, ribosomal structure and biogenesis	172	4
Posttranslational modification, protein turnover, chaperones	170	4
Carbohydrate transport and metabolism	166	4
Inorganic ion transport and metabolism	161	4
*Secondary metabolites biosynthesis, transport and catabolism	126	3
Defense mechanisms	107	3
Intracellular trafficking, secretion, and vesicular transport	84	2
Lipid transport and metabolism	83	2
Nucleotide transport and metabolism	69	2
Cell motility	62	1
Cell cycle control, cell division, chromosome partitioning	41	1
Chromatin structure and dynamics	2	0.05
Cytoskeleton	2	0.05
General function prediction only	536	13
Function unknown	495	12
Not in COGs	3858	51

Table 5.S3: 16S rRNA matches in combined Sanger-454 assembly.

scf id	match id	Description	Tax group	% ident.	align length	scf % G+C	scf pct 454 reads
47221	108542	Uncultured Bacteroidetes	Bacteroidetes	98	432	42	100%
52016	22139	*Lyngbya sp.*	Cyanobacteria	98	248	44	62%
52023	22139	*Lyngbya sp.*	Cyanobacteria	98	360	44	57%
52112	22139	*Lyngbya sp.*	Cyanobacteria	99	1482	44	69%
45926	336986	Uncultured bacterium	Gamma-proteobacteria	99	1182	47	100%
46341	108542	Uncultured Bacteroidetes	Bacteroidetes	97	747	49	100%
51929	154709	*Alteromonas macleodii*	Gamma-proteobacteria	97	906	49	98%
46429	315029	Uncultured bacterium	Alpha-proteobacteria	99	918	51	56%
51845	100014	Uncultured bacterium	Bacteroidetes	98	609	52	100%
50749	317859	*Muricauda aquimarina*	Gamma-proteobacteria	96	1151	52	100%
48764	361451	Uncultured *Hyphomonas*	Alpha-proteobacteria	99	1410	53	0%
48160	264115	*Phaeobacter caerulensis*	Alpha-proteobacteria	98	1420	56	11%

Table 5.S4a: Most readily identified proteins in soluble protein *L. majuscula* 3L extracts.

NCBI gene locus ID	Protein ID	Average spectral count (S.E.M.)
LYNGBM3L_16560	phycocyanin, beta subunit	1951 (133)
LYNGBM3L_56140	phycobilisome protein	1902.75 (93)
LYNGBM3L_15380	allophycocyanin, beta subunit	1506.5 (221)
LYNGBM3L_16550	phycocyanin, alpha subunit	1089.75 (108)
LYNGBM3L_56060	phycobilisome protein	1054.25 (138)
LYNGBM3L_15440	phycobilisome protein	1019.75 (78)
LYNGBM3L_56030	phycobilisome protein	776.75 (79)
LYNGBM3L_40910	putative multicopper oxidase	560.25 (34)
LYNGBM3L_58690	putative phospholipid-binding domain	336.25 (78)
LYNGBM3L_35040	fructose-biphosphate aldolase	330.75 (69)

Table 5.S4b: Expressed proteins in *L. majuscula* 3L predicted to function in natural product biosynthesis.

*Less than 2 unique peptides identified by tryptic digest

NCBI gene locus ID	Protein ID	Average spectral count (S.E.M.)
LYNGBM3L_06570	BarA*	0.5 (0.5)
LYNGBM3L_06560	BarB1	4 (0.8)
LYNGBM3L_06550	BarB2	8.75 (3.2)
LYNGBM3L_06530	BarC	9.5 (4.0)
LYNGBM3L_06520	BarD	5.75 (2.3)
LYNGBM3L_06510	BarE	51.25 (12.1)
LYNGBM3L_06500	BarF	8.75 (0.8)
LYNGBM3L_06490	BarG	29.75 (4.0)
LYNGBM3L_06480	BarH	61 (5.8)
LYNGBM3L_74570	CurA	29 (4.1)
LYNGBM3L_74560	CurB	9.75 (2.5)
LYNGBM3L_74550	CurC	7.5 (2.2)
LYNGBM3L_74540	CurD	4.25 (2.1)
LYNGBM3L_74530	CurE	3.5 (0.5)
LYNGBM3L_74510	CurF	4 (2.0)
LYNGBM3L_74500	CurG	8.5 (3.3)
LYNGBM3L_74480	CurH	10.5 (2.7)
LYNGBM3L_74470	CurI	7 (1.7)
LYNGBM3L_74460	CurJ	10.25 (2.5)
LYNGBM3L_74450	CurK	1.5 (0.9)
LYNGBM3L_74440	CurL	7.75 (2.1)
LYNGBM3L_74580	CurM*	1.75 (1.75)
LYNGBM3L_66100	Carmabin Scf 52117: NADPH quinone reductase and related Zn-dependent oxidoreductases	2.25 (0.9)
LYNGBM3L_66070	Carmabin Scf 52117: amino acid adenylation domain	10.5 (2.7)
LYNGBM3L_66060	Carmabin Scf 52117: amino acid adenylation domain	2 (1.4)
LYNGBM3L_66050	Carmabin Scf 52117: amino acid adenylation domain	14.5 (3.7)
LYNGBM3L_66040	Carmabin Scf 52117: amino acid adenylation domain	10.75 (1.7)
LYNGBM3L_03580	Carmabin Scf 51865: Phosphopantetheine attachment site	4.5 (2.6)
LYNGBM3L_67160	Scf 52117: amino acid adenylation domain	3.25 (1.2)

CHAPTER 6

Dissertation Conclusions and Future Directions of Natural Products Research in Filamentous Marine Cyanobacteria

Filamentous marine cyanobacteria continue to be rich and reliable sources of novel bioactive compounds that are valuable in drug discovery and biotechnology. The relative ease of collecting these macroscopic bacteria by hand in tropical, shallow water environments with snorkeling and SCUBA has allowed for an incredible diversity of molecules to be isolated and described over the past 40 years. As mentioned in this dissertation, several of these natural products, such as curacin A (Verdier-Pinard et al. 1998), somocystinamide A (Wrasidlo et al. 2008), and apratoxin A (Liu et al. 2009) have received considerable attention for their potency in specific assays and are of serious preclinical interest for disease treatment. Continued field collections of filamentous cyanobacteria, including in geographic areas that have received relatively less attention, should maintain a healthy supply of novel chemical entities for the foreseeable future. On the other hand, because cyanobacteria tend to proliferate in areas where coral reefs have been damaged or compromised due to anthropogenic impacts (Paul et al. 2005), increased cyanobacterial populations in some well populated areas may be an unintended but indirectly positive side effect of reef damage and decline and offer increased amounts of biomass for natural product investigations. Whether an adequate level of cyanobacterial diversity for continued natural product discovery would be maintained under these conditions may depend on additional environmental and ecological factors.

The natural product potential of marine cyanobacteria has been immeasurably increased over the past 10 years as a result of the enormous improvements in genetics and genomic technologies. The biosynthetic gene clusters identified to date for natural products produced by *Lyngbya* strains encode genes that perform enzymatic reactions rarely or never before seen in microbial life (Jones et al. 2010). Use of these enzymes for directed engineering of natural products should continue to add to the available pool of molecules being considered as new drugs, and could be used in concert with pathways from other bacteria to generate novel analogs of molecules already of high value. The rapidly decreasing cost of sequencing will quickly reveal a significant number of gene clusters responsible for producing *Lyngbya* molecules that have already been characterized, as well as an assortment of unknown, orphan pathways that will require further investigation. General improvements in culturing techniques for these cyanobacterial strains may lead to conditions that promote the expression of these pathways and unlock even more molecular diversity.

Despite the proven track record and anticipated future success of *Lyngbya* and other filamentous marine cyanobacteria to provide natural products useful in human applications, there are a number of limitations inherent in these organisms that must be addressed to fully capture and unlock this potential. Several of these limitations, including low compound yields, slow growth rates, lack of genetic techniques, and unavailability of genomic data have been the subject of research presented in this dissertation. Chapters 2 and 3, focusing on the transcription and regulation of the neurotoxic jamaicamides, represent the first such efforts to understand how expression of gene clusters from marine *Lyngbya* strains are controlled by genetic and

environmental factors (Jones et al. 2009, Esquenazi et al. 2011). Each factor is of significant importance in determining how to improve natural product yields in laboratory cultures. Additional study of the promoter regions identified in the jamaicamide pathway, including how regulatory proteins such as 7968 and 5335 interact with these regions, could yield specific targets for manipulation once genetic techniques for these strains are available. The ability to monitor turnover of *Lyngbya* natural products *in vivo* using MALDI-TOF mass spectrometry and stable isotope feeding as presented in chapter 3 will likely become the standard method used by our laboratory and others to evaluate the impact different culture conditions make on the rate of molecule biosynthesis. The requirement for minute amounts of biomass and the simultaneous access to all ionizing compounds present in a MALDI extract should allow for a wide variety of these conditions such as temperature, nutrients, co-culture, and elicitor panels to be tested rapidly and at relatively low cost. Inclusion of other isotopes into this process may increase the applications of this approach to polyketides or other non-alkaloid cyanobacterial natural products.

Probably the principal impediment to accessing large amounts of cyanobacterial natural products without organic synthesis is the total absence of any genetic techniques for these strains. As discussed in chapter 4, all attempts to introduce DNA into *Lyngbya* strains proved unsuccessful. Most likely, establishment of a heterologous host for these natural products will precede development of these techniques. The thick polysaccharide sheath, slow growth rates, and close association with heterotrophic organisms are highly significant obstacles to generating *Lyngbya* mutants. As success with producing lyngbyatoxin proteins in *Escherichia coli* and

Streptomyces coelicolor indicates, other bacterial strains are the best options at present for overexpression of *Lyngbya* gene clusters and platforms for targeted gene replacement, mutasynthesis and other biochemical investigations. The genetic differences between actinobacteria and cyanobacteria may mean that actinobacteria are not the optimal long term choice for these *Lyngbya* pathways. With proper engineering, use of other cyanobacteria such as *Synechococcus* or *Synechocystis* (Roberts et al. 2009) may avoid GC content or codon bias problems as were anticipated in using *Streptomyces*. In the future, development of a highly modified (potentially synthetic) bacterial "superhost" may make many of these limitations trivial; until that time, it would be prudent to expand on the work presented here to continue to develop heterologous expression strategies for *Lyngbya* natural product gene clusters.

The genome sequence of *Lyngbya majuscula* 3L represents a monumental effort on the part of many scientists over approximately 5 years to assemble the information presented in chapter 5. This particular strain was selected because of its natural product components and because it is probably the best characterized from the Gerwick culture archives. This strain was also thought to produce both antillatoxin (Orjala et al. 1995b) and malyngamide H (Orjala et al. 1995a) – molecules for which gene cluster information would have been particularly valuable because of the attention each molecular class has received on account of bioactivity (Jabba et al. 2010, Villa et al. 2010). The lack of these pathways (in addition to an unexpected absence of any other large natural product gene clusters beyond those already characterized) is one of the best indications to date that the many hundreds of natural

products attributed to marine *Lyngbya* species are scattered among many strains worldwide and could provide the most accurate method of taxonomy in light of such a high degree of phylogenetic ambiguity (Sharp et al. 2009, Engene et al. 2010). Indeed, revision of the genus of this marine *Lyngbya* lineage is already underway (Engene et al., unpublished data) and these efforts will provide more guidance in what natural products different strains may produce and direct future genome sequencing efforts. However, even with these improvements it will be impossible to gauge the full natural product arsenal of these strains without genomic information (Jones et al. 2011), as each cluster may be expressed only under specific genetic and environmental conditions either in the field or in culture. In this regard, the future is incredibly bright for marine cyanobacteria genome sequencing efforts. Each strain represents a metagenomic data set that can be mined and exploited for its assortment of natural product pathways – especially upon development of improved heterologous expression techniques. One of the best insights from these efforts will be to better understand the bacteria associated with cyanobacterial filaments. What role do these bacteria serve in the ecology and physiology of cyanobacteria? What is the natural product capacity of these strains? And, what percentage of natural products attributed to cyanobacteria are in fact produced by these so-called "nuisance organisms?"

Filamentous cyanobacteria from the marine realm are ancient organisms that continue to offer contemporary inspiration for using natural sources to solve problems facing humans today. Improvements in technology over time will eventually allow for the full potential of marine cyanobacteria to be accessed and limitations inherent in their morphology and life history characteristics to be overcome. Work presented in

this dissertation addresses some of the most significant current problems natural products chemists and biologists face in working with marine cyanobacteria, and results presented here provide direction on how we can make progress in accessing greater amounts of cyanobacterial natural products, manipulate cyanobacterial genetic information, and unlock the full natural product potential of these remarkable life forms.

References

Engene N, Coates RC, Gerwick WH (2010) 16S rRNA gene heterogeneity in the filamentous marine cyanobacterial genus *Lyngbya*. *J Phycol* 46:591-601.

Esquenazi E, Jones AC, Byrum T, Dorrestein PC, Gerwick WH (2011) Temporal dynamics of natural products biosynthesis in marine cyanobacteria. *Proc Natl Acad Sci USA* 108:5226-5231.

Jabba SV, Prakash A, Dravid SM, Gerwick WH, Murray TF (2010) Antillatoxin, a novel lipopeptide, enhances neurite outgrowth in immature cerebrocortical neurons through activation of voltage-gated sodium channels. *J Pharmacol Exper Therap* 332:698-709.

Jones AC, Monroe EA, Eisman EB, Gerwick L, Sherman DH, Gerwick WH (2010) The unique mechanistic transformations involved in the biosynthesis of modular natural products from marine cyanobacteria. *Nat Prod Rep* 27:1048-1065.

Jones AC, Monroe EA, Podell S, Hess WR, Klages S, Esquenazi E, Niessen S, Hoover H, Rothmann M, Kube M, Reinhardt R, Lasken RS, Yates III JR, Burkart MD, Allen EE, Dorrestein PC, Gerwick WH, Gerwick L (2011). Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proc Natl Acad Sci USA* doi:10.1073/pnas.1101137108.

Jones AC, Gerwick L, Gonzalez D, Dorrestein PC, Gerwick WH (2009) Transcriptional analysis of the jamaicamide gene cluster from the marine cyanobacterium *Lyngbya majuscula* and identification of possible regulatory proteins. *BMC Microbiol* 9:247

Liu Y, Law BK, Luesch H (2009) Apratoxin A reversibly inhibits the secretory pathway by preventing cotranslational translocation. *Mol Pharmacol* 76:91-104.

Orjala J, Nagle D, Gerwick WH (1995a) Malyngamide H, an ichthyotoxic amide possessing a new carbon skeleton from the Caribbean cyanobacterium *Lyngbya majuscula*. *J Nat Prod* 58:764-768.

Orjala J, Nagle DG, Hsu V, Gerwick WH (1995b) Antillatoxin: An exceptionally ichthyotoxic cyclic lipopeptide from the tropical cyanobacterium *Lyngbya majuscula*. *J Am Chem Soc* 117:8281-8282.

Paul VJ, Thacker RW, Banks K, Golubic S (2005) Benthic cyanobacterial bloom impacts the reefs of South Florida (Broward County, USA). *Coral Reefs* 24:693-697.

Roberts AA, Copp JN, Marahiel MA, Neilan BA (2009) The *Synechocystis* sp. PCC6803 sfp-type phosphopantetheinyl transferase does not possess characteristic broad-range bioactivity. *ChemBioChem* 10:1869-1877.

Sharp K, Arthur KE, Gu L, Ross C, Harrison G, Gunasekera SP, Meickle T, Matthew S, Luesch H, Thacker RW, Sherman DH, Paul VJ (2009) Phylogenetic and chemical diversity of three chemotypes of bloom-forming *Lyngbya* species (Cyanobacteria: Oscillatoriales) from reefs of Southeastern Florida. *Appl Environ Microbiol* 75:2879-2888.

Verdier-Pinard P, Lai J-Y, Yoo H-D, Yu J, Marquez B, Nagle DG, Nambu M, White JD, Falck JR, Gerwick WH, Day BW, Hamel E (1998) Structure-activity analysis of the interaction of curacin A, the potent colchicine site antimitotic agent, with tubulin and effects of analogs on the growth of MCF-7 breast cancer cells. *Mol Pharmacol* 53:62-76.

Villa FA, Lieske K, Gerwick L (2010) Selective MyD88-dependent pathway inhibition by the cyanobacterial natural product malyngamide F acetate. *Eur J Pharmacol* 629:140-146.

Wrasidlo W, Mielgo A, Torres VA, Barbero S, Stoletov K, Suyama TL, Klemke RL, Gerwick WH, Carson DA, Stupack DG (2008) The marine lipopeptide somocystinamide A triggers apoptosis via caspase 8. *Proc Natl Acad Sci USA* 105:2313-2318.