


11-2009

# Absolute Configuration and Biosynthesis of Pahayokolide A from *Lyngbya* sp. Strain 15-2 of the Florida Everglades

Li Liu

*Florida International University*, [lliu004@fiu.edu](mailto:lliu004@fiu.edu)

Follow this and additional works at: <http://digitalcommons.fiu.edu/etd>

 Part of the [Analytical Chemistry Commons](#), [Biochemistry Commons](#), and the [Medicinal-Pharmaceutical Chemistry Commons](#)

---

## Recommended Citation

Liu, Li, "Absolute Configuration and Biosynthesis of Pahayokolide A from *Lyngbya* sp. Strain 15-2 of the Florida Everglades" (2009). *FIU Electronic Theses and Dissertations*. Paper 134.  
<http://digitalcommons.fiu.edu/etd/134>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact [dcc@fiu.edu](mailto:dcc@fiu.edu).

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

ABSOLUTE CONFIGURATION AND BIOSYNTHESIS OF PAHAYOKOLIDE A  
FROM *LYNGBYA* SP. STRAIN 15-2 OF THE FLORIDA EVERGLADES

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CHEMISTRY

by

Li Liu

2009

To: Dean Kenneth G. Furton  
College of Arts and Sciences

This dissertation, written by Li Liu, and entitled Absolute Configuration and Biosynthesis of Pahayokolide A from *Lyngbya* sp. Strain 15-2 of the Florida Everglades, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

---

Watson Lees

---

Fenfei Leng

---

José Almirall

---

David W. Lee

---

Kathleen S. Rein, Major Professor

Date of Defense: November 10, 2009

The dissertation of Li Liu is approved.

---

Dean Kenneth G. Furton  
College of Arts and Sciences

---

Dean George Walker  
University Graduate School

Florida International University, 2009

© Copyright 2009 by Li Liu

All rights reserved.

## DEDICATION

I dedicate this dissertation to my parents, who have helped so much with baby-sitting and have given me their fullest support; particularly to my husband, who has always supported me with his unconditional love, and to our precious daughter Yuxi, who is the joy of our lives.

## ACKNOWLEDGMENTS

I am grateful for all the support I have received during researching and writing up this dissertation. My sincere gratitude goes to Dr. Kathleen Rein, my major professor, for her support and creative advice throughout my Ph.D study at FIU. This dissertation would not have been possible without her expert guidance of every step of the way. Not only was she readily available for me, but she always read and responded to the drafts of each chapter of my work more quickly than I could have hoped. I deeply appreciate my committee members, Dr. Watson Lees, Dr. Fenfei Leng, Dr. José Almirall and Dr. David Lee for their helpful comments during my stay at FIU.

I wish to thank the people in my laboratory, Fujiang, Wentian, Richard, Robert, Jamie, Alphard, Daisy, Vianessa, Tianying, Ryan and Dragon for their friendship and support. Many people on the faculty and staff of the Department of Chemistry & Biochemistry and FIU Graduate School assisted and encouraged me in various ways. I am especially grateful to Dr. Yong Cai and Dr. Xiaotang Wang for their friendly support. My research for this dissertation was made more efficient through the use of several electronic resources. Thus I gladly express my gratitude to the staff in the departmental spectroscopy room: Yali, Myron and Dr. Yaru Song.

Thanks to the FIU Graduate School for financial support with the FIU Presidential Fellowship and Dissertation Year Fellowship. The financial support from National Institute of Environmental Health Sciences is gratefully acknowledged.

Finally, I would like to thank my family, especially my husband, for their patience, understanding, support, and most of all love.

ABSTRACT OF THE DISSERTATION  
ABSOLUTE CONFIGURATION AND BIOSYNTHESIS OF PAHAYOKOLIDE A  
FROM *LYNGBYA* SP. STRAIN 15-2 OF THE FLORIDA EVERGLADES

by

Li Liu

Florida International University, 2009

Miami, Florida

Professor Kathleen S. Rein, Major Professor

Pahayokolides A-D are cytotoxic cyclic polypeptides produced by the freshwater cyanobacterium *Lyngbya* sp. strain 15-2 that possess an unusual  $\beta$ -amino acid, 3-amino-2,5,7,8-tetrahydroxy-10-methylundecanoic acid (Athmu). The absolute configuration of pahayokolides A-D was determined using advanced Marfey's method. It was also confirmed that a pendant *N*-acetyl-*N*-methyl leucine moiety in pahayokolide A was absent in pahayokolides B and pahayokolides C-D were conformers of pahayokolide A. Feeding experiments indicated that the biosynthesis of the Athmu sidechain arises from leucine or  $\alpha$ -ketoisovalerate, however could not be further extended by three rounds of condensation with malonate units. Putative four peptide and one unique polyketide synthetases in *Lyngbya* sp. strain 15-2 were identified by using a PCR method and degenerate primers derived from conserved core sequences of known NRPSs and PKSs. Identification of one unique KS domain conflicted with the logic rule that the long side chain of Athmu was assembled by three rounds of ketide extensions if PKSs were involved. A gene cluster (*pah*) encoding a peptide synthetase putatively producing pahayokolide was cloned, partially sequenced and characterized. Seven modules of the

non-ribosomal peptide synthetase (NRPS) were identified. Ten additional opening reading frames (ORFs) were found, responsible for peptide resistance, transport and degradation. Although the predicted substrate specificities of NRPS agreed with the structure of pahayokolide A partially, the disagreement could be explained. However, no PKS gene was found in the *pah* gene cluster.



## TABLE OF CONTENTS

CHAPTER	PAGE
Scope of the dissertation .....	1
1. Polypeptides Isolated from <i>Lyngbya</i> Species .....	2
1.1 Introduction.....	2
1.1.1 Acyclic polypeptides.....	5
1.1.2 Cyclic polypeptides.....	14
1.1.3 Cyclic polypeptides with branches .....	35
1.1.4 Absolute configuration of peptides.....	43
1.2 Results and Discussion .....	46
1.2.1 Absolute configuration of pahayokolides A and B.....	46
1.2.2 Pahayokolides C and D is conformers of pahayokolide A .....	55
1.3 Materials and Methods.....	59
1.3.1 Culture conditions of <i>Lyngbya</i> sp. strain 15-2.....	59
1.3.2 Purification of pahayokolides A-D .....	59
1.3.3 Preparation of FDLA derivatives.....	60
1.3.4 HPLC/PDA conditions .....	61
1.3.5 ESI LC/MS Conditions .....	61
1.3.6 Spike-in experiments .....	61
1.4 Conclusions.....	61
2. Biosynthetic Pathway for the Construction of the Pahayokolide A .....	63
2.1 Introduction.....	63
2.1.1 Functional domains of non-ribosomal peptide synthetases .....	67
2.1.1.1 Activation by the adenylation domain .....	67
2.1.1.2 Intermediate transport by the peptidyl carrier protein .....	72
2.1.1.3 Peptide elongation by the condensation domain.....	73
2.1.2 Peptide release .....	76
2.1.3 Modifying domains.....	80
2.1.3.1 Epimerization domain.....	80
2.1.3.2 Cyclization domain.....	81
2.1.3.3 <i>N</i> -methyltransferase domain.....	82
2.1.4 Deviations from NRPS co-linearity rule.....	82
2.1.5 Polyketide synthetases associated with nonribosomal peptide synthesis .....	86
2.1.6 Biosynthesis of non-ribosomal peptides in cyanobacteria.....	87
2.2 Results and Discussion .....	89
2.2.1 The biosynthetic origin of the Athmu moiety in pahayokolide A .....	89
2.2.2 Identification putative peptide and polyketide synthetases in <i>Lyngbya</i> sp. strain 15-2 .....	98

2.2.3 Cloning and partial characterization of the NRPS genes for pahayokolide A production in <i>Lyngbya</i> sp. strain 15-2.....	101
2.3 Materials and Methods.....	108
2.3.1 Culture conditions of <i>Lyngbya</i> sp. strain 15-2.....	108
2.3.2 Stable isotope incorporation experiments.....	109
2.3.3 NMR experiment.....	109
2.3.4 DNA manipulations.....	109
2.3.5 Construction of the fosmid library.....	110
2.3.6 Screening of the fosmid library.....	111
2.3.7 Sequence analysis.....	112
2.4 Conclusions.....	112
Significance and Future Work.....	114
References.....	116
VITA.....	125

## LIST OF TABLES

TABLE	PAGE
Table 1 - 1. Availability of amino acids standards and reactions with FDLA.....	46
Table 1 - 2. LC/MS analysis of pahayokolide A hydrolysates derivatized with D,L-FDLA and L-FDLA .....	54
Table 2 - 1. Representative Examples of Nonproteinogenic Constituents of Peptide Antibiotics.....	66
Table 2 - 2. Highly Conserved Core Motifs of the A Domain of Peptide Synthetases.....	69
Table 2 - 3. Highly Conserved Core Motifs of the C Domain of Peptide Synthetases.....	74
Table 2 - 4. Highly Conserved Core Motifs of the Epimerization Domain of Peptide Synthetases.....	81
Table 2 - 5. Highly Conserved Core Motifs of the N-methylation Domain of Peptide Synthetases.....	82
Table 2 - 6. Isotopic enrichments based on <sup>13</sup> C NMR for pahayokolide A (98) .....	92
Table 2 - 7. Activated substrates and the binding pocket amino acid residues of A domains in pLN2, pLN6 and pLN8 .....	100
Table 2 - 8. Partial sequencing of Fos11, 45, 22, 83.....	101
Table 2 - 9. Deduced functions of encoded proteins in the pah gene cluster of Lyngbya sp. strain 15-2.....	103
Table 2 - 10. Deduced functions of the opening frame reading frames in the pah gene cluster .....	107
Table 2 - 11. primer sequences .....	110

## LIST OF FIGURES

FIGURE	PAGE
Figure 1 - 1. The reaction of D, L-FDLA with amino acids .....	45
Figure 1 - 2. HPLC/PDA chromatogram of standard amino acid mixture derivatized with D,L-FDLA.....	46
Figure 1 - 3. HPLC/PDA analysis of pahayokolide A hydrolysates derivatized with D,L-FDLA (top) and L-FDLA (bottom).....	48
Figure 1 - 4. HPLC/PDA chromatogram of pahayokolide A (98) hydrolysate derivatized with L-FDLA and spiked with N-Me-L-Leucine derivatized with L-FDLA (top) and D-FDLA (bottom).....	49
Figure 1 - 5. HPLC/PDA chromatogram of pahayokolide A (98) hydrolysate derivatized with L-FDLA and spiked with D,L-Homophe derivatized with L-FDLA. ....	50
Figure 1 - 6. HPLC/PDA chromatogram of pahayokolide A (98) hydrolysate derivatized with L-FDLA and spiked with D-Phe derivatized with L-FDLA..	51
Figure 1 - 7. Mass chromatograms of the L-FDLA derivatives of pahayokolide A (98) hydrolysate monitored at the m/z values of the M-1 ions (A, m/z 572, L- Phe; B, m/z 552, N-Me-D,L-Leu; C, m/z 586, D-Homophe) ions using ESI LC/MS in negative ion mode.....	52
Figure 1 - 8. Mass chromatograms of the L-FDLA (A-D) and D,L-FDLA (E) derivatives of pahayokolide A (98) hydrolysate monitored at the m/z values of the M-1 ions (A, m/z 512, L-Ser; B, m/z 526, L-Thr; C, m/z 554, D-Glu; D, m/z 686, D-AthmU and E, m/z 686, D-AthmU) using ESI LC/MS in negative ion mode.....	53
Figure 1 - 9. The absolute configuration of pahayokolides A (98) and B (99).....	53
Figure 1 - 10. HPLC/PDA chromatogram of pahayokolides A-D.....	55
Figure 1 - 11. HPLC/PDA chromatogram of pahayokolide A heated at 39 °C and monitored at 220 nm (red) and 254 nm (blue) at 0, 1, 2, 3, and 48 hours. ....	57
Figure 1 - 12. HPLC/PDA chromatogram of a complex with pahayokolide C as the major component heated at 39 °C and monitored at 220 nm (red) and 254 nm (blue) at 0, 1, 2, 3, and 48 hours. ....	58

Figure 1 - 13. HPLC/PDA chromatogram of a complex with pahayokolide D as the major component heated at 39 °C and monitored at 220 nm (red) and 254 nm (blue) at 0, 1, 2, 3, and 48 hours. ....	59
Figure 2 - 1 Linear, cyclic and branched cyanobacterial peptides and their bioactivities.....	64
Figure 2 - 2. Domain-catalyzed reactions. ....	68
Figure 2 - 3 Ribbon diagram of the phenylalanine-activating A domain (PheA) of gramicidin S synthetases I (GrsA).....	70
Figure 2 - 4 Posttranslational phosphopantetheinylation.....	72
Figure 2 - 5 Structure of the PCP domain derived from the third module of the <i>B. brevis</i> tyrocidine synthetase. ....	73
Figure 2 - 6 Crystal structure of VibH.....	75
Figure 2 - 7 Suggested mechanism for the condensation/elongation reaction in peptide synthesis. ....	75
Figure 2 - 8. Structure of Srf TE.....	76
Figure 2 - 9 Mechanism of thioesterase domain catalysis. ....	77
Figure 2 - 10. Naturally occurring macrolactones and macrolactams. ....	79
Figure 2 - 11. Enterobactin NRPS as an example for iterative NRPS.....	84
Figure 2 - 12 Vibriobactin NRPS as an example for nonlinear NRPS.....	85
Figure 2 - 13. Proposed biosynthesis of Adda.....	90
Figure 2 - 14. The structure of pahayokolide A-B and <sup>13</sup> C labeled substrates. ....	91
Figure 2 - 15. Natural abundance (top) and [1- <sup>13</sup> C]-L-leucine labeled (bottom) <sup>13</sup> C NMR of pahayokolides A (98) (66~76 ppm).....	94
Figure 2 - 16. Natural abundance (top) and [1- <sup>13</sup> C]-L-leucine labeled (bottom) <sup>13</sup> C NMR of pahayokolides A (98) (170~175 ppm).....	94
Figure 2 - 17. Natural abundance (top) and [1,2- <sup>13</sup> C]-L-leucine labeled (bottom) <sup>13</sup> C NMR of pahayokolide A (98) (66~76 ppm).....	95

Figure 2 - 18 Natural abundance (top) and [1,2- <sup>13</sup> C]-L-leucine labeled (bottom) <sup>13</sup> C NMR of pahayokolides A (98) (170~175 ppm) .....	95
Figure 2 - 19. Natural abundance (top) and [1- <sup>13</sup> C]-acetate labeled (bottom) <sup>13</sup> C NMR of pahayokolides A (98) (165~180 ppm) .....	96
Figure 2 - 20. Natural abundance (top) and [1- <sup>13</sup> C]-acetate labeled (bottom) <sup>13</sup> C NMR of pahayokolides A (98) (55~75 ppm) .....	96
Figure 2 - 21. Biosynthesis of L-leucine from valine .....	97
Figure 2 - 22. Proposed biosynthesis of Athmu.....	97
Figure 2 - 23. Structure of three amino sugars.....	98
Figure 2 - 24. PCR of genomic DNA from Lyngbya sp. Strain 15-2 using degenerate NRPS and PKS primers.....	99
Figure 2 - 25. NotI and Bsu36I restriction map of four NRPS-positive fosmids of Lyngbya sp. strain 15-2. ....	102
Figure 2 - 26 A predicted model for the biosynthesis of pahayokolide A templated by the Pah synthtase.....	105
Figure 2 - 27. Screening of the genomic library of Lyngbya sp. strain 15-2.....	111
Figure 2 - 28. Preparation of 20 pools of cells for the genomic library screening.	112

## Scope of the dissertation

*Lyngbya*, a common and accessible genus of cyanobacteria, produce numerous and structurally diverse secondary metabolites, including nonribosomal peptides and polyketides in particular. Many novel structures have been elucidated as a result of systematic screenings for anticancer, antibacterial, antifungal, and protease inhibitory effects (Moore, et al, 1996; Namikoshi, et al, 1996). Some *Lyngbya* compounds prove potential for development into commercial drugs. However, few biosynthetic gene clusters of *Lyngbya* peptides have been characterized to date and their heterologous expression needs to be established.

Pahayokolides A-D are branched-cyclic peptides isolated from freshwater *Lyngbya* sp. strain 15-2 of the Florida Everglades. Revealing the absolute configuration of pahayokolide A is the basis for understanding its bioactivities. The study of pahayokolide A and their biosynthetic origins has great value not only for drug discovery but also for harnessing these pathways to create new molecules of pharmaceutical interesting (Gerwick, et al., 2001).

The specific aims of this study were to determine the absolute configuration of pahayokolides A-D (Chapter 1), identify the biosynthetic origin of the Athmu moiety in pahayokolide A (Chapter 2), and clone and characterize the genes coding for the biosynthesis of pahayokolide A in *Lyngbya* sp. strain 15-2. (Chapter 2)

# 1. Polypeptides Isolated from *Lyngbya* Species

## 1.1 Introduction

Cyanobacteria, also called blue-green algae, are ancient aquatic and photosynthetic prokaryotes. The oldest known fossils cyanobacteria are from Archaean rocks of Western Australia, dated at 3.5 billion years (Stanley, 2004). Cyanobacteria share traits with both plants and bacteria. Like algae, cyanobacteria form oxygen during photosynthesis. Like other bacteria, cyanobacteria lack a nucleus and other internal organelles bound by membranes. However, there is an important difference between cyanobacteria and algae. Some cyanobacteria can fix nitrogen from the atmosphere. Therefore, they can become dominant over eukaryotic algae in nitrogen limited waters, because phosphorus and nitrogen are required for algal and cyanobacterial growth (Patočka, 2001).

A high degree of biological adaptation affords cyanobacteria the ability to produce a unique range of defensive secondary metabolites [Gerwick et al., 1994 and 2001; Namikoshi et al., 1996; Mayer et al., 2003]. Over 300 nitrogen-containing secondary metabolites from marine cyanobacteria have been reported in the literature (Tan, 2007). These secondary metabolites often enable cyanobacteria to compete effectively in a variety of environments and successfully inhabit extreme niches. Cyanobacterial secondary metabolites exhibit a variety of molecular structures and biological activities, especially those associated with inhibiting growth and development in other organisms. Cyanobacteria as a potential source for drug have generated scientific interest. A number of cyanobacterial secondary metabolites have been discovered as potential lead



compounds for further drug development. One compound already in clinical trials for the treatment of cancer is based on the cryptophycins isolated from terrestrial cyanobacteria (Trimurtulu et al.; 1994).

Cyanobacterial secondary metabolites are often isolated by bioassay guided fractionation (BGF) using a variety of biological screening methods. The main purpose of the BGF strategy is to divide the fractions of a cyanobacterial extract into two classes: the amenable and discarded fractions. The amenable fractions show a reasonable expectation in the bioactivity test and will be separated, typically chromatographically, into several subfractions. The discarded fractions do not contain a discrete, identifiable component responsible for the observed activity and follow-up efforts are usually discontinued. Subfractions giving a positive response in the confirming bioassay will be followed up for purification of active components (Phillipson, et al., 2002).

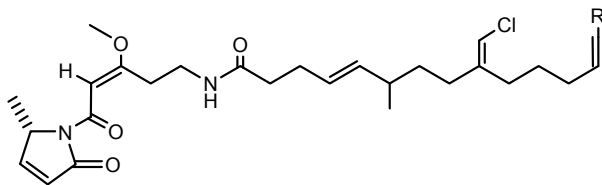
Some cyanobacterial secondary metabolites were also isolated from marine invertebrates. The presence of a natural product within a particular cell type cannot provide unequivocal proof for the biosynthetic origin of a given compound. Conceptually, demonstrating the biosynthesis of a natural product in a given cell type can be achieved on chemical, protein, or genetic levels.

The wide occurrence of toxin producing cyanobacteria in freshwater and seawater creates problems for all life forms. Cyanobacterial toxins are often responsible for acute lethal, chronic and sub-chronic poisonings of wild and/or domestic animals and humans (Carmichael, 1994, 1997; Jochimsen et al., 1998; Falconer, 1999; Daranas et al., 2001). Cyanobacterial toxins may be classified by how they affect the human body (Patočka, 2001). Hepatotoxins, are produced by some strains of the cyanobacteria *Microcystis*,

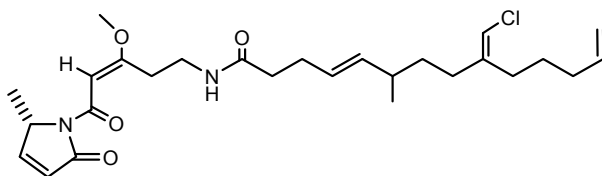
*Anabaena*, *Oscillatoria*, *Nodularia*, *Nostoc*, *Cylindrospermopsis* and *Umezakia*. Neurotoxins, produced by some strains of *Aphanizomenon* and *Oscillatoria*, affect the nervous system. *Cylindrospermopsis raciborski* may also produce toxic alkaloids, causing gastrointestinal symptoms or kidney disease in humans. Not all cyanobacteria form toxins, and it is likely that there are as yet many unrecognized toxins.

The taxonomy of cyanobacteria is complex and controversial (Patočka, 2001). Cyanobacteria of toxicological significance belong to the genus *Anabaena*, *Oscillatoria*, *Microcystis*, *Nodularia*, *Cylindrospermopsis* and *Lyngbya*. *Lyngbya* sp. and *Microcystis* sp. are easily collected and cultured in laboratory so that the isolation of compounds in mg range, while laborious and time consuming (Welker et al., 2006), is achievable. *Lyngbya* is a common and accessible genus of cyanobacteria, which is distributed worldwide throughout tropical and subtropical regions. The unbranched filaments of *Lyngbya* are cylindrical and usually wider than 6  $\mu\text{m}$ . The straight, slightly wavy, or rarely coiled *Lyngbya* filaments usually form large, layered, leathery mats of varied thickness, and then form large benthic and surface blooms in freshwater and sea water. The most important species of genus *Lyngbya* are *L. majuscula*, *L. martensiana*, *L. aestuarri* and *L. wollei*. There are an increasing number of *Lyngbya* species which have been found to produce an impressive array of structurally varied compounds with diverse biological activities [Gerwick et al, 2001]. A major group of cyanobacterial secondary metabolites are peptides or contain peptide substructures. A total of 108 polypeptides isolated from *Lyngbya* species, published between January 2001 and October 2009, are reviewed with emphasis on their structural diversity and biological activities. The gross structures of these molecules were determined using 1D and 2D NMR spectroscopy. The

relative and/or absolute configurations were elucidated by chemical manipulation including the advanced Marfey's method, successful application of the *J*-based configuration analysis, Mosher analysis, and organic synthesis of the stereoisomers (Tan, 2007).



Jamaicamide A (**1**) R=Br, Jamaicamide B (**2**) R=H



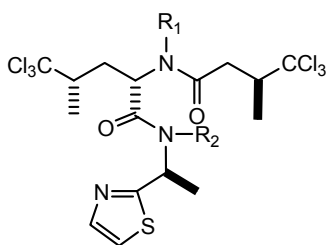
Jamaicamide C (**3**)

### 1.1.1 Acyclic polypeptides

A screening program for bioactive compounds from marine cyanobacteria led to the isolation of jamaicamides A-C from a dark green strain of *Lyngbya majuscula* found in Hector's Bay, Jamaica (Edwards et al., 2004). Jamaicamides A-C (**1-3**) are highly functionalized lipopeptides containing a terminal alkynyl or alkene, vinyl chloride,  $\beta$ -methoxy enone system, and pyrrolinone ring. Jamaicamides A-C (**1-3**) exhibited cytotoxicity ( $LC_{50} = 15 \mu M$ ) to both the H-460 human lung and Neuro-2a mouse neurocamide cell lines, sodium channel blocking activity and goldfish toxicity. Jamaicamide B (**2**) was the most active fish toxin and followed by jamaicamide C (**3**).

Interestingly, the nonaromatic ring-containing alkynoic lipopeptide jamaicamide B (**2**) showed weak antimalarial activity ( $IC_{50} = 18.4 \mu M$ ). However, the terminal bromoacetylene homologue jamaicamide A was inactive in this assay ( $IC_{50} > 50 \mu M$ ).

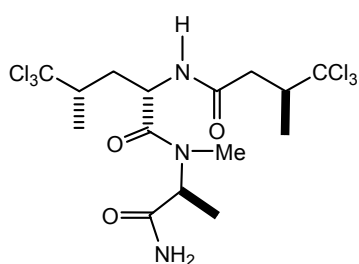
Three polychlorinated metabolites, pseudodysidenin (**5**), nordysidenin (**6**), and dysidenamide (**7**) were identified from a marine free-living red cyanobacterium, *Lyngbya majuscula*, collected at Boca del Drago Beach, Bocas del Toro, Panama (Jiménez et al., 2001). Compounds **5-7** are closely related to dysidenin (**4**) and isodysidenin (**8**) isolated from Australian and New Guinean specimens of *Dysidea herbacea*, respectively. This is the first report strengthening that sponge-based dysidenins are biosynthesized by associated cyanobacteria. The relative and absolute configurations of compounds **5-7** were established by chiral HPLC and optical rotation comparison to other chlorine-containing metabolites. Pseudodysidenin (**5**) exhibits cytotoxicity against four neoplastic cell lines P-388, A-549, HT-29, and MEL-28 ( $IC_{50} > 1 \mu g/mL$ ).



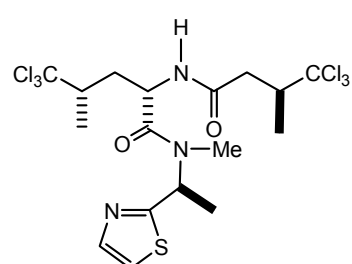
Dysidenin (**4**)  $R_1=CH_3$ ,  $R_2=H$

Pseudodysidenin (**5**)  $R_1=H$ ,  $R_2=CH_3$

Nordysidenin (**6**)  $R_1=H$ ,  $R_2=H$



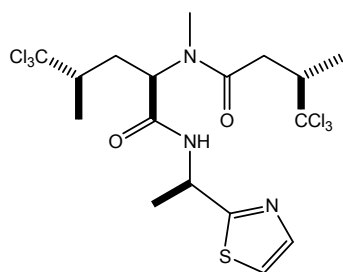
Dysidenamide (**7**)



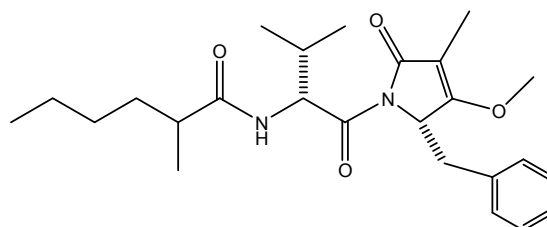
Isodysidenin (**8**)

The apratoxin-producing *Lyngbya* sp. NIH309 from Palau also afforded several other nitrogenous metabolites, including palau'imide (**10**) (Luesch et al., 2002b).

Palau'imide was unique and structurally unrelated to any other compounds isolated from the organism from Guam. The *N*-acylpyrrolinone unit in palau'imide (**10**) has precedence in dysidin (**9**) from the sponge *Dysidea herbacea*, albeit most likely produced by a symbiotic cyanobacterium (Hofheinz, et al., 1977). Palau'imide exhibited IC<sub>50</sub> values of 1.4 and 0.36 mM for in vitro cytotoxicity against KB and LoVo cells, respectively.



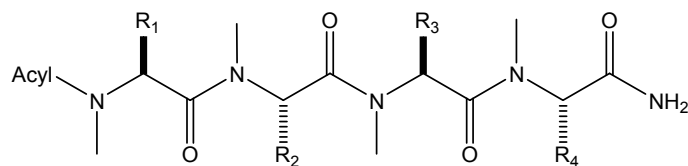
Dysidin (**9**)



Palau'imide (**10**)

Six structurally-related linear lipopeptides, including dragonamide A (**11**) (Jiménez et al., 2001), carmabin A (**13**), dragomabin (**14**) and dragonamide B (**12**) (Mcphail et. al., 2007), dragonamide C (**15**) and D (**16**) (Gunasekera et al., 2008a) were isolated from the marine cyanobacteria *Lyngbya majuscula* and *Lyngbya polychroa*. A unique C8-alkynoate unit has been found in carmabin A, dragomabin, dragonamides A and B. Dragonamide A (**11**) exhibits cytotoxicity against four neoplastic cell lines P-388, A-549, HT-29, and MEL-28 (IC<sub>50</sub> > 1 µg/mL). To the best of our knowledge, carmabin A (**13**), dragomabin (**14**), and dragonamide A (**11**) were the only *Lyngbya* metabolites showing good antimalarial activity (IC<sub>50</sub> = 4.3, 6.0, and 7.7 µM, respectively), whereas the nonaromatic analogue, dragonamide B (**12**), was inactive. The lack of activity for

dragonamide B (**12**) suggests that an aromatic amino acid at the carboxy terminus is necessary for antimalarial activity in this compound series. Carmabin A (**13**) was more

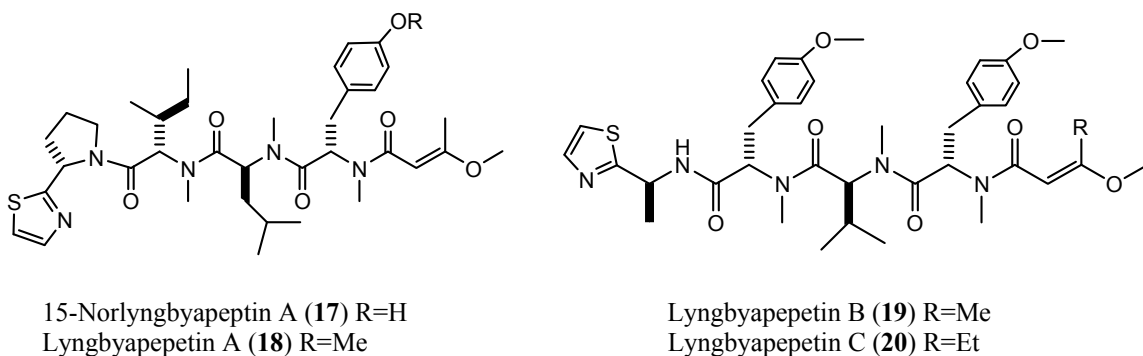


Name	R1	R2	R3	R4	Acyl
Dragonamide A ( <b>11</b> )	i-Pr	i-Pr	i-Pr	4-(MeO)PhCH <sub>2</sub> -	
Dragonamide B ( <b>12</b> )	i-Pr	i-Pr	i-Pr	i-Pr	
Dragonamide C ( <b>15</b> )	i-Pr	i-Pr	i-Pr	i-Pr	
Dragonamide D ( <b>16</b> )	i-Pr	i-Pr	i-Pr	i-Pr	
Carmabin A ( <b>13</b> )	Bn	Me	Me	4-(MeO)PhCH <sub>2</sub> -	
Dragomabin A ( <b>14</b> )	Me	Bn	Me	Me	

cytotoxic to Vero cells ( $IC_{50} = 9.8 \mu M$ ) than dragomabin ( $IC_{50} = 182.3 \mu M$ ) or dragonamide A ( $IC_{50} = 67.8 \mu M$ ), and thus, dragomabin possesses the best differential toxicity between parasite and mammalian cells. The *Vero cell* line is derived from kidney epithelial cells of the African Green Monkey. It appears that the presence of three extra carbons in the aliphatic chain leads to the increase in cytotoxicity of carmabin A (**13**) over that of dragomabin. Dragonamides C (**15**) and D (**16**) showed weak activity in cancer cell viability assays, with  $GI_{50}$  values of 56 and 59  $\mu M$  against U2OS

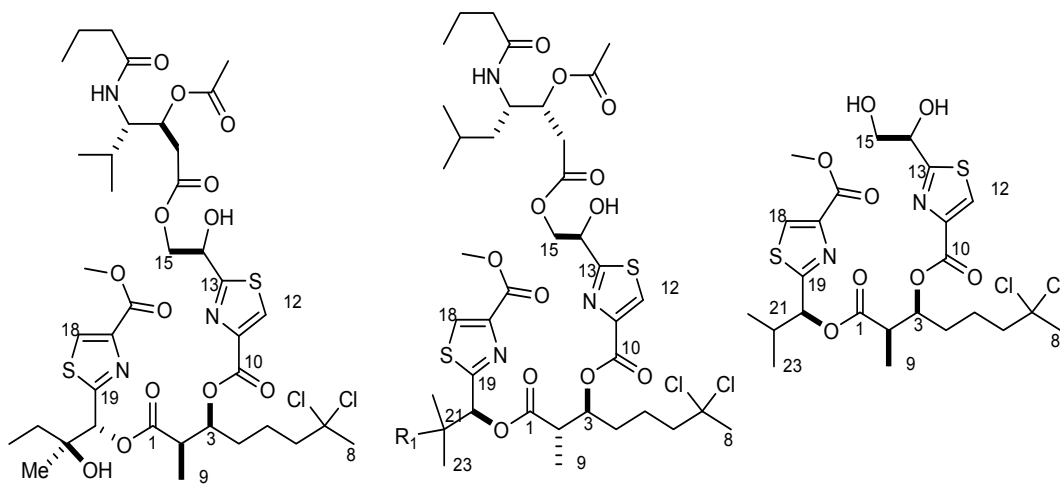
osteosarcoma cells, 22 and 32  $\mu\text{M}$  against HT29 colon adenocarcinoma cells, and 49 and 51  $\mu\text{M}$  against IMR-32 neuroblastoma cells, respectively. These data are similar to cytotoxicity data reported for dragonamides A (**11**) and B (**12**) against other cell lines.

Several collections of *Lyngbya* sp. VP417, from Finger's Reef, Apra Harbor, Guam and *Lyngbya* sp. NIH309 from Palau have yielded three more members of the lyngbyapeptin family, termed 15-norlyngbyapeptin A (**17**) (Williams et al., 2003c), lyngbyapeptins B and C (**19-20**) (Luesch et al., 2002b). In comparison to lyngbyapeptin A (**18**), the peptides **19** and **20** possess the same (for **19**) or similar (for **20**) characteristic C and N-terminal modification and differ by containing other amino acid units in between. Like lyngbyapeptin A (**18**), compounds **19** and **20** were noncytotoxic against KB and LoVo cells at  $<5$  mM.



Three acyclic members of the lyngbyabellin family, D (**21**) (Williams et al., 2003c), F (**22**) and I (**23**) (Han et al., 2005) have been isolated from *Lyngbya* sp. VP417 collected from Finger's Reef, Apra Harbor, Guam and *L. majuscula* collected in Hector Bay, Jamaica and Alotau Bay, Papua New Guinea, respectively. Structurally, lyngbyabellins D (**21**), F and I resemble dolabellin (**24**) isolated from the Japanese sea

hare *Dolabella auricularia* (Sone et al., 1995). Lyngbyabellin D (**21**), containing a rare  $\gamma$ -amino- $\beta$ -hydroxy acid, is the largest member of the lyngbyabellin family. There is no general method for determining the configuration of the chiral centers in  $\gamma$ -amino- $\beta$ -hydroxy acids. However, Rinehart reported that 4-amino-3-hydroxy-5-methylheptanoic acid underwent an epimerization at C-3 via an acid-catalyzed dehydration/hydration sequence, and that significant quantities of the intermediate  $\alpha$ ,  $\beta$ -unsaturated acid existed after prolonged hydrolysis (Rinehart et al., 1992). The absolute configuration of C-29 could be determined by acid hydrolysis and subsequent ozonolysis, with oxidative workup, to *D*- or *L*-valine. Chiral HPLC showed unambiguously that C-29 was derived from *L*-valine. H-28 and H-29 exhibited an 8.4 Hz



Lyngbyabellin D (**21**)

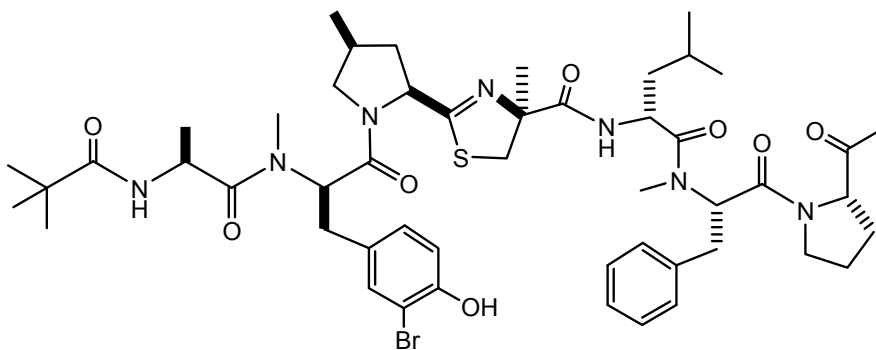
Lyngbyabellin F (**22**) R<sub>1</sub>=OH  
Lyngbyabellin I (**23**) R<sub>1</sub>=H

Dolabellin (**24**)

coupling, enabling the relative configuration to be determined by a NOE experiment. Thus the stereochemistry of C-28 and C-29 in lyngbyabellin D (**21**) was assigned as 28*S*, 29*S*. Because of limited amounts of lyngbyabellin I (**23**), the absolute stereochemistry



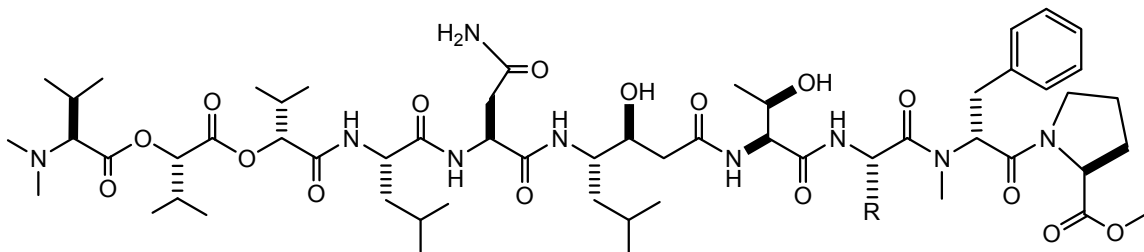
was not established by chemical methods, however, it was proposed to belong to the same enantiomeric series as lyngbyabellin F (**22**). Lyngbyabellin D (**21**) displayed an IC<sub>50</sub> value of 0.1 μM against the KB cell line. Lyngbyabellin F and I were cytotoxicity to NCI-H460 human lung tumor and neuro-2a mouse neuroblastoma cells and had LC<sub>50</sub> values between 0.7 and 1.8 mM.



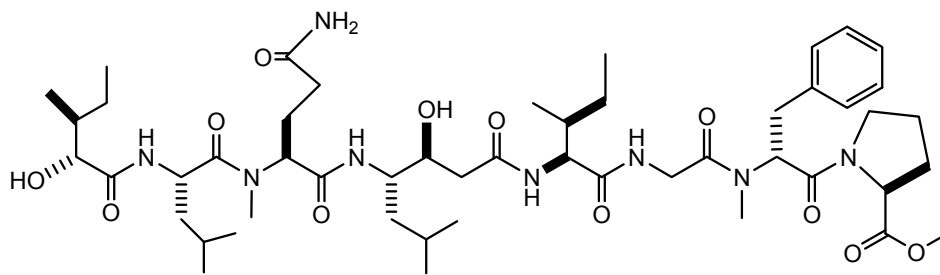
Bisebromoamide (**25**)

The efforts toward finding novel marine cyanobacterial metabolites with antitumor activity led to the isolation of bisebromoamide from a *Lyngbya* sp. harvested in Okinawa Prefecture (Teruya et al., 2009). Bisebromoamide (**25**) possesses a dense combination of unusual structural features, including a 2-substituted thiazoline-4-methyl-4-carboxylic acid unit fused to a methyl-proline. Another unusual structural element is the 2-(1-oxopropyl) pyrrolidine (Opp) residue and *N*-methyl-bromine-tyrosine. The Opp unit in bisebromoamide (**25**) is unprecedented in natural products. In addition, bisebromoamide (**25**) possesses an *N*-pivalamide moiety. Bisebromoamide (**25**) exhibited cytotoxicity against HeLa S3 cells with an IC<sub>50</sub> value of 0.04 μg/mL, and the average 50% growth inhibition (GI<sub>50</sub>) value against a panel of 39 human cancer cell lines (termed JFCR39) at

40 nM. In addition, bisbromoamide (**25**) exhibited potent protein kinase inhibition: the phosphorylation of ERK (extracellular signal regulated protein kinase) in NRK cells by PDGF (platelet-derived growth factor)-stimulation was selectively inhibited by treatment with 10 to 0.1  $\mu$ M of bisbromoamide (**25**).



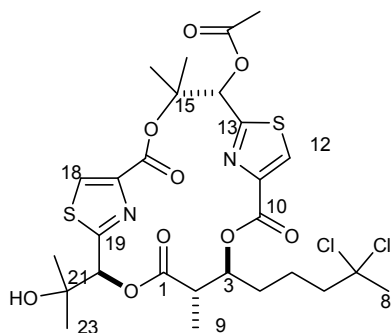
Grassystatin A (**26**) R=Me, Grassystatin B (**27**) R=Et



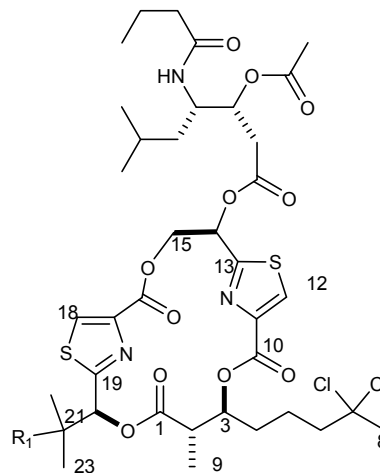
Grassystatin C (**28**)

Three statin unite-containing linear peptides, termed grassystatins A-C (**26-28**), have been isolated from the marine cyanobacteria *Lyngbya confervoides* collected off Grassy Key (Kwan et al., 2009). Grassystatins A and B (**26-27**) showed similar potency and selectivity against cathepsins D ( $IC_{50}$ s of 26.5 nM and 7.27 nM, respectively) and E ( $IC_{50}$ s of 886 pM and 354 pM, respectively). The increased affinity for cathepsin E over cathepsin D may be interpreted by the interaction of the polar residue asparagine in grassystatins A and B (**26-27**) with glutamine-303 in cathepsin E, which replaces the

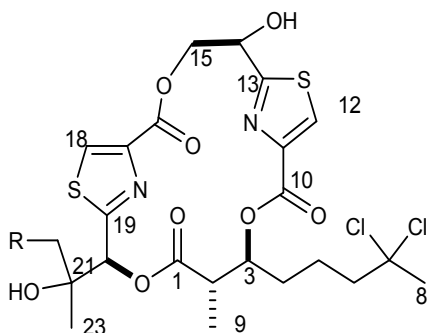
nonpolar residue methionine-307 in cathepsin D. Grassystatin A was observed to reduce antigen presentation by dendritic cells, a process thought to rely on cathepsin E. The truncated peptide analogue grassystatin C (**28**), which consists of two fewer residues than



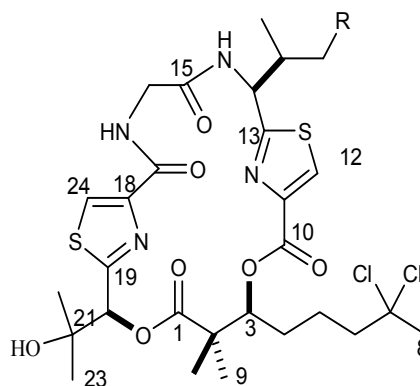
Hectochlorin (**29**)



Lyngbyabellin E (**31**) R=OH  
Lyngbyabellin H (**33**) R=H



Lyngbyabellin C (**30**) R=Me  
Lyngbyabellin G (**32**) R=H



Lyngbyabellin A (**34**) R=Me  
Lyngbyabellin B (**35**) R=H

grassystatins A and B (**26-27**), was less potent against both cathepsins D and E, but still selective for cathepsin E. The selectivity of grassystatins A-C (**26-28**) for cathepsin E

over D (20-38-fold) suggests that these natural products may be useful tools to probe cathepsin E function.

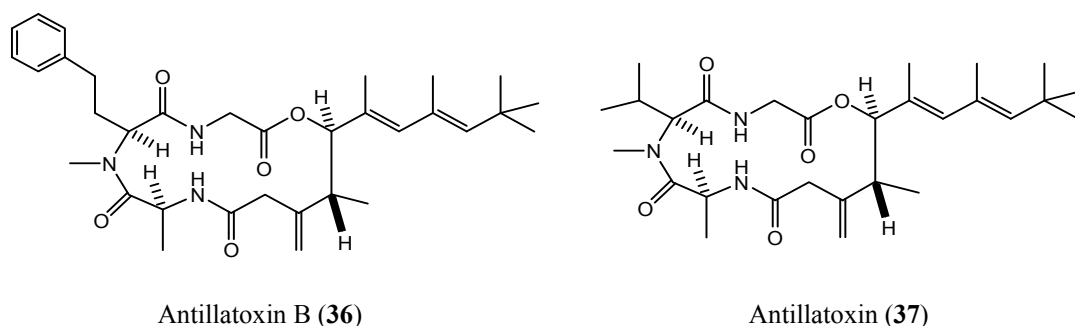
### 1.1.2 Cyclic polypeptides

Hectochlorin (**29**) (Marquez et al., 2002), lyngbyabellins C (**30**) (Luesch et al., 2002b), E (**31**), G-H (**32-33**) (Han et al., 2005) have been reported from *Lyngbya*. Structurally, these five cyclic compounds resemble lyngbyabellins A (**34**) (Luesch et al., 2000a) and B (**35**) (Luesch et al., 2000b). Hectochlorin possesses the same absolute configuration as dolabellin and lyngbyabellins E, G and H except at the C14. Hectochlorin has potent antifungal activity against *Candida albicans* and antiproliferative activity in the NCI 60-cell line assay. The NCI 60-cell line assay utilized 60 different human tumor cell lines, representing leukemia, melanoma and cancers of the lung, colon, brain, ovary, breast, prostate, and kidney. Hectochlorin had an average GI<sub>50</sub> against the NCI 60 cell lines of 5.1  $\mu$ M and showed greatest potency against the colon, melanoma, ovarian, and renal panels. The IC<sub>50</sub> values for in vitro cytotoxicity for lyngbyabellin C were 2.2 mM against KB and 3.3 mM against LoVo cells. The lyngbyabellins E-I were tested for cytotoxicity to NCI-H460 human lung tumor and neuro-2a mouse neuroblastoma cells and had LC<sub>50</sub> values between 0.2 and 4.8 mM. Intriguingly, lyngbyabellin E (**31**) and H (**33**) appeared to be more active against the H460 cell line with LC<sub>50</sub> values of 0.4 and 0.2 mM, respectively, compared to LC<sub>50</sub> values of 1.2 and 1.4 mM in the neuro-2a cell line. Lyngbyabellin I (**23**) was the most toxic to neuro-2a cells (LC<sub>50</sub> = 0.7 mM), whereas lyngbyabellin G (**32**), was the least cytotoxic of all compounds to either cell line. On the basis of this limited screening, it appears that lung tumor cell toxicity is enhanced in the

cyclic representatives with an elaborated side chain. However, the fact that the acyclic compounds (lyngbyabellin F and I and dolabellin) are still relatively potent cytotoxins suggests two possibilities. One explanation is that the acyclic compounds may adopt a conformation similar to the cyclic forms at their target. It is also possible that the acyclic forms are produced in vitro by intracellular esterases and represent the truly bioactive forms of these molecules. If this latter hypothesis is correct, then differences in potency between the cyclic and acyclic forms could be the result of differences in their permeability to cells. Unfortunately, the results are difficult to compare with those obtained for lyngbyabellins A–D, dolabellin, and a few other analogs produced synthetically or semi-synthetically because different cell lines were used (e.g., much of the present work is reported using the KB cell line, which is considerably more drug sensitive than the H-460 and neuro-2a cell lines). However, the trend in the data supports the conclusion drawn above that an overall cyclic constitution is not required for potent cytotoxic properties in this drug class. Lyngbyabellin E (**31**) was also tested for the cytoskeletal-disrupting activity because lyngbyabellin A and hectochlorin have been reported to be strong actin-disrupting agents (Luesch et al., 2000a; Marquez et al., 2002). Lyngbyabellin E (**31**), at concentrations of 0.01–6.0 mM, disrupted the cellular microfilament network in A-10 cells. Additionally, at the higher concentrations tested, many cells contained two nuclei, consistent with the inhibition of cytokinesis that often occurs after disruption of the microfilament network. The effects were specific for microfilaments, as there was no evidence of microtubule loss at these concentrations.

Antillatoxin B (**36**) was isolated from two *Lyngbya majuscula* collections from Collado Reef, Puerto Rico, and Bush Key, Dry Tortugas (Nogle et al., 2001a). Its

structure was deduced from 2D NMR and data comparisons with antillatoxin (**37**). Antillatoxin B contains an unusual *N*-methyl homophenylalanine which was only shown in a few cyanobacterial metabolites. Both antillatoxin (**37**) and antillatoxin B (**36**) are potent activators of the voltage-sensitive sodium channel in mouse neuro-2a neuroblastoma cells with EC<sub>50</sub> values of 0.18 and 1.77 μM, respectively. In addition, antillatoxin B (**36**) is strongly ichthyotoxic to goldfish, with an LC<sub>50</sub> of 1.0 μM. However,

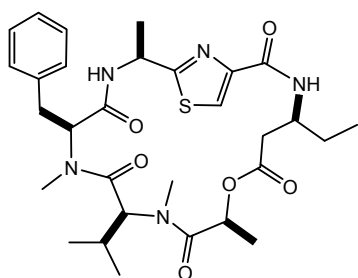


antillatoxin (**37**) is more toxic to goldfish than antillatoxin B (**36**), with a LC<sub>50</sub> value at 0.1 μM. It is interesting to note that in both measures of the biological activity of antillatoxin B's (**36**), a 10-fold decrease in potency is observed compared to antillatoxin (**37**). It seems likely that substitution of a larger *N*-methyl homophenylalanine residue for an *N*-methyl valine residue accounts for this decrease.

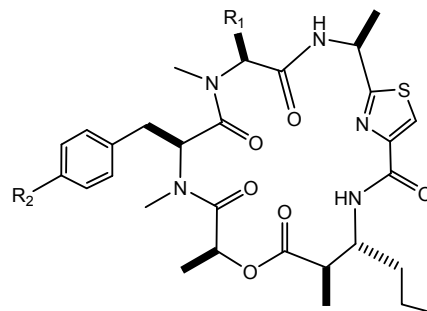
Obyanamide (**38**) was isolated from the marine cyanobacterium *Lyngbya confervoides* collected in Saipan, Commonwealth of the Northern Mariana Islands (Williams et al., 2002). Obyanamide (**38**) is a novel depsipeptide containing 2 *N*-methyl amino acids, an Ala-thiazole unit, and a β-amino acid. The absolute stereochemistry of (**38**) was deduced by chiral chromatography of the hydrolysis products and comparison of

the retention time with authentic and synthetic standards. Obyanamide (**38**) was cytotoxic against KB and LoVo cells with an  $IC_{50}$  of 0.58 and 3.14  $\mu\text{g/mL}$ , respectively.

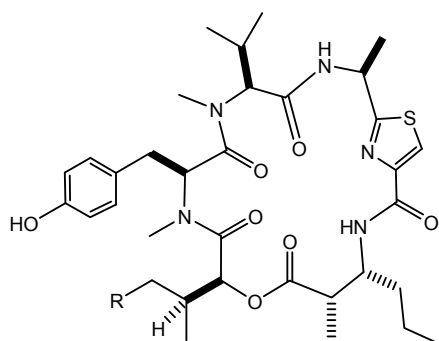
Six  $\beta$ -amino acid-containing cyclic depsipeptides, termed ulongamides A-F (**39-44**), were isolated from collections of apratoxin-producing *Lyngbya* sp. NIH309 from



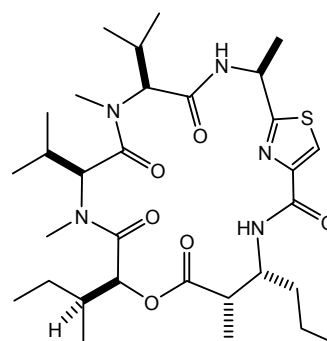
Obyanamide (**38**)



Ulongamide A (**39**)  $R_1=i\text{-Pr}$ ,  $R_2=\text{H}$   
 Ulongamide B (**40**)  $R_1=i\text{-Pr}$ ,  $R_2=\text{OH}$   
 Ulongamide C (**41**)  $R_1=\text{Bn}$ ,  $R_2=\text{OH}$



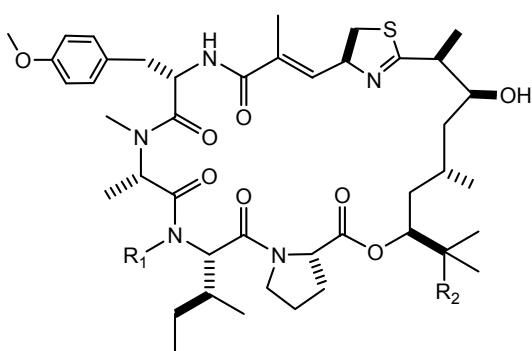
Ulongamide D (**42**)  $R=\text{H}$   
 Ulongamide E (**43**)  $R=\text{Me}$



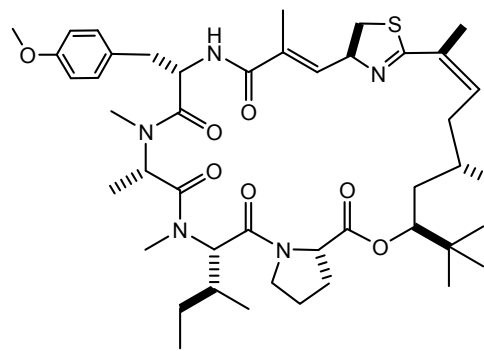
Ulongamide F (**44**)

Palau (Luesch et al., 2002a). The stereochemistry of the  $\beta$ -amino acid moiety, 3-amino-2-methylhexanoic acid, was established by advanced Marfey analysis of the acid hydrolyzates and found to be  $2R$ ,  $3R$  in compounds **39-41** but  $2S$ ,  $3R$  in compounds **42-44**. In comparison to obyamide (**38**), which bears a 3-aminopentanoic acid moiety, compounds **39-44** contain a different  $\beta$ -amino acid unit. Remarkably, ulongamide A (**39**)

differs from obyanamide otherwise merely in the residue sequence, with the *N*-methyl-valine and the *N*-methyl-phenylalanine residues being switched. The great structural variety of the ulongamides is intriguing and might suggest that enzymes with relaxed substrate specificity are involved in their biosynthesis or that slight genetic divergence among individual patches of this cyanobacterium might be responsible for the diversity. All compounds except **44** were weakly cytotoxic against KB cells. It suggests that the lack of an aromatic amino acid or the *N*-methyl group adjacent to the hydroxyl acid is detrimental to activity.



Apratoxins A (**45**)  $R_1=Me$ ,  $R_2=Me$   
 Apratoxins B (**46**)  $R_1=H$ ,  $R_2=Me$   
 Apratoxins C (**47**)  $R_1=Me$ ,  $R_2=H$

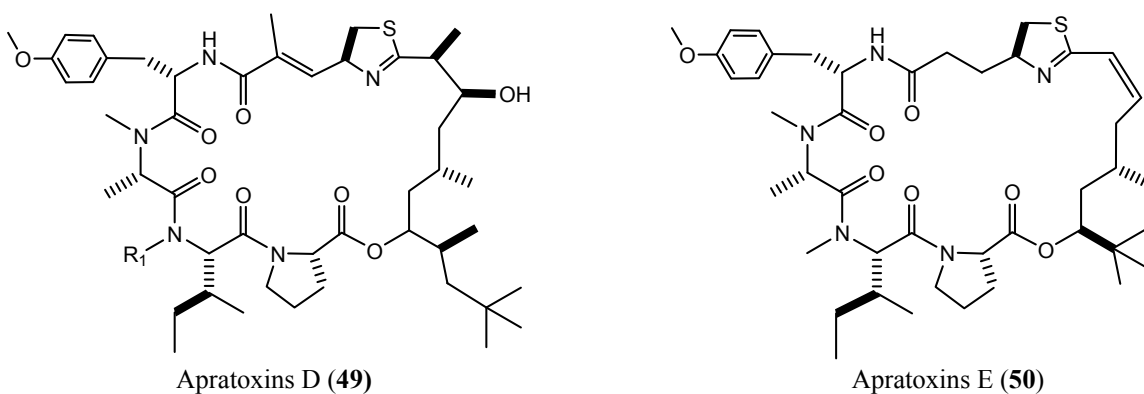


*E*-dehydroapratoxin A (**48**)

Two collections of *Lyngbya* sp. from Guam and Palau that both afforded the potent cytotoxin apratoxin A (**45**) (Luesch et al., 2001a) yielded different apratoxins analogues, apratoxins B (**46**) and C (**47**) (Luesch et al., 2002c). Apratoxins A-C (**45-47**) are presumably of mixed peptide-polyketide biogenesis. Apratoxin A (**45**) possesses  $IC_{50}$  values for in vitro cytotoxicity against human tumor cell lines ranging from 0.36 to 0.52 nM. It was only marginally active in vivo against a colon tumor and ineffective against a



mammary tumor. However, its potential as an antitumor agent was limited by the lack of selectivity. Apratoxins B (**46**) and C (**47**), were evaluated for their in vitro cytotoxicity along with semisynthetic *E*-dehydroapratoxin A (**48**) to identify key structural elements responsible for the cytotoxicity of this novel family of depsipeptides. All analogues **46-48** displayed weaker cytotoxicity than **45**, but to different extents. The in vitro cytotoxicity of apratoxin C (**47**) almost approached that of apratoxin A (**45**), indicating that the *tert*-butyl group is not essential for activity and can be replaced at least by the sterically less demanding isopropyl group. Apratoxin B (**46**) and especially *E*-dehydroapratoxin A (**48**)



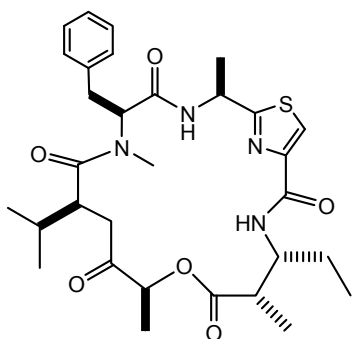
were significantly less cytotoxic than apratoxin A (**45**) and apratoxin C (**47**). Cytotoxicity decreasing of (**46**) and (**48**) suggested that the *N*-methyl group of the Ile residue might play a role in the cytotoxicity as well as the 1,2-methine system (C34–C35) bearing the hydroxyl group. The NMR data also revealed a conformational change for compounds 2 and 4 in CDCl<sub>3</sub>. Molecular modeling studies for apratoxin A (**45**) in CDCl<sub>3</sub> demonstrated the spatial proximity of the *N*-methyl group of Ile and the hydroxyl group of the Dtena (3,7-dihydroxy-2,5,8,8-tetramethylnonanoic acid) unit. Perhaps this conformational

behavior is important under aqueous assay conditions and contributes to the observed activity loss as well.

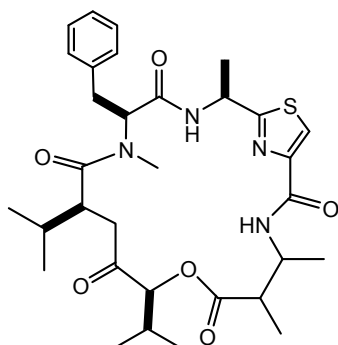
Apratoxin D was isolated from the Papua New Guinea marine cyanobacteria *Lyngbya majuscula* and *Lyngbya sordida* (Gutiérrez et al., 2008). Apratoxin D (**49**) contains the same macrocycle as apratoxins A (**45**) and C (**47**) but possesses the novel 3,7-dihydroxy-2,5,8,10,10-pentamethylundecanoic acid as the polyketide moiety. Apratoxin D (**49**) showed potent in vitro cytotoxicity against H-460 human lung cancer cells with an IC<sub>50</sub> value of 2.6 nM, which is nearly equipotent to that of apratoxin A (**45**). The similar cytotoxicity of apratoxin A (**45**) and Apratoxin D (**49**) indicates that the cytotoxicity of apratoxins is not strongly impacted by the larger lipopeptide tail. This relationship could be important to the design of analogue structures for probing the mechanism of action of the apratoxins. The isolation of apratoxin D from two distinct species of the genus *Lyngbya* supports the hypothesis that genetic transfer of natural product biosynthetic pathways between different marine cyanobacteria is common (Simmons et al., 2008).

A collection of the marine cyanobacterium *Lyngbya bouillonii* from Guam afforded apratoxin E (**50**), an apratoxin cytotoxin (Matthew et al., 2008b). Apratoxin E (**50**) displayed strong cytotoxicity against several cancer cell lines derived from colon, cervix, and bone, ranging from 21 to 72 nM, which is lower than its closest analogues, *E*-dehydroaparotoxin (**48**), yet 5- to 15-fold less active than apratoxin A (**45**). It suggests that the  $\alpha,\beta$ -unsaturation of the modified cysteine residue is not essential for apratoxin activity. The 5- to 15-fold reduced activity of apratoxin E (**50**) compared with apratoxin

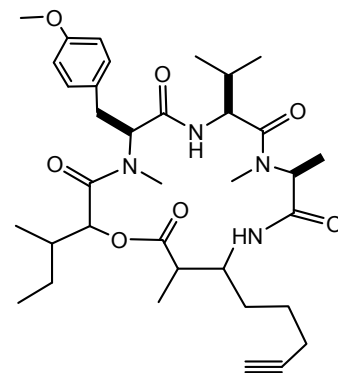
A (45) is attributed to the dehydration in the long-chain polyketide unit, which could affect the conformation of the molecule.



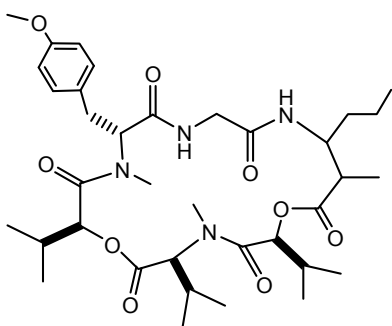
Guineamide A (51)



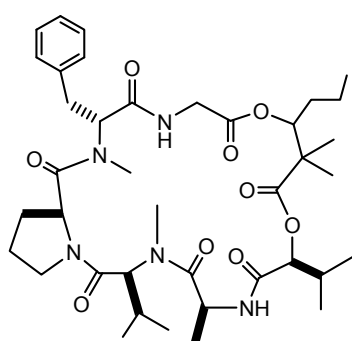
Guineamide B (52)



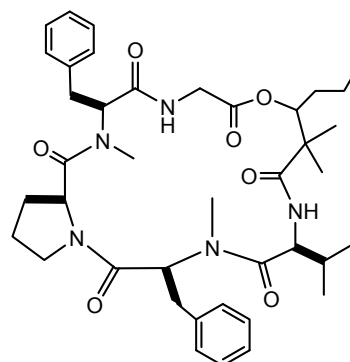
Guineamide C (53)



Guineamide D (54)



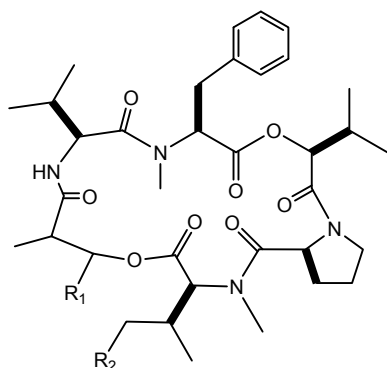
Guineamide E (55)



Guineamide F (56)

Guineamides A-F (51-56) are novel cyclic depsipeptides isolated and characterized from a Papua New Guinea collection of the marine cyanobacterium *Lyngbya majuscula* (Tan et al., 2003). All six possessed *N*-methylated amino acids and  $\beta$ -amino acids, or  $\beta$ -hydroxyl groups. The  $\beta$ -amino and  $\beta$ -hydroxyl residues are an increasingly common feature in marine cyanobacterial metabolites. The three  $\beta$ -amino acid residues present in

the guineamides [Mapa (2-methyl-3-aminobutanoic acid) in **51**, Maba (2-methyl-3-aminobutanoic acid) in **52**, and Maoya (2-methyl-3-amino-oct-7-ynoic acid) in **53**] are also present as constituents in molluscan metabolites (e.g., dolastatins D, **11**, and **12** and onchidin A). It was speculated that these substances are originated from the cyanobacterial diet of the mollusc. Dmhha (2,2-dimethyl-3-hydroxyhexanoic acid) in guineamides E (**55**) and F (**56**) was reported for the first time. Guineamides B (**52**) and C (**53**) possess moderate cytotoxicity to a mouse neuroblastoma cell line with IC<sub>50</sub> values of 15 and 16 μM, respectively.



Antanapeptin A (**57**)  
 $R_1 = \text{CH}_2\text{CH}_2\text{CH}_2\text{C}\equiv\text{CH}$ ,  $R_2 = \text{CH}_3$

Antanapeptin B (**58**)  
 $R_1 = \text{CH}_2\text{CH}_2\text{CH}_2\text{CH}=\text{CH}_2$ ,  $R_2 = \text{CH}_3$

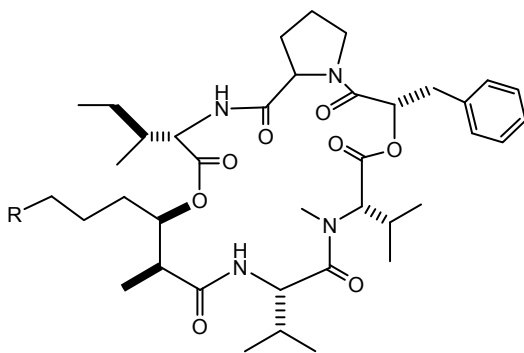
Antanapeptin C (**59**)  
 $R_1 = \text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$ ,  $R_2 = \text{CH}_3$

Antanapeptin D (**60**)  
 $R_1 = \text{CH}_2\text{CH}_2\text{C}\equiv\text{CH}$ ,  $R_2 = \text{H}$

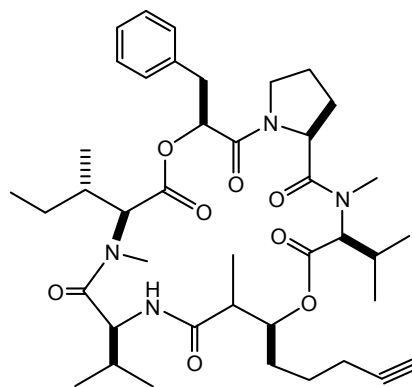
A series of depsipeptides, antanapeptins A-D (**57-60**) were isolated from a *Lyngbya majuscula* collection from Antany Mora, Madagascar (Nogle et al., 2002). The antanapeptins did not exhibit activity in several biological assays. Similarly, the related metabolite georgamide, isolated from an Australian cyanobacterial collection, did not possess biological properties (Wan et al., 2001). However, two other related metabolites, yanucamides A and B from a Fijian collection of *L. majuscula*, showed potent brine shrimp toxicity (Sitachitta et al., 2000). Thus, a specific biological function for this

structural class of compounds remains uncertain.

Trungapeptins A-C (**61-63**) were isolated from the marine cyanobacterium *Lyngbya majuscula* collected from Trung Province, Thailand (Bunyajetpong, et al., 2006). The structures of compounds **61-63** are closely related to the antanapeptins, a series of depsipeptides isolated from a Madagascan collection of *L. majuscula*. The structures of compounds **61-63** contain the unique 3-hydroxy-2-methyl-7-octynoic acid (Hmoya), 3-hydroxy-2-methyl-7-octenoic acid (Hmoea), and 3-hydroxy-2-methyl-7-octanoic acid (Hmoaa) residues, respectively. The relative stereochemistry of 3-hydroxy-2-methyl-7-octynoic acid (Hmoya) of trungapeptin A was elucidated by application of the *J*-based configuration analysis (Matsumori et al., 1999). Its absolute stereochemistry



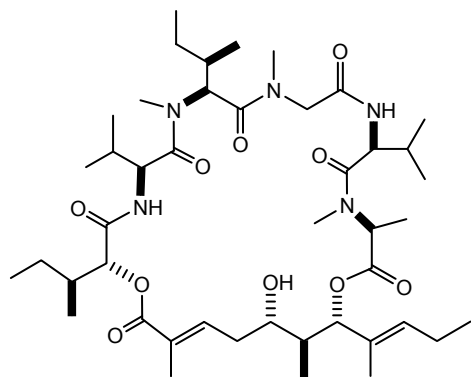
Trungapeptin A (**61**) R=CH<sub>2</sub>C≡CH  
Trungapeptin B (**62**) R= CH<sub>2</sub>CH=CH<sub>2</sub>  
Trungapeptin C (**63**) R= CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>



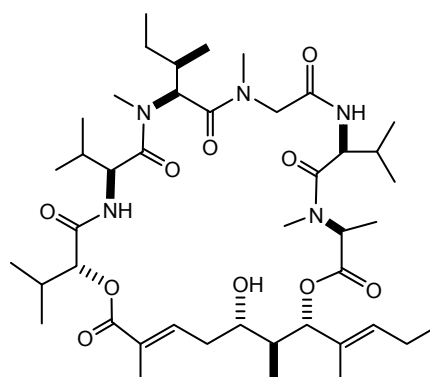
Huntupeptin A (**64**)

was established to be 2*S*, 3*R* by Mosher's method. Trungapeptin A (**61**) showed mild brine shrimp toxicity at 10 ppm and mild ichthyotoxicity at 6.25 ppm but did not exhibit cytotoxicity against KB and LoVo cells at 10 μg/mL. Due to a limited supply of compounds **62** and **63**, their biological activities were not evaluated.

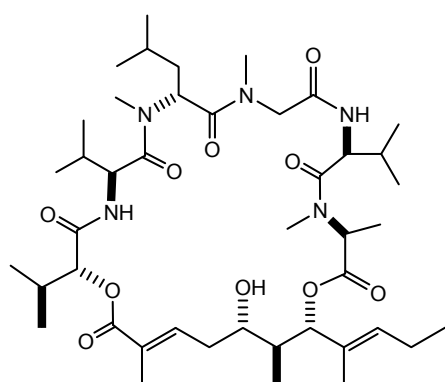
The marine *Lyngbya majuscula* from Pulau Hantu Besar, Singapore afforded a cyclodepsipeptide, hantupeptin A (Tripathi et al, 2009). Hantupeptin A (**64**) showed 100% brine shrimp toxicity at 100 and 10 ppm, which was significantly higher than the activity reported for its close analogue trungapeptin A (**61**), which exhibited mild brine



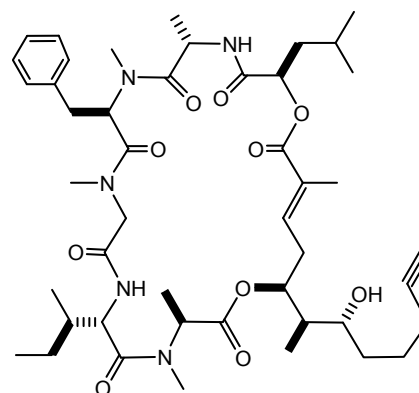
Aurilide B (**65**)



Aurilide C (**66**)



Aurilide (**67**)



Palau'amide (**68**)

shrimp toxicity. Furthermore, *in vitro* cytotoxicity testing of **64** against the leukemia cell line MOLT-4 with an  $IC_{50}$  value of 32 nM. Compound **64** also displayed significant cytotoxicity in the breast cancer cell line MCF-7 with an  $IC_{50}$  value of 4.0  $\mu$ M. In

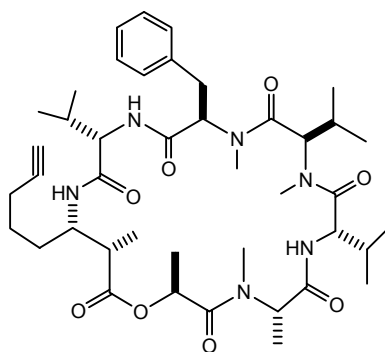
comparison, trungapeptin A (**61**), was reported to be inactive when tested at 10  $\mu\text{g/mL}$  against KB or LoVo cells.

Aurilides B (**65**) and C (**66**) were isolated from a strain of the marine cyanobacterium *Lyngbya majuscula* collected from Alotau Bay, Papua New Guinea (Han et al., 2006). Aurilides B and C are closely related to aurilide (**67**), originally isolated from the sea hare *Dolabella auricularia*. The isolation of aurilides B (**65**) and C (**66**) from the marine cyanobacterium *L. majuscula* clearly indicates that the sea hare *D. auricularia* obtains and accumulates aurilide (**67**) from its diet of marine cyanobacteria. Both aurilides B and C showed in vitro cytotoxicity toward NCI-H460 human lung tumor and the neuro-2a mouse neuroblastoma cell lines, with  $\text{LC}_{50}$  values between 0.01 and 0.13  $\mu\text{M}$ . Aurilide B (**65**) was evaluated in the NCI 60 cell line panel and exhibited a high level of cytotoxicity (the mean panel  $\text{GI}_{50}$  concentration was less than 10 nM) and to be particularly active against leukemia, renal and prostate cancer cell lines.

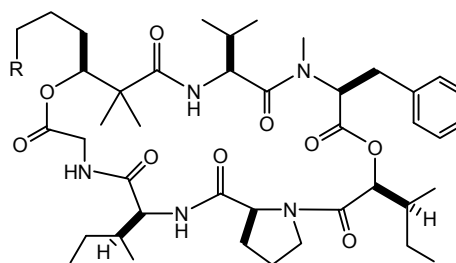
A species of *Lyngbya* from Ulong Channel in Palau has yielded palau'amide (**68**), which displayed strong cytotoxicity against KB cells with an  $\text{IC}_{50}$  value of 13 nM (Williams et al., 2003b).

Ulongapeptin (**69**), a cyclic depsipeptide, was isolated from a Palauan marine cyanobacterium *Lyngbya* sp. (Williams et al., 2003d). The absolute and relative stereochemistry of the  $\beta$ -amino acid, 3-amino-2-methyl-7-octynoic acid (AMO) in **69** was determined by synthesis of the saturated  $\alpha$ -alkyl- $\beta$ -amino acid and Marfey's analysis of the acid hydrolysate of tetrahydro-**69**. Two of the four valine units in ulongapeptin (**69**) are *N*-methylated, a percentage that mirrors that found in the literature. The presence of two *D*-configuration amino acids is unusual though since over 90% of the amino acids

incorporated into cyanobacterial metabolites have an *L* configuration. The AMO unit found in **69** has been previously identified in the molluskan metabolite onchidin (Fernandez-Suarez et al., 1997). Such  $\beta$ -amino acid units appear to be ubiquitous to cyanobacterial metabolites, and their appearance in compounds isolated from other marine organisms has often been suggested a dietary or symbiotic relationship between the two organisms. Ulongapeptin (**69**) was cytotoxic against KB cells at an IC<sub>50</sub> value of 0.63  $\mu$ M.



Ulongapeptin (**69**)

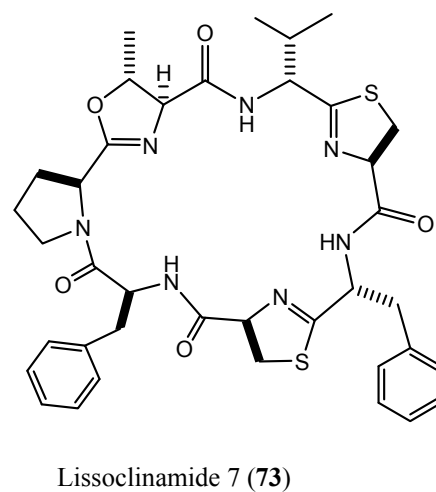
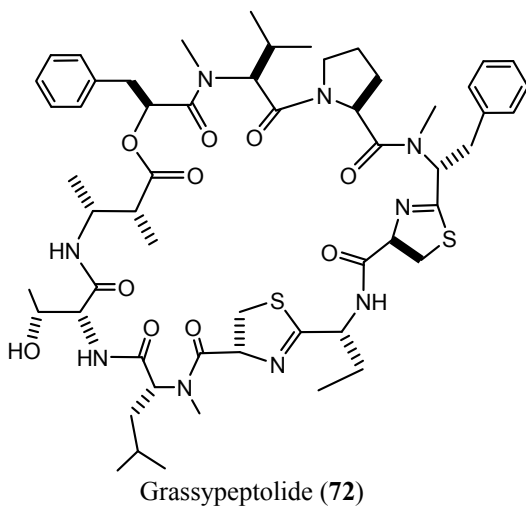


Pitipeptolide A (**70**) R= CH<sub>2</sub>C≡CH  
 Pitipeptolide B (**71**) R= CH<sub>2</sub>CH=CH<sub>2</sub>

Two cyclodepsipeptides, pitipeptolides A (**70**) and B (**71**), have been isolated from the marine *Lyngbya majuscula* collected at Piti Bomb Holes, Guam (Luesch et al., 2001b). The most distinctive features of pitipeptolides A (**70**) and B (**71**), the 2,2-dimethyl-3-hydroxy-7-octynoic acid (Dhoya) and 2,2-dimethyl-3-hydroxy-7-octenoic acid (Dhoea) units, are found in some other marine natural products such as the kulolides, isolated from marine mollusks known for accumulating cyanobacterial metabolites via their diet, and yanucamides A and B (Dhoya), directly obtained from a cyanobacterial source. The isolation of compounds **70** and **71** provides more evidence that Dhoya and



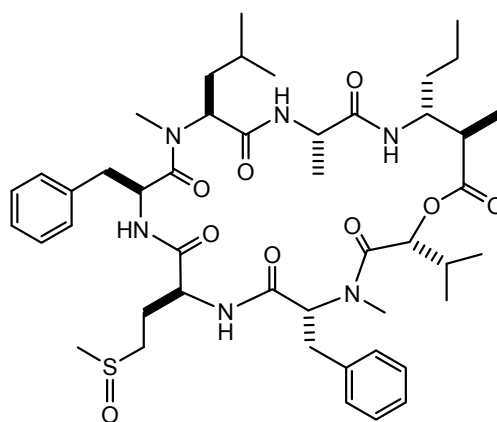
Dhoha units are diagnostic for metabolites of cyanobacterial origin and that they can be considered biosynthetic signatures. Pitiptolides A (**70**) and B (**71**) exhibit weak cytotoxicity against LoVo cancer cells, but possess moderate antimicrobial activity and stimulate elastase activity.



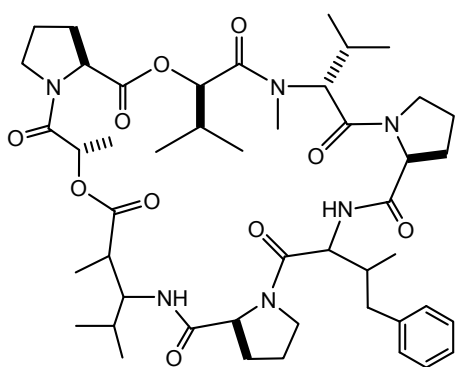
A collection of the cyanobacterium *Lyngbya confervoides* from the Florida Keys yielded grassypeptolide (**72**), a macrocyclic depsipeptide with unusually high *D*-amino acid content, two thiazolines, and one  $\beta$ -amino acid (Kwan et al., 2008). Grassypeptolide (**72**) inhibited four cancer cell lines derived from human osteosarcoma (U2OS), cervical carcinoma (HeLa), colorectal adenocarcinoma (HT29), and neuroblastoma (IMR-32) with  $IC_{50}$  values from 1.0 to 4.2  $\mu$ M. These data are within the range of  $IC_{50}$  values reported for lissoclinamide 7 (**73**) (53.7 nM to 21.5  $\mu$ M), but in different cell lines (Hawkins et al., 1990; Wipf et al., 1998). Lissoclinamide 7 (**73**), the most cytotoxic of the lissoclinamide series, has two thiazoline rings with the same arrangement and stereoconfiguration as **72**. It has been shown that the thiazolines of **73** are important to its

cytotoxic activity (Wipf et al., 1998). It is tempting to speculate that the thiazolines might also be responsible for the activity of **72**, and that it might indicate a shared mechanism of action with the lissoclinamides (**73**).

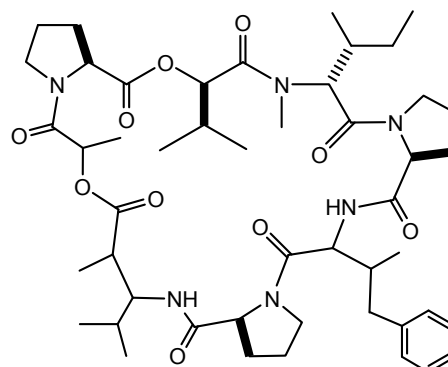
A cyclodepsipeptide, termed carriebowmide (**74**) was isolated from *Lyngbya polychroa* collected from Carrie Bow Cay, Belize (Gunasekera et al., 2008b). Carriebowmide (**74**) contains two rare amino acids, 3-amino-2-methylhexanoic acid and methionine sulfoxide.



Carriebowmide (**74**)



Dolastatin 16 (**75**)



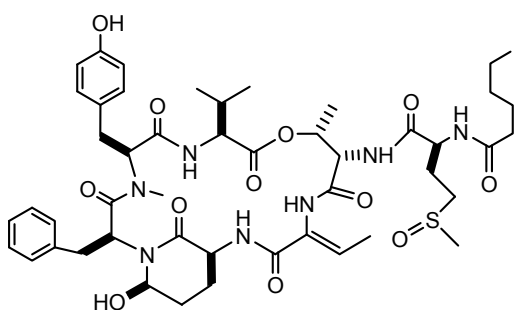
Homodolastatin 16 (**76**)

The known dolastatin 16 (**75**) was isolated from a *Lyngbya majuscula* collection from Antany Mora, Madagascar (Nogle et al., 2002). The isolation of dolastatins 3, 10,

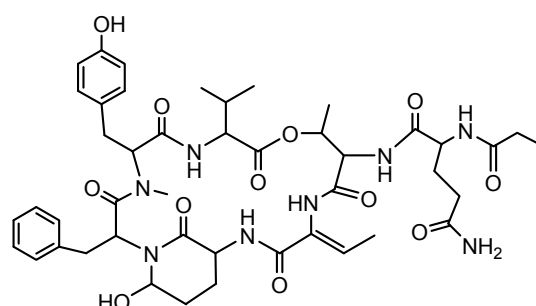
12, and 16 from various microalgal collections demonstrate their origin in cyanobacteria. In addition to a known cyclic depsipeptide antanapeptin A (**57**), a bioactive cyclic depsipeptide, homodolastatin 16 (**76**) has been isolated from the cyanobacterium *Lyngbya majuscula*, collected from Wasini Island off the southern Kenyan coast (Davies-Coleman et al., 2003). The stereochemistry of the 3-hydroxy-2-methyl-7-octynoic acid (Hmoya) in **57** and the dolaphenvaline (Dpv) and dolamethylleucine (Dml) fragments in **76** remains unassigned. The *L*-configuration of the lactate residue in dolastatin 16 (**75**) was originally established from chiral HPLC data. However, several attempts, using both chiral GC analysis and HPLC (with precolumn derivatization), failed to reproducibly establish the stereochemistry of the lactate residue in homodolastatin 16 (**76**). In spite of the strong activity of dolastatin 16 (**75**) against a wide variety of cancer cell lines, e.g., lung (NCI-H460: GI<sub>50</sub>=0.00096 µg/mL), colon (KM20L2 GI<sub>50</sub>= 0.0012 µg/mL), brain (SF-295, GI<sub>50</sub>= 0.0052 µg/mL), and melanoma (SK-MEL5 GI<sub>50</sub>= 0.0033 µg/mL) (Nakao et al., 1998), homodolastatin 16 (**76**), as a higher homologue of dolastatin 16 (**75**), exhibited moderate activity against esophageal (WHCO1 IC<sub>50</sub>=4.3 µg/mL and WHCO6 IC<sub>50</sub>=10.1 µg/mL) and cervical cancer cell lines (ME180 IC<sub>50</sub>=8.3 µg/mL). One possible reason for the difference in their anticancer activities resulted from the different cancer cell lines used for homodolastatin 16 (**76**) and dolastatin 16 (**75**). Unfortunately, the comparative analysis using esophageal and cervical cancer cell lines was carried out because the Davies-Coleman's group had no dolastatin 16 (**75**).

Examination of the mixed assemblages of *Lyngbya majuscula* and *Schizothrix* sp. from the Fijian Islands led to the isolation of somamides A (**77**) and B (**78**) which are analogues in structure to symprostatin 2 (**80**) and dolastatin 13 (**79**) (Nogle et al., 2001b;

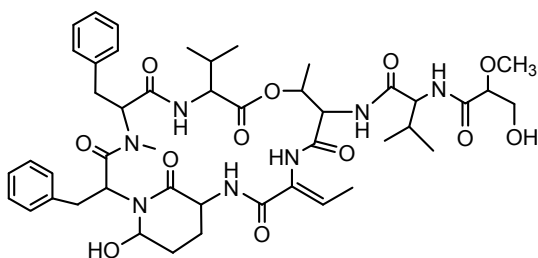
Pettit et al., 1989; Harrigan, et al., 1999). Dolastatin 13 previously been reported from marine invertebrates, including the sea hare, *D. auricularia*, well known for its ability to accumulate compounds acquired from its diet of cyanobacteria (Harrigan, et al., 2000). Thus, isolation of these depsipeptides provides further evidence that such metabolites are truly of cyanobacterial origin. So far dolastatin 13 has been found to strongly inhibit



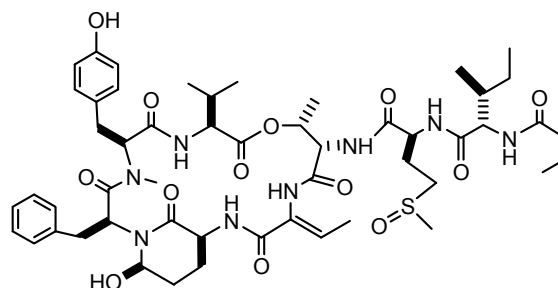
Somamides A (77)



Somamides B (78)



Dolastatin 13 (79)

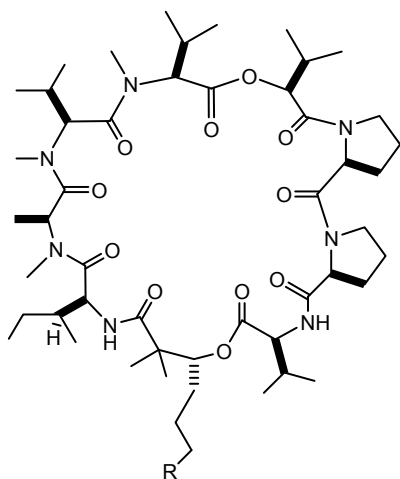


Symplostatin 2 (80)

growth of the PS cell line derived from pig kidney and exhibit an  $ED_{50}$  value of  $0.013\mu\text{g/mL}$ , whereas dehydrolastatin 13 proved to be marginally inactive in this system and revealed the first structure/activity insight. Pig kidney cells of the PS line

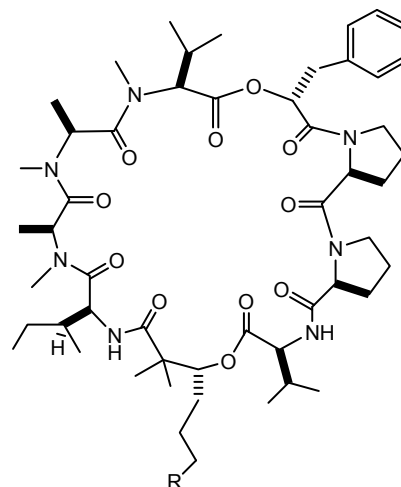
have proved useful for the study of group B arboviruses (Miller et al., 1969; Westaway, 1966). There is no report about the bioactivity of somamide A (**77**) and B (**78**).

Four depsipeptides, wewakpeptins A-D (**81-84**), were isolated from the marine cyanobacterium *Lyngbya semiplena* collected from Papua New Guinea (Han et al., 2005). The wewakpeptins represent an unusual arrangement of amino and hydroxy acid subunits



Wewakpeptins A (**81**) R=CH<sub>2</sub>C≡CH

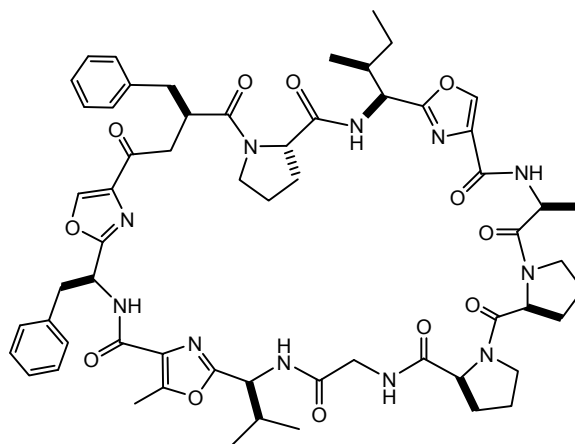
Wewakpeptins B (**82**) R=CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>



Wewakpeptins C (**83**) R=CH<sub>2</sub>C≡CH

Wewakpeptins D (**84**) R=CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>

relative to known cyanobacterial peptides, and possess a bis-ester, a 2,2-dimethyl-3-hydroxy-7-octynoic acid (Dhoya) or 2,2-dimethyl-3-hydroxyoctanoic acid (Dhoaa) residue, and a dipropyl group reminiscent of dolastatin 15. Wewakpeptin A (**81**) and B (**82**) were the most cytotoxic among these four depsipeptides with an LC<sub>50</sub> of approximately 0.4 μM to both the NCI-H460 human lung tumor and the neuro-2a mouse neuroblastoma cell lines. Intriguingly, wewakpeptins C (**83**) and D (**84**) were approximately 10-fold more toxic than A and B to these cell lines.

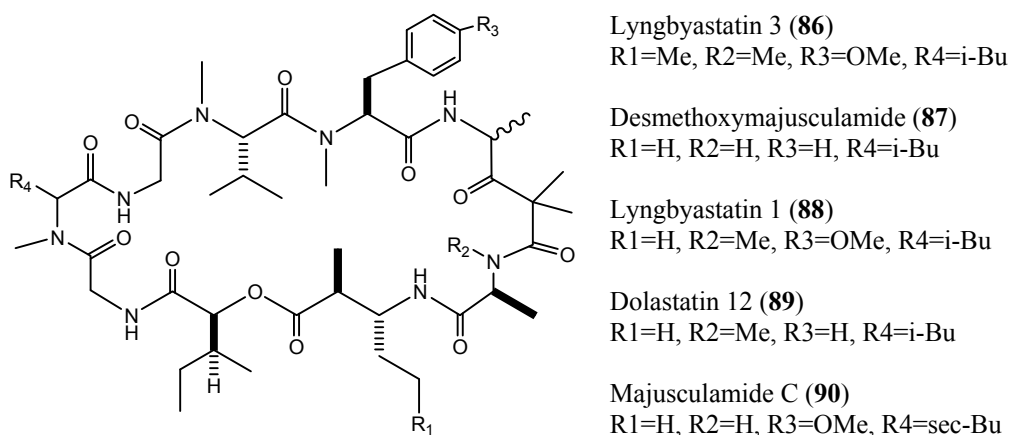


Wewakazole (**85**)

Examination of a Papua New Guinea collection of *Lyngbya majuscula* led to the discovery of a cyclic dodecapeptide, wewakazole (**85**) (Nogle et al., 2003). Wewakazole derives solely from amino acids and features three oxazole units which originate from serine. There was no report on the biological activity of wewakazole.

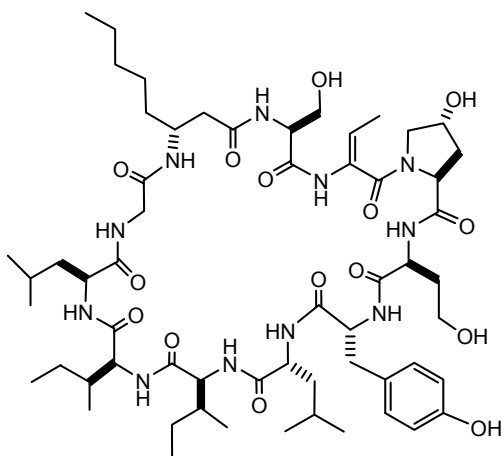
Lyngbyastatin 3 (**86**) and desmethoxymajusculamide C (**87**) were isolated from *Lyngbya majuscula* collected in Apra Harbor, Guam (Williams et al, 2003a) and Fijian (Simmons et al., 2009). The absolute configurations of the two unusual amino acid residues, the 3-amino-2-methylhexanoic acid (Amha) and 4-amino-2,2-dimethyl-3-oxopentanoic acid units (Adoa) in lyngbyastatin 3 (**86**), have been established by chemical degradation. The cyanobacterial compounds, lyngbyastatin 3 (**86**), lyngbyastatin 1 (**88**), and dolastatin 12 (**89**), are mixtures of Adoa epimers [*R* (major) and *S* (minor)], whereas the structurally related majusculamide C (**90**) is a single diastereomer with an *S*-Adoa unit. The presence of a minor diastereomer obviously broadens the 2D NMR and 1D TOCSY spectra of **86**, **88**, and **89** considerably, but it has also been suggested that the *R*- and *S*-Adoa diastereomers probably have appreciably different

amounts of *cis* and *trans* conformers around the (Adoa)-(N-methyl-alanine) bond. Also in the case of the *R*-Adoa compounds, the overall shape of the *cis* and *trans* conformers differs significantly. All of these factors likely contribute to the unusual degree of broadness in the NMR spectra. The doubling and broadening of several specific  $^1\text{H}$  and  $^{13}\text{C}$  NMR signals were not observed in the spectra for desmethoxymajusculamide (**87**),

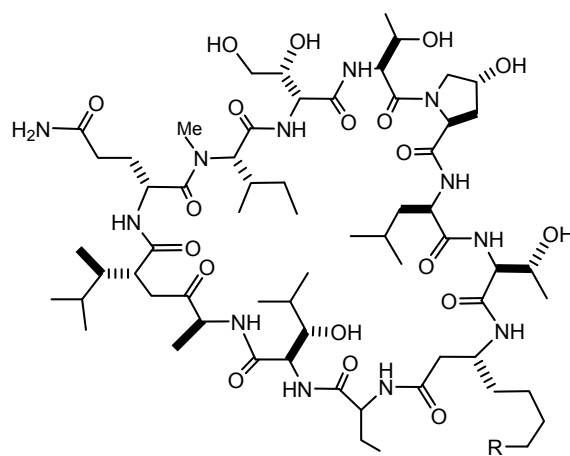


thus suggesting the presence of a single Adoa epimer. The pure compound **86** had  $\text{IC}_{50}$  values of 32 and 400 nM against KB and LoVo cell lines, respectively. When tested in vivo against colon adenocarcinoma #38 or mammary adenocarcinoma #16/C in mice, **86** was poorly tolerated and exhibited only marginal or no antitumor activity. Desmethoxymajusculamide (**87**) and its ring-opened form (generated by breaking the ester bond) demonstrated equivalent efficacy and solid tumor selectivity against four cell lines, including HCT-116, H-460, MDA-MB-435 and Neuro-2A. HCT-116 was the most sensitive cell line, with  $\text{IC}_{50}$  values of 20 and 16 nM, respectively. These findings were

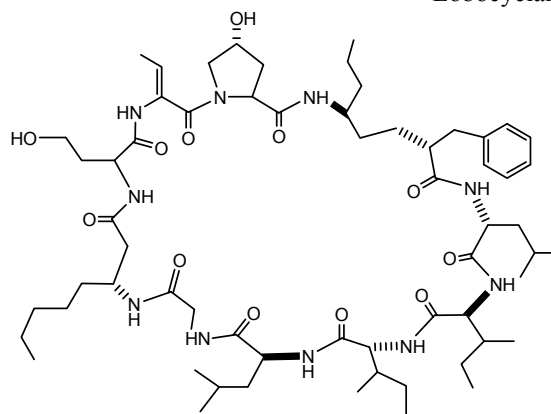
discordant with the common viewpoint that the cyclic form is the bioactive form of the many peptides.



Lobocyclamide A (**91**)



Lobocyclamide B (**92**) R=Et  
Lobocyclamide C (**93**) R=H



Hormothamnin (**94**)

The *in vitro* antifungal activity against *Candida albicans* of methanol extracts from *Lyngbya confervoides* led to the isolation of lobocyclamides A-C (**91-93**) (MacMillan et al., 2002). Lobocyclamide A (**91**) was very similar to the known peptides hormothamnin (**94**) isolated from *Hormothamnion enteromorphoides* (Gerwick et al., 1992) and laxaphycin A, isolated from the terrestrial cyanobacterium *Anabaena laxa* (Frankmölle et

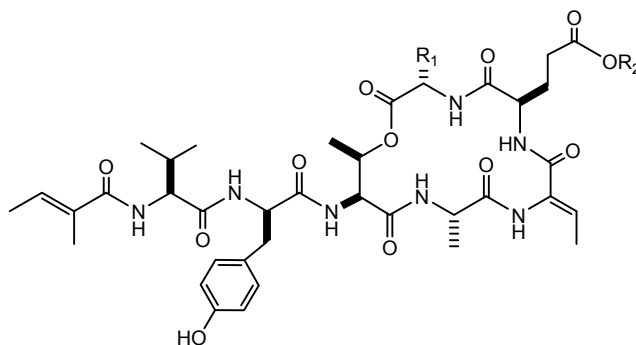


al., 1992a&b; Bonnard et al., 1997). Lobocyclamides A-C (**91-93**) exhibited modest antifungal activity when tested against fluconazole-resistant fungi *C. albicans* 96-489 (150 µg/disk, **91**; 7 mm zone; **92**, 8 mm; **93**, 10 mm) and *C. glabrata* (**92**, 6 mm; **93**, 8 mm, each at 150 µg/disk) in disk diffusion assays on Sabouraud agar plates or microbroth dilution assay (*C. albicans* 96-489 MIC **91**, 100 µg/disk; **92**, 30-100 µg/mL). Mixtures of lobocyclamides **91** and **92** exhibited significant synergism with higher activity (e.g., 1:1 mixture of **91** + **92**, MIC 10-30 µg/mL) than either of the pure compounds alone. A similar phenomenon also noted by Moore and co-workers for the laxaphycins A and B (Frankmölle et al., 1992a).

### 1.1.3 Cyclic polypeptides with branches

Tiglicamides A-C (**95-97**) were discovered from the Florida marine *Lyngbya confervoides*, with their analogues largamides A-C (**98-100**) (Matthew et al., 2009a and 2009b). Tiglicamides A-C (**95-97**) and largamides A-C (**98-100**) differ from each other by one amino acid residue within the cyclic core structure, suggesting an unusually relaxed substrate specificity of the nonribosomal peptide synthetase that is the putative biosynthetic enzyme responsible for the corresponding amino acid incorporation. Tiglicamides A-C (**95-97**) moderately inhibited porcine pancreatic elastase in vitro with IC50 values from 2.14 to 7.28 µM, two to three orders of magnitudes less potent against the same enzyme or other mammalian elastases than lyngbyastatins 4–7 (**108-111**). The activities of two other serine proteases tested (chymotrypsin, trypsin) were not compromised by compounds **95-97** at concentrations up to 50 µM. These results are

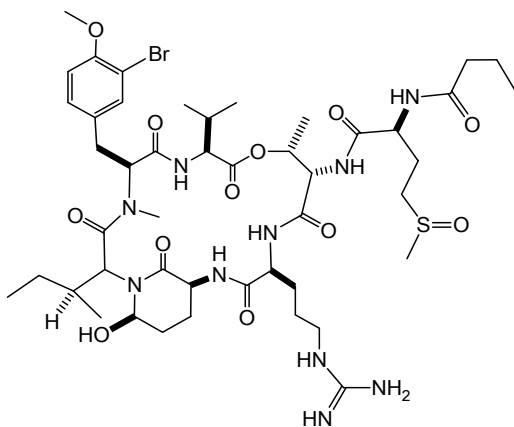
consistent with those previously reported for their analogues, largamides A-C (**98-100**). The corresponding largamide methyl esters **101-103** retained low-micromolar inhibitory activity, indicating that the carboxylic acid residue is not a requisite element for elastaseinhibitory activity.



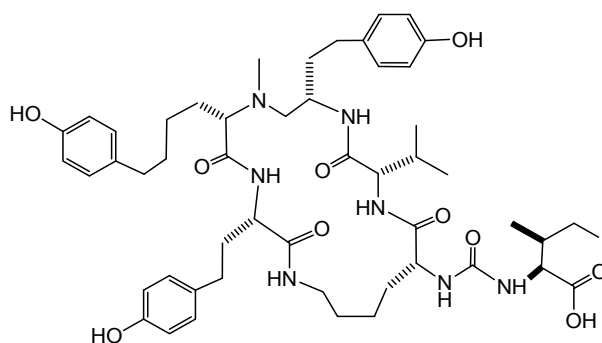
Name	R1	R2	Name	R1	R2
Tiglicamide A ( <b>95</b> )	4-(OH)PhC <sub>2</sub> H <sub>4</sub> -	H	Largamide A methyl ester ( <b>101</b> )	i-Bu	CH <sub>3</sub>
Tiglicamide B ( <b>96</b> )	PhCH <sub>2</sub> -	H			
Tiglicamide C ( <b>97</b> )	CH <sub>3</sub> SO C <sub>2</sub> H <sub>4</sub> -	H	Largamide B methyl ester ( <b>102</b> )	4-(OH)PhC <sub>3</sub> H <sub>5</sub> -	CH <sub>3</sub>
Largamide A ( <b>98</b> )	i-Bu	H			
Largamide B ( <b>99</b> )	4-(OH)PhC <sub>3</sub> H <sub>5</sub> -	H	Largamide C methyl ester ( <b>103</b> )	4-(OH)PhC <sub>4</sub> H <sub>8</sub> -	CH <sub>3</sub>
Largamide C ( <b>100</b> )	4-(OH)PhC <sub>4</sub> H <sub>8</sub> -	H			

The investigation of the marine cyanobacterium *Lyngbya confervoides* collected from the southeastern coast of Florida led to the isolation of 3-amino-6-hydroxy-2-piperidone (Ahp) containing peptolide, pompanopeptin A (**104**), and a novel cyclic pentapeptide, pompanopeptin B (**105**) (Matthew et al., 2008a). Pompanopeptin B (**105**) contains *N*-methyl-2-amino-6-(4'-hydroxyphenyl) hexanoic acid (*N*-methyl-Ahpha),

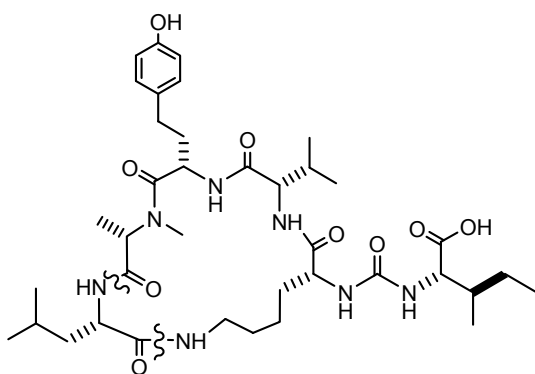
structurally related to carboxypeptidase-A inhibitors anabaenopeptins I (**106**) and J (**107**) isolated from the cyanobacterium *Aphanizomenon flos-aquae* in which *L*-leucine/*L*-phenylalanine and *N*-methyl-*L*-alanine residues were replaced by *L*-Htyr and *N*-methyl-*L*-Alpha, respectively (Murakami et al., 2000). Pompanopeptin A (**104**) selectively



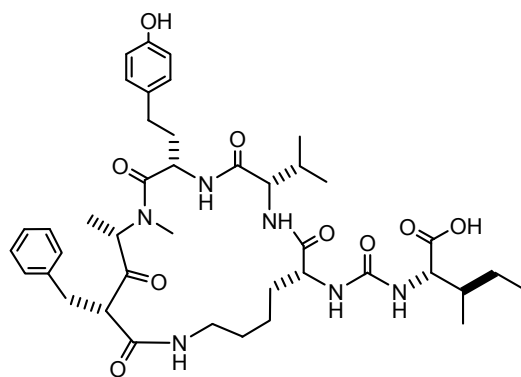
Pompanopeptin A (**104**)



Pompanopeptin B (**105**)



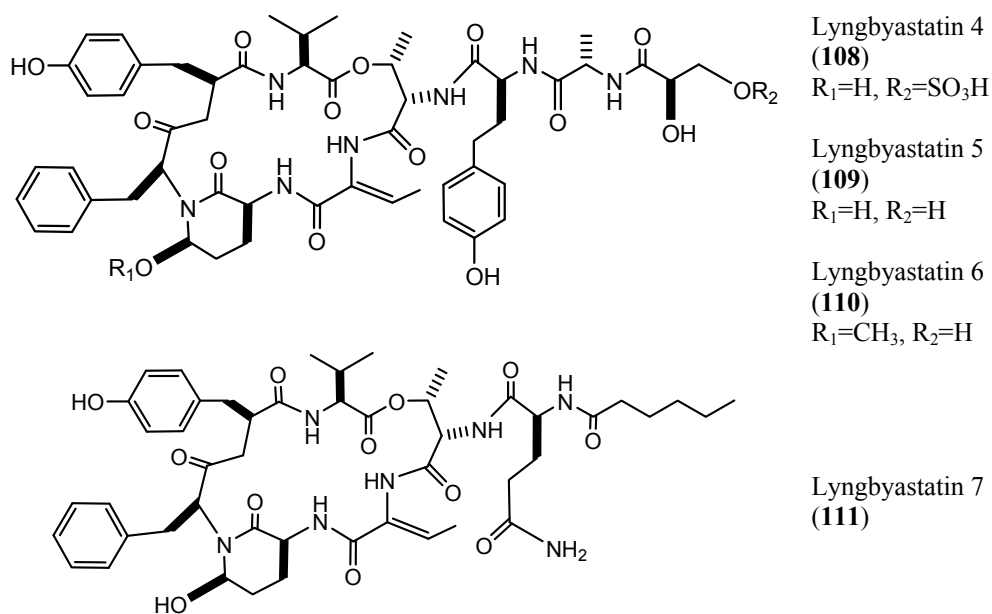
Anabaenopeptin I (**106**)



Anabaenopeptin J (**107**)

inhibited trypsin over elastase and chymotrypsin, with an  $IC_{50}$  value of 2.4 mM; selectivity is conferred by an arginine residue in the cyclic core.

Lyngbyastatins 4-6 (**108-110**), three depsipeptides containing the unusual amino acid homotyrosine and a 3-amino-6-hydroxy-2-piperidone (Ahp) or 3-amino-6-methoxyl-2-piperidone (Amp) residue, were isolated from the marine cyanobacterium *Lyngbya confervoides* off the Florida Atlantic coast (Matthew et al., 2007; Taori et al., 2007). A sample of the marine cyanobacterium *Lyngbya* sp. collected from a mangrove channel at Summerland Key in the Florida Keys produced two lyngbyastatin analogues, termed lyngbyastatin 7 (**111**) and a previously identified cyanobacterial metabolite somamide B (**78**) (Taori et al., 2007; Nogle et al., 2001b). Lyngbyastatin 4-7 (**108-111**) and somamide

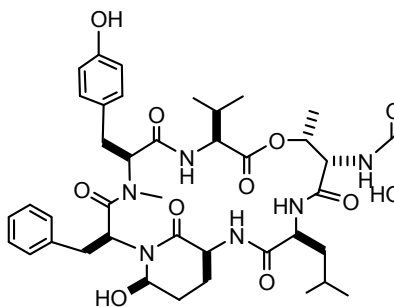


B (**78**) selectively inhibit elastase and chymotrypsin *in vitro* over another serine protease, trypsin with IC<sub>50</sub> values of 3.2~13.9 nM and 2.5~4.3 μM, respectively. Along with lyngbyastatin 4-7, somamide B and scyptolin A, numerous related Ahp-containing cyanobacterial metabolites are assumed to be enzyme substrate mimics to inhibit serine

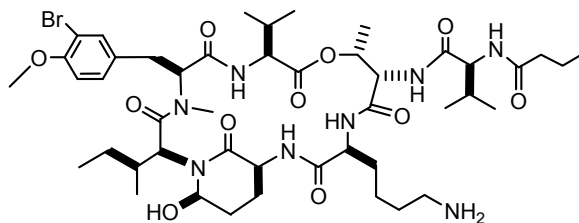
proteases (Matern et al., 2003). Since the Abu moiety appears to contribute strongly to the observed selectivity for elastase (S1 subsite = recognition pocket), the cyclic core structure for **108-111** represents a potent inhibitor prototype. The residues Abu (2-amino-2-butenic acid), Thr, Htyr (homotyrosine) and Ala bind at subsites through S1 to S4 of the enzyme. The side chain of component amino acids in related inhibitors, such as **108-111**, has been postulated to provide additional interaction points for hydrogen bonding with the enzyme. The Thr unit that forms the ester bond to yield the cyclodepsipeptide core occupies the S2 subsite of the protease. The two consecutive residues, Htyr and Ala, located *N*-terminal to this Thr residue are important determinants for efficient elastase-inhibitor complexes based on the co-crystal structures for FR901277 and scyptolin A with the enzyme (S3 and S4 subsites). However, comparable bioassay data for cyclodepsipeptides **108-111** indicate that the corresponding compositional difference in the side chain between **108** and **111** (Htyr-Ala) versus **110** and **111** (Gln-Ha/Ba) is overall less influential on the elastase-inhibitory activity. Remarkably, the fact that the protease-inhibitory activity is retained in the *O*-methylated (Amp) derivative, lyngbyastatin 6 (**110**), demonstrates that the hydroxyl proton in the Ahp unit is not critical for the inhibition of elastase or chymotrypsin.

Two cyclodepsipeptides, kempopeptins A (**112**) and B (**113**), were isolated from a marine cyanobacterium *Lyngbya* sp. from the Florida Keys (Taori et al., 2008). This *Lyngbya* sp. previously afforded the structurally related potent elastase inhibitors lyngbyastatin 7 (**111**) and somamide B (**78**). Many Ahp-containing cyclodepsipeptides isolated from cyanobacteria are known inhibitors of serine proteases such as elastase, chymotrypsin, and trypsin. Kempopeptin A (**112**) inhibited elastase ( $IC_{50}=0.32 \mu\text{M}$ ) with

a slight selectivity over chymotrypsin ( $IC_{50}=2.6 \mu\text{M}$ ). However, kempopeptin B (**113**) inhibited only trypsin activity ( $IC_{50}=8.4 \mu\text{M}$ ). These results are in accordance with previous crystallographic and structure-activity relationship data, suggesting that the amino acid residue between Thr and Ahp binds to the enzyme's specificity pocket and thus plays an important role in determining the selectivity toward serine proteases (Linington et al., 2008). A hydrophobic amino acid at the positions of Thr and Ahp commonly confers preference for chymotrypsin and elastase inhibition (leucine in **112**), while a basic amino acid such as lysine or arginine is necessary for trypsin inhibition (lysine in **113**). The discovery of four different protease inhibitors with three distinct



kempopeptin A (**112**)

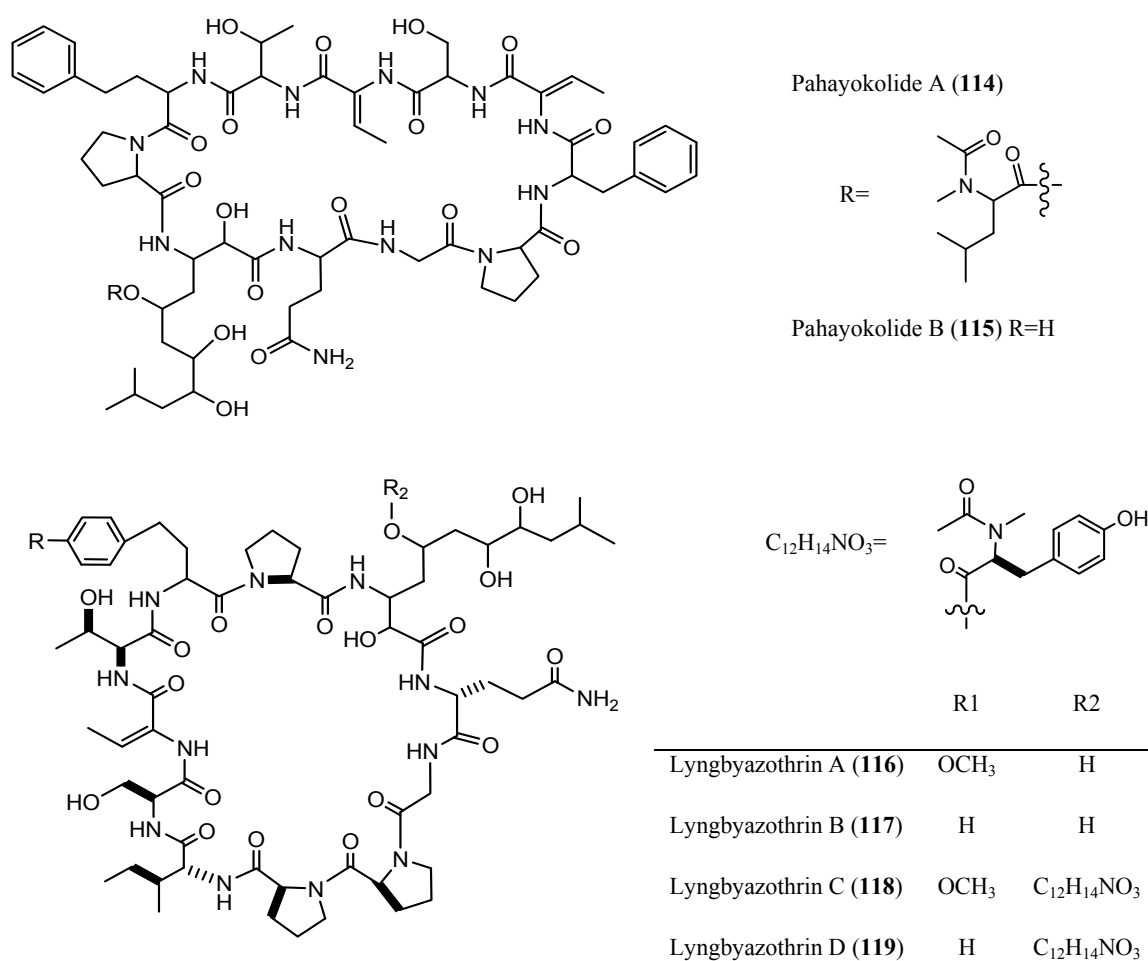


kempopeptin B (**113**)

selectivity profiles from a single homogeneous cyanobacterial collection illustrates the potential of cyanobacteria to execute their own combinatorial biosynthesis and structure optimization. While the role of these protease inhibitors in nature is not fully understood, a defensive function against other microorganisms or consumers is probable.

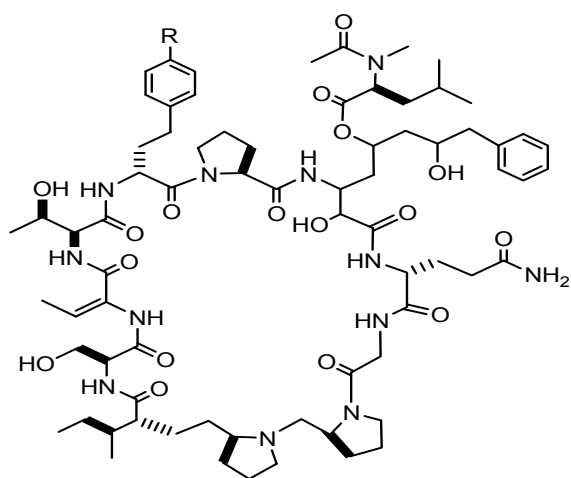
Pahayokolides A-B (**114-115**) and lyngbyazothrins A-D (**116-119**) were produced by a freshwater *Lyngbya* sp. isolated from the Florida Everglades and the cultured

*Lyngbya* sp. 36.91 respectively (An et al., 2007; Zainuddin et al., 2009). Pahayokolides A and B contain the same cyclic undecapeptide core and pahayokolide A contains a pendant *N*-acetyl-*N*-methyl leucine moiety which is absent in pahayokolide B. The pahayokolides are remarkably similar in structure to schizotrin A (**122**), cyclic peptides isolated from *Schizotrix* sp. (Pergament et al., 1994). *Z*-dehydrobutyrate, threonine, homophenylalanine and *N*-acetyl-*N*-methyl leucine of pahayokolide A replace the valine, serine,



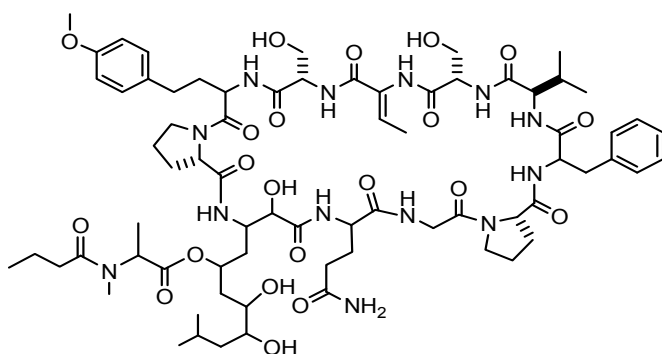
homotyrosine methyl ether and *N*-butyroyl-*N*-methyl alanine of schizotrin A, respectively. The lyngbyazothrins are analogous in structure with tychonamides A (**120**)

and B (**121**) isolated from *Tchyonema* sp. (Mehner et al., 2008). All the pahayokolides, lyngbyazothrins and schizotrin A have an unusual  $\beta$ -amino acid, 3-amino-2,5,7,8-tetrahydroxy-10-methylundecanoic acid (Athmu). To the best of our knowledge, this amino acid has not been observed elsewhere. However, a similar  $\beta$ -amino acid, 3-amino-2,5,7,-trihydroxy-8-phenyloctanoic acid, has been recently identified in tychonamides A and B. Like pahayokolide A, tychonamides A and B contain a pendant *N*-acetyl-*N*-methyl



Tychonamides A (**120**) R=OMe

Tychonamides B (**121**) R=H



Schizotrin A (**122**)

leucine connected via an ester linkage to the 5-hydroxy group of the  $\beta$ -amino acid. Pahayokolide A inhibits a number of cancer cell lines over a range of concentrations (IC<sub>50</sub> varied from 2.13 to 44.57  $\mu$ M). Pahayokolide A was acutely toxic to zebrafish



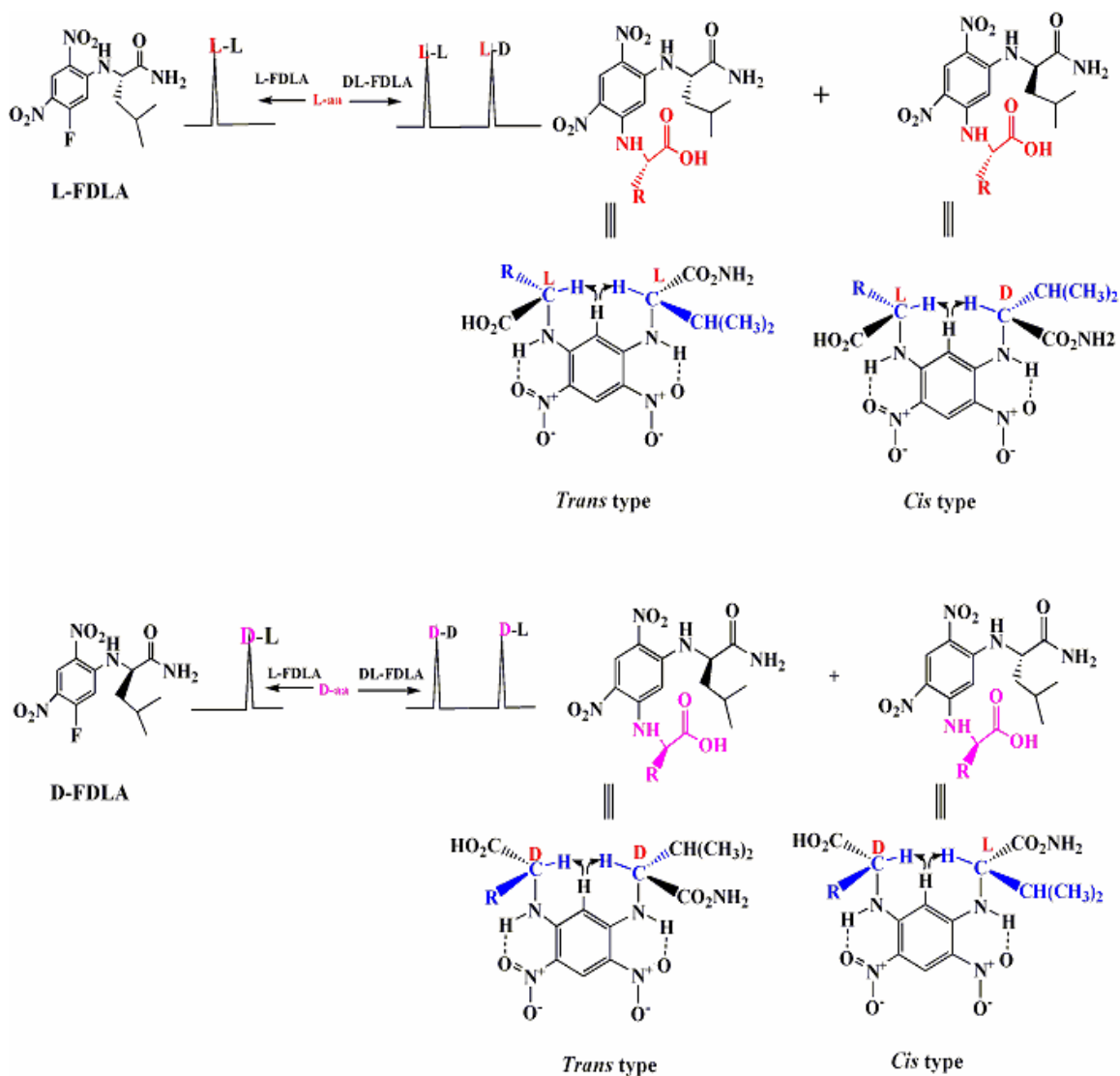
embryos ( $LC_{50}=2.15 \mu\text{M}$ ), however, only marginally toxic against brine shrimp at the highest concentrations tested (1 mg/mL) (Berry et al., 2004). The mixture of lyngbyazothrins A (**116**) and B (**117**) showed only low antimicrobial activity against *Micrococcus flavus*, whereas the mixture of lyngbyazothrins C (**118**) and D (**119**) was active against *Bacillus subtilis*, *Escherichia coli*, *Pseudomonas aeruginosa*, and *Serratia marcescens*. It seems that the acyl residue at C-5 of the Aound unit plays an important role in antimicrobial activity. This assumption was supported by the activity and structure of pahayokolide A, where the 5-hydroxy group is also substitute by *N*-acetyl-*N*-methyl-leucine.

Various polypeptides isolated from *Lyngbya* species have been reviewed and their bioactivities discussed. It is common that cyanobacteria, including *Lyngbya*, often produce more than one member of a certain structural class. This phenomenon is partially a result of relaxed specificity of biosynthetic enzymes (Magarvey et al., 2006). Particularly, not only polyketide synthases (PKS) of prokaryotes can show considerable substrate tolerance (Watts et al., 2007), but also adenylation domains of non-ribosomal peptide synthetase (NRPS) may accept different amino acids within a group of polar or nonpolar amino acids (Challis et al., 2000). Natural structural diversification may also be the result of slight genetic variation or a slightly changed environment.

#### **1.1.4 Absolute configuration of peptides**

To resolve the structure of polypeptides, the natural products are characterized with an amino acid analysis, mass spectrometry and NMR. Full assignment of the  $^1\text{H}$ - and  $^{13}\text{C}$ -

NMR chemical shifts is achieved on the basis of 2D-NMR experiments including TOCSY (total correlated spectroscopy) and HMBC (heteronuclear multiple bond correlation experiment). The 2D TOCSY spectrum identifies the chemical shifts of pairs of protons that are within the same amino acid and unknown amino acids. The sequence of the component is deduced from HMBC. The *J*-based configuration analysis is proven to be practical for natural products with complicated structures (Matsumori et al., 1999).



### Figure 1 - 1. The reaction of D, L-FDLA with amino acids

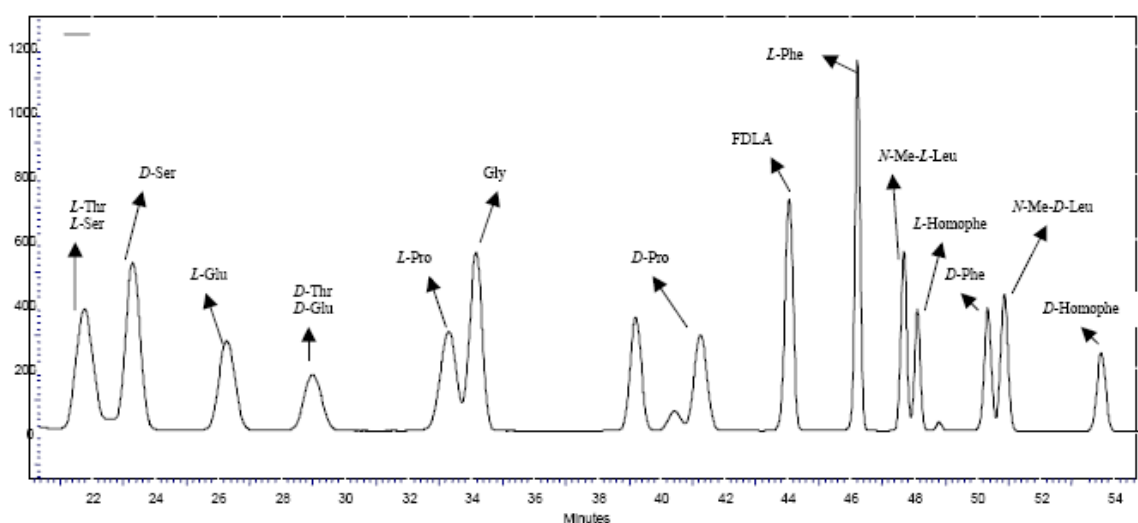
The absolute configuration of the secondary alcohol is established by Marfey's method or advanced Marfey's method. It is difficult to apply Marfey's method to a peptide containing non-proteinogenic amino acids when standards are not available. The advantage of advanced Marfey's method over Marfey's method is that the former does not require authentic *D* and *L* standards.

The basis of the advanced Marfey's method is shown in Fig. 1-1. The *L*-FDLA (1-fluoro-2,4-dinitrophenyl-5-*L*-leucinamide) derivatives of *D*-amino acids (*D-L* type; the first letter refers to amino acids and the second letter refer to FDLA) are diastereomers of the *L*-FDLA derivatives of *L*-amino acids (*L-L* type). Generally the diastereomers have different retention times during chromatography. In this method, the resolution of *D*- and *L*-amino acid derivatives with FDLA results from the difference in the hydrophobicity of the diastereomeric derivatives. The hydrophobic difference is a result of the *relative* arrangement of the two hydrophobic substituents at both C $\alpha$  of the test amino acids (except Gly) and *L*-leucine of the *L*-FDLA. The FDLA derivative of the *cis* (*Z*)-type arrangement interacts more strongly with ODS silica gel and has a longer retention time than that of the *trans* (*E*)-type arrangement. Therefore, the *trans* (*E*)-derivatives (*L-L* and *D-D*) are usually eluted first from the column. The relationships between *L-D* types and *D-L* types, and between *D-D* types and *L-L* types are enantiomeric. Each pair of enantiomers shows the same retention times. The retention times of the aa-FDLA derivatives on C-18 chromatogram is:  $L-L = D-D < L-D = D-L$  (Kiyonaga, 1997).

## 1.2 Results and Discussion

### 1.2.1 Absolute configuration of pahayokolides A and B

Pahayokolides A (**114**) and B (**115**) are two branched cyclic peptides produced by the fresh water cyanobacterium *Lyngbya* sp. strain 15-2 from the Florida Everglades. The absolute configuration of the component amino acids in pahayokolides A and B were



**Figure 1 - 2.** HPLC/PDA chromatogram of standard amino acid mixture derivatized with *D,L*-FDLA.

Assignments are made for derivatives with *L*-FDLA.

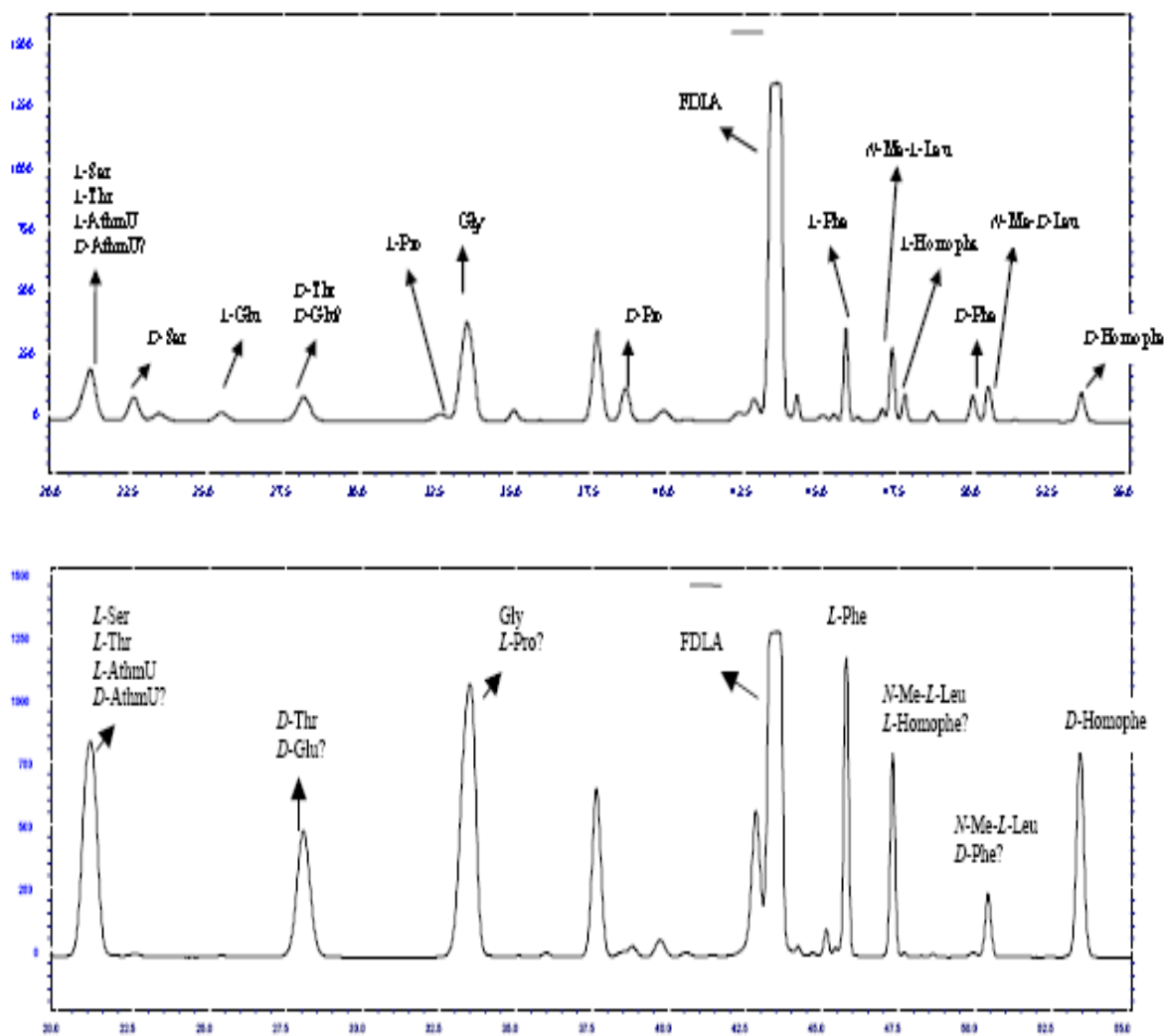
determined using the advanced Marfey's method. The resulting amino acid derivatives are separated by the  $C_{18}$  reverse-phase HPLC, and detected by PDA UV at 340 nm and mass spectrometry. In order to optimize the HPLC condition, authentic *D* and *L* amino

acid standards were derivatized with *L*-FDLA and analyzed on HPLC/PDA, with the exception of Athmu and *N*-Me-*D*-leucine which were not commercially available (Table 1-1, Fig. 1-2). In the case of *N*-methyl-leucine only the *L* isomer was commercially available. According to the Marfey's method, the derivative of *N*-Me-*D*-leucine with *L*-FDLA may be replaced with its enantiomer, the derivative of *N*-Me-*L*-Leu with *D*-FDLA, because both of them had the same retention time on the C<sub>18</sub> HPLC column.

**Table 1 - 1. Availability of amino acids standards and reactions with FDLA**

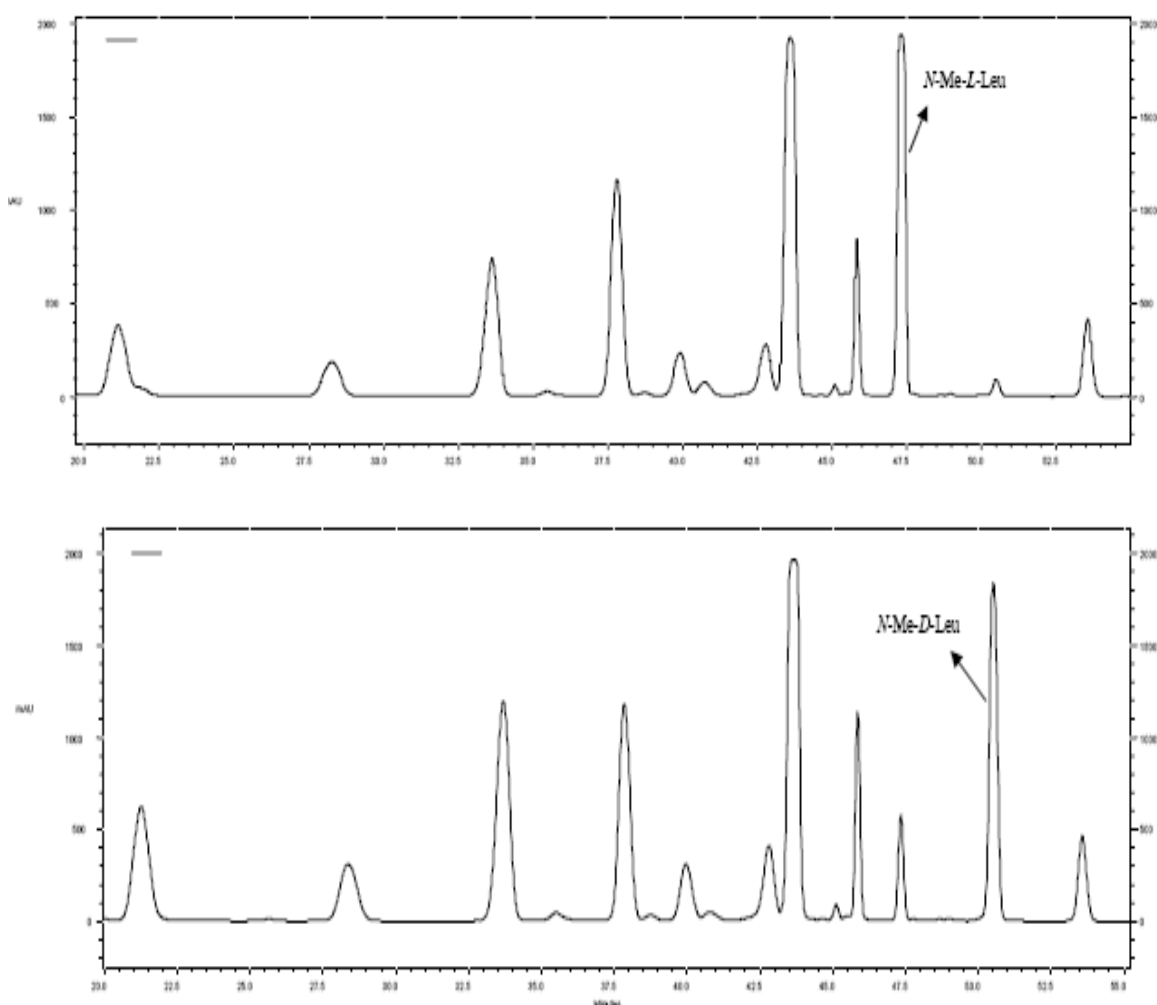
Standards	Availability	FDLA	
<i>L</i> -Ser	Yes	<i>L</i> -FDLA	/
<i>D</i> -Ser	Yes	<i>L</i> -FDLA	/
<i>L</i> -Thr	Yes	<i>L</i> -FDLA	/
<i>D</i> -Thr	Yes	<i>L</i> -FDLA	/
<i>L</i> -Athmu	No	/	/
<i>D</i> -Athmu	No	/	/
<i>L</i> -Glu	Yes	<i>L</i> -FDLA	/
<i>D</i> -Glu	Yes	<i>L</i> -FDLA	/
<i>L</i> -Pro	Yes	<i>L</i> -FDLA	/
<i>D</i> -Pro	Yes	<i>L</i> -FDLA	/
<i>L</i> -Phe	Yes	<i>L</i> -FDLA	/
<i>D</i> -Phe	Yes	<i>L</i> -FDLA	/
<i>L</i> -Homophe	Yes	<i>L</i> -FDLA	/
<i>D</i> -Homophe	Yes	<i>L</i> -FDLA	/
<i>N</i> -Me- <i>L</i> -Leu	Yes	<i>L</i> -FDLA	<i>D</i> -FDLA
<i>N</i> -Me- <i>D</i> -Leu	No	/	/
Gly	Yes	<i>L</i> -FDLA	/

As expected, the Marfey analysis supported the presence only of the amino acids indicated in the structures of pahayokolides A (**114**) and B (**115**) (Fig. 1-3). The absolute configuration was assigned as follows using PDA detector at 340 nm: *L*-serine, *D*-glutamine, and *L*-proline (2), respectively. In the hydrolysis process, free glutamine in the hydrolysate was converted into free glutamic acid by hydrolysis of the amide linkage.



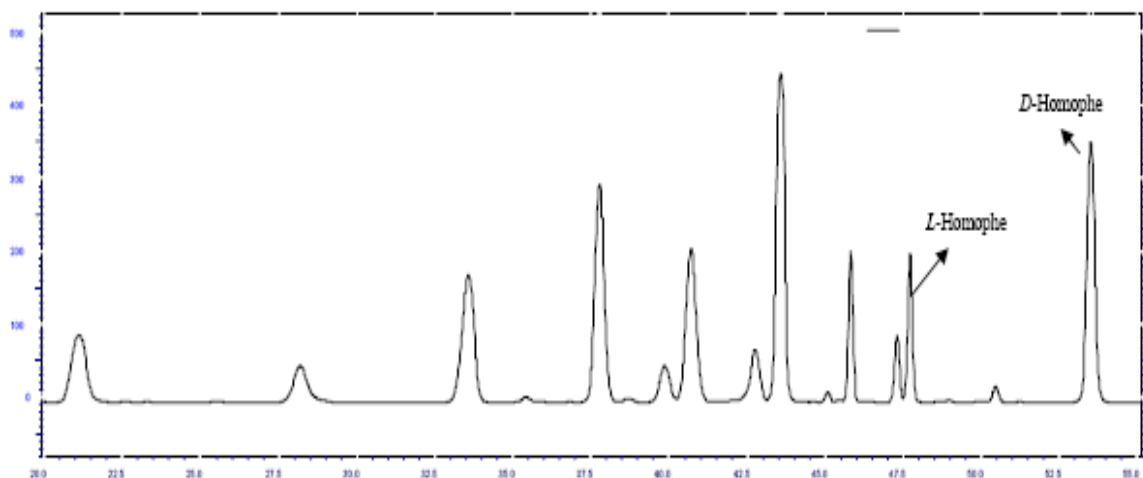
**Figure 1 - 3.** HPLC/PDA analysis of pahayokolide A hydrolysates derivatized with *D,L*-FDLA (top) and *L*-FDLA (bottom).

Dhb (dehydrobutyrate) was not stable as a free amino acid; therefore, it could not be derivatized with *L*-FDLA standard and remained undetected in the amino acid analysis (Dahlmann, et al., 2003). Comparison of the diastereomeric derivatives of *N*-methyl-*L*-leucine prepared with both *L*-FDLA and *D*-FDLA, with the *L*-FDLA derivatives of the acid hydrolysates of pahayokolides A (**114**) and B (**115**), indicated that pahayokolide B (**114**) differed from pahayokolide A (**115**) in possessing an *N*-methyl-leucine.



**Figure 1 - 4.** HPLC/PDA chromatogram of pahayokolide A hydrolysate derivatized with *L*-FDLA and spiked with *N*-Me-*L*-Leucine derivatized with *L*-FDLA (top) and *D*-FDLA (bottom).

The separate experiments, in which the diastereomeric derivatives of *N*-methyl-*L*-leucine with *L*-FDLA and *D*-FDLA were separately spiked into the pahayokolide A hydrolysate, indicated that the presence of both *N*-methyl-*L*-leucine and its *D* isomer in pahayokolide A (Fig. 1-4). However, the *D/L* ratio was lower under milder hydrolysis conditions. When pahayokolide A was hydrolyzed in 6M HCl, the *L: D* ratio of *N*-methyl-leucine was 2.75:1. Hydrolysis of pahayokolide A in 4M HCl, resulted in a *L: D* ratio of 5.5:1. This possibly resulted from the acid sensitivity of the *N*-methyl-leucine residue. It was concluded that the *D* isomer arose from partial racemization of *N*-methyl-*L*-leucine. A similar observation was made for the tychonamides (Mehner et al., 2008).

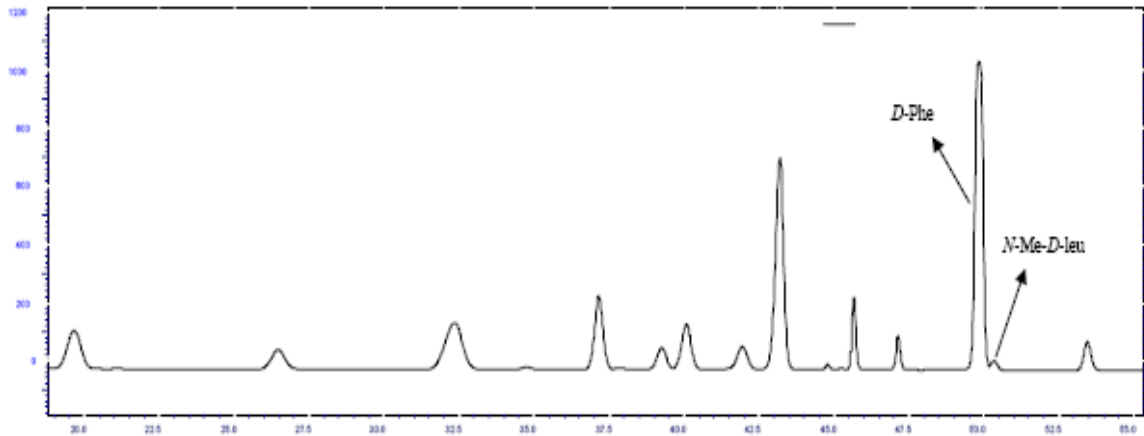


**Figure 1 - 5. HPLC/PDA chromatogram of pahayokolide A hydrolysate derivatized with *L*-FDLA and spiked with *D,L*-Homophe derivatized with *L*-FDLA.**

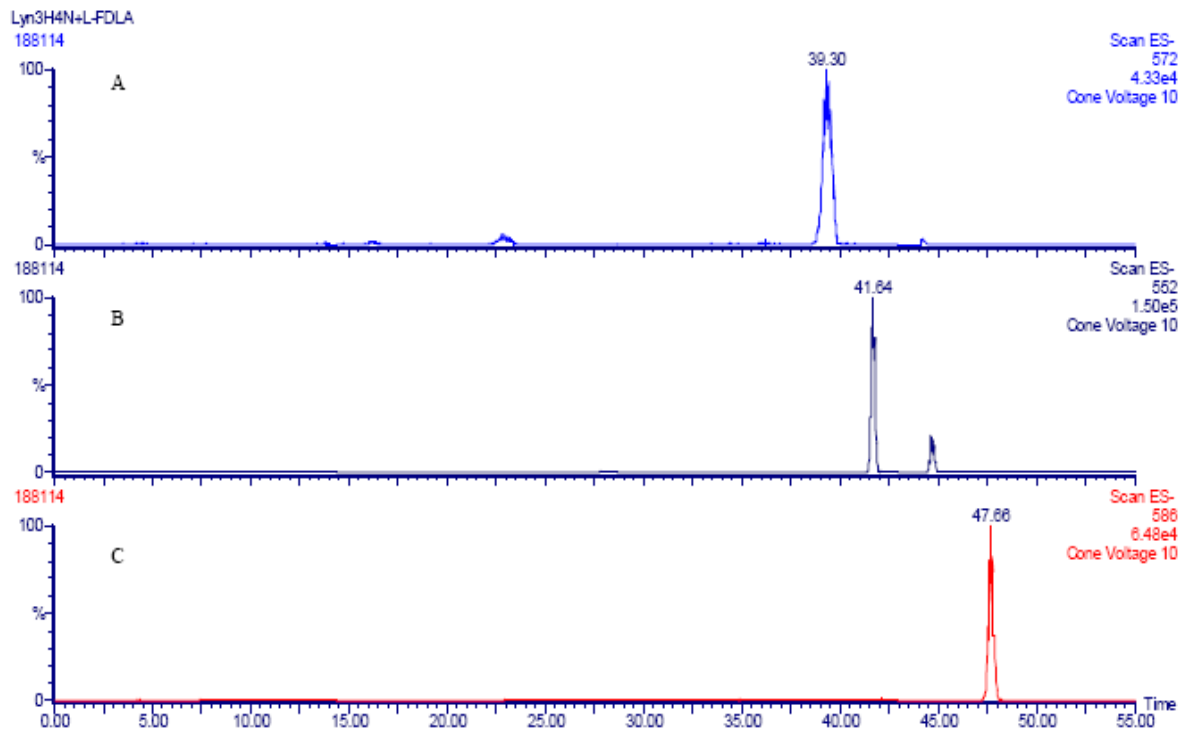
The retention times for the *L*-FDLA derivatives of *N*-methyl-*L*-leucine and *L*-homophenylalanine were close at 47.3 and 47.7 min., respectively. In separate experiments, the *L*-FDLA derivatized pahayokolide A hydrolysate was spiked with the *L*-



FDLA derivatives of *N*-methyl-*L*-leucine or *L*-homophenylalanine, confirming the presence of *N*-methyl-*L*-leucine and the absence of *L*-homophenylalanine (Fig. 1-5).

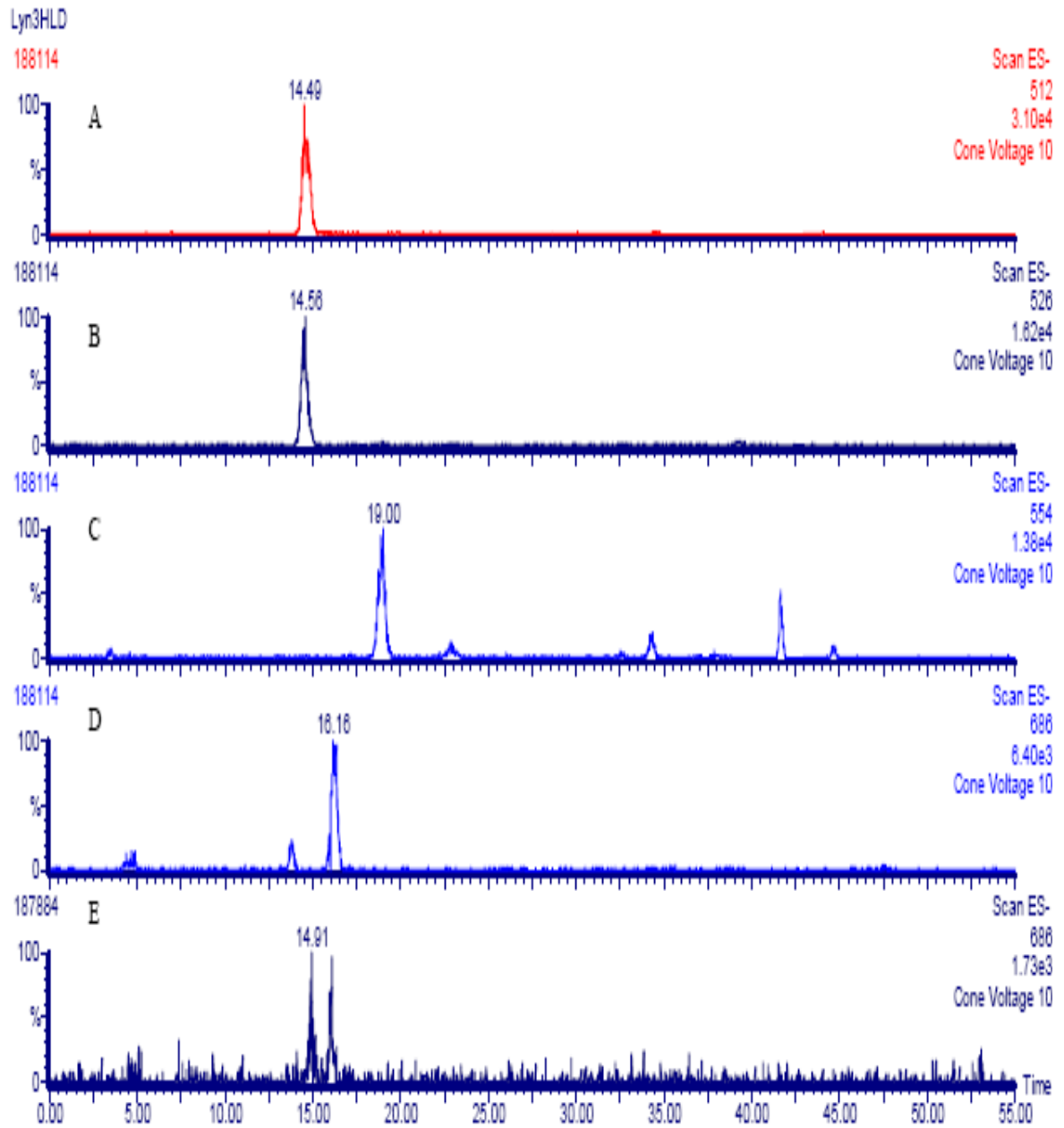


**Figure 1 - 6.** HPLC/PDA chromatogram of pahayokolide A hydrolysate derivatized with *L*-FDLA and spiked with *D*-Phe derivatized with *L*-FDLA.



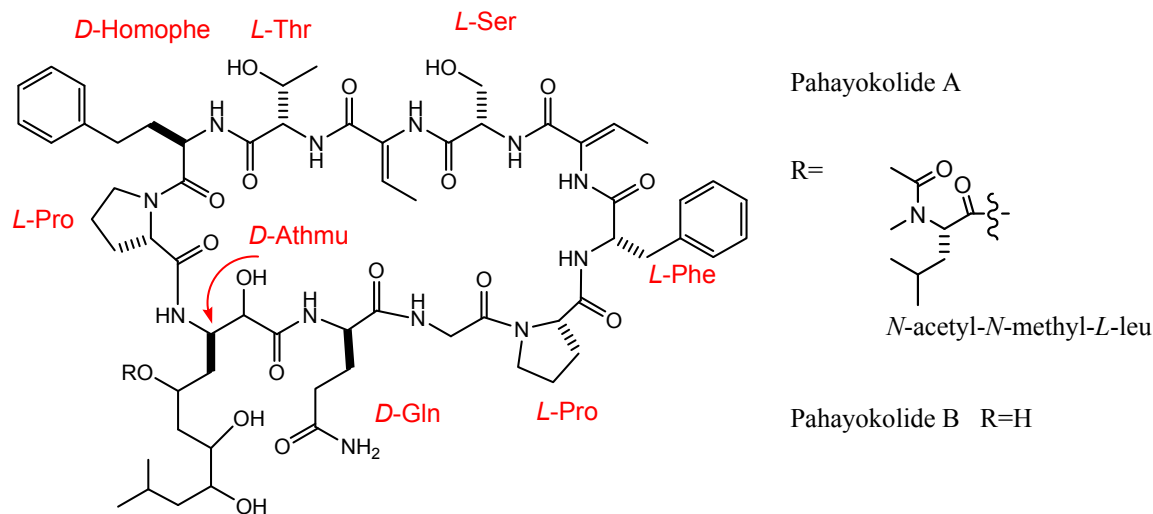
**Figure 1 - 7. Mass chromatograms of the *L*-FDLA derivatives of pahayokolide A hydrolysate monitored at the *m/z* values of the M-1 ions (A, *m/z* 572, *L*-Phe; B, *m/z* 552, *N*-Me-*D,L*-Leu; C, *m/z* 586, *D*-Homophe) ions using ESI LC/MS in negative ion mode**

Similarly, the retention times for the *L*-FDLA derivatives of *D*-phenylalanine and *N*-methyl-*D*-leucine were very close at 50.0 and 50.2 min. respectively. In separate



**Figure 1 - 8.** Mass chromatograms of the *L*-FDLA (A-D) and *D,L*-FDLA (E) derivatives of pahayokolide A hydrolysate monitored at the *m/z* values of the M-1 ions (A, *m/z* 512, *L*-Ser; B, *m/z* 526, *L*-Thr; C, *m/z* 554, *D*-Glu; D, *m/z* 686, *D*-AthmU and E, *m/z* 686, *D*-AthmU) using ESI LC/MS in negative ion mode.

experiments, the *L*-FDLA derivatized pahayokolide A hydrolysate was spiked with the *L*-FDLA-*D*-phenylalanine derivative and the *D*-FDLA-*N*-methyl-*L*-leucine derivative, confirming the presence of *N*-methyl-*D*-leucine and the absence of *D*-phenylalanine (Fig. 1-6). Marfey's analysis using MS confirmed the assignments: *L*-phenylalanine, *N*-methyl-*L*-leucine and *D*-homophenylalanine (Fig. 1-7).



**Figure 1 - 9.** The absolute configuration of pahayokolides A and B

The absolute configuration at C $\beta$  of Athmu was tentatively assigned as *R* (or *D*) by Marfey's analysis with ESI-MS (electrospray ionization mass spectrometry) (Fig. 1-8 D, E). The assignment was based on relative retention times of the diastereomeric Athmu derivatives, assuming that *L*-FDLA-*L*-Athmu elutes before *L*-FDLA-*D*-Athmu. The absolute configuration at C $\beta$  of two similar  $\alpha$ -hydroxy- $\beta$ -amino acids, 3-amino-2,5,9-

trohydroxy-10-phenyldecanoic acid (Ahda) (Helms et al., 1988) and 3-amino-2,5-dihydroxy-8-phenyloctanoic acid (Ahoa) (Fujii et al., 1999), was assigned by this way.

The FDLA derivative of Athmu was observed only when the hydrolysis was performed

**Table 1 - 2. LC/MS analysis of pahayokolides A-D hydrolysates derivatized with *D,L*-FDLA and *L*-FDLA**

	Pahayokolide A hydrolysates + <i>D,L</i> -FDLA		Pahayokolide A hydrolysates + <i>L</i> -FDLA		Pahayokolide B hydrolysates + <i>L</i> -FDLA		Pahayokolides C&D hydrolysates + <i>L</i> -FDLA	
	[M+TFA-H] <sup>-</sup>	T <sub>R</sub> (min)	[M+TFA-H] <sup>-</sup>	T <sub>R</sub> (min)	[M+TFA-H] <sup>-</sup>	T <sub>R</sub> (min)	[M+TFA-H] <sup>-</sup>	T <sub>R</sub> (min)
<i>L</i> -Athmu	686	14.9	/	/	/	/	/	/
<i>D</i> -Athmu		16.0	686	16.2	686	17.9	686	17.6
<i>L</i> -Ser	512	15.4	512	14.5	512	15.2	512	15.5
<i>D</i> -Ser		16.3	/	/	/	/	/	/
<i>L</i> -Thr	526	15.4	526	14.6	526	16.0	526	15.4
<i>D</i> -Thr		22.7	/	/	/	/	/	/
<i>L</i> -Glu	554	18.0	/	/	/	/	/	/
<i>D</i> -Glu		19.7	554	19.0	554	21.0	554	20.0
<i>L</i> -Pro	522	24.1	522	24.6	522	24.8	522	24.4
<i>D</i> -Pro		31.9	/	/	/	/	/	/
<i>L</i> -Phe	572	39.4	572	39.3	572	40.8	572	40.2
<i>D</i> -Phe		45.8	/	/	/	/	/	/
<i>N</i> -Me- <i>L</i> -Leu	552	42.0	552	41.6	552	42.5	552	42.0
<i>N</i> -Me- <i>D</i> -Leu		45.6	552	45.0	552	46.0	552	45.5
<i>L</i> -Homophe	586	42.6	/	/	/	/	/	/
<i>D</i> -Homophe		49.4	586	47.7	586	49.0	586	48.2
Gly	482	23.7	482	24.2	482	24.4	482	24.0
FDLA	427	35.8	427	35.7	427	35.9	427	35.5

M = amino acid + FDLA.

in 4M HCl. When the hydrolysis was carried out in 6M HCl a derivative having a

molecular ion at  $m/z$  668 was observed, corresponding to a loss of water from the FDLA-Athmu derivative. Either one of the alcohols is dehydrated to an alkene or the lactone is formed under these conditions. Marfey's analysis using ESI-MS confirmed *L*-serine, *L*-threonine, *D*-glutamine (Fig. 1-8 A-C), and was consistent with the other results detected using PDA detector (Table 1-2). Except for the absence of *N*-methyl-*L*-leucine, the component amino acids in pahayokolide B have the same absolute configuration as pahayokolide A (Table 1-2, Fig. 1-9).

### 1.2.2 Pahayokolides C and D is conformers of pahayokolide A

Two more cyclic peptides, pahayokolides C and D, were isolated with pahayokolides A and B from *Lyngbya* sp. strain 15-2 (Fig. 1-10). High resolution mass spectroscopy suggested pahayokolides A, C and D had the same  $m/z$  (HRESIFTMS  $m/z$   $[M + Na]^+ 1494.748$ ). Marfey's analysis indicated that pahayokolides C and D contained

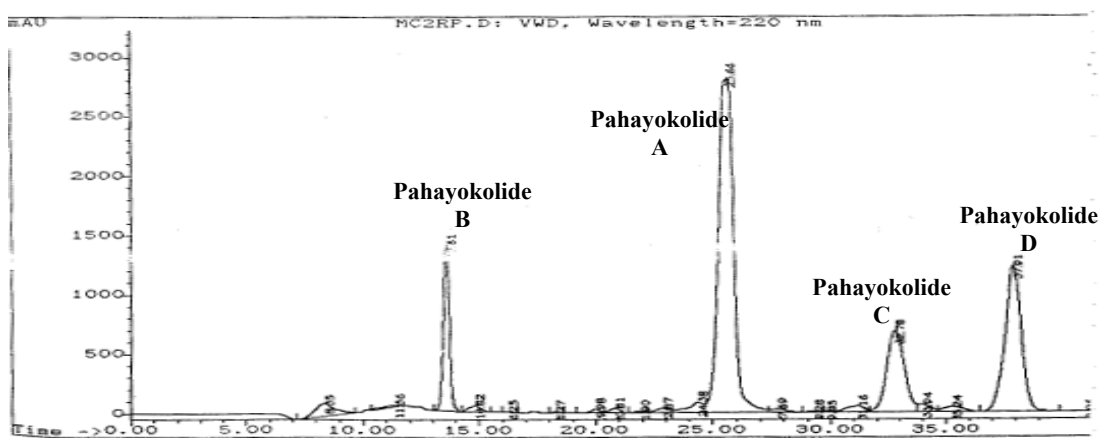
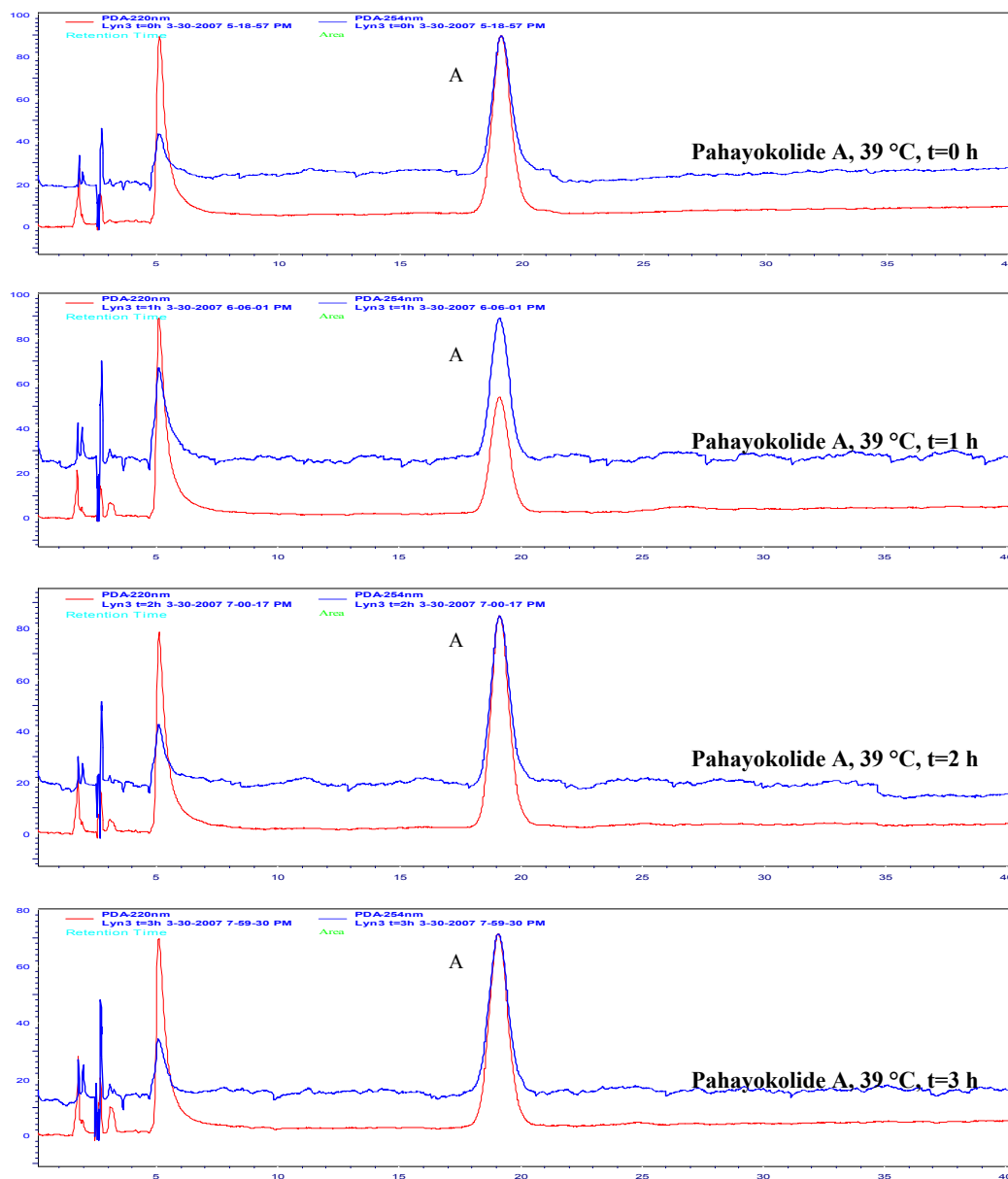
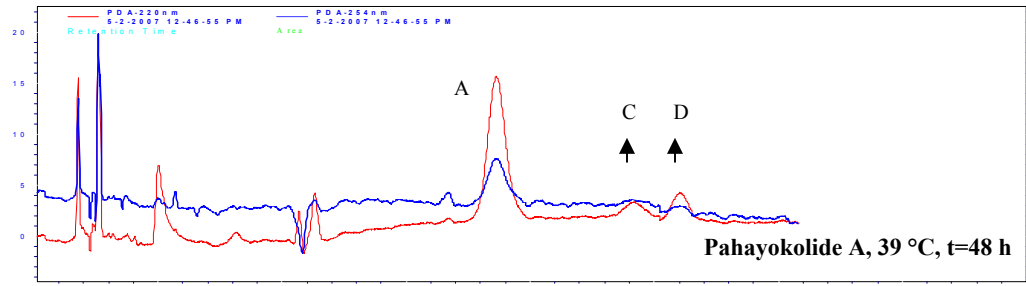


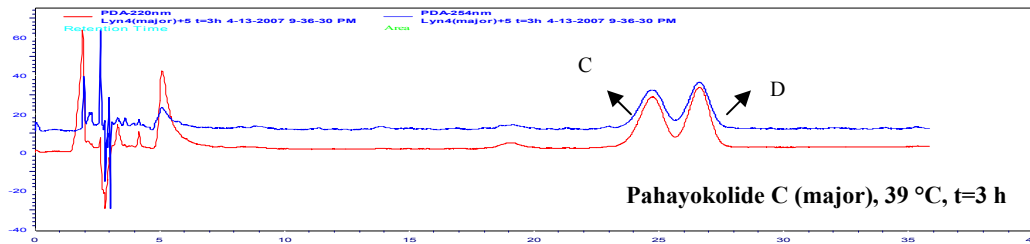
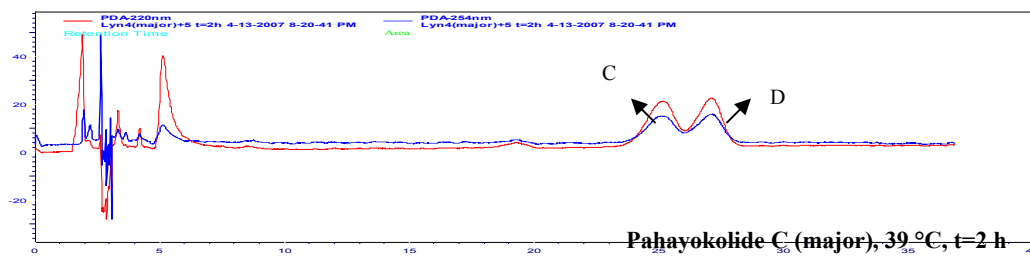
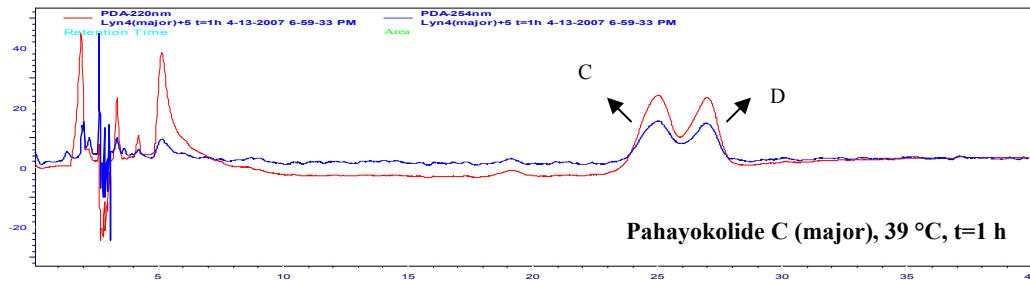
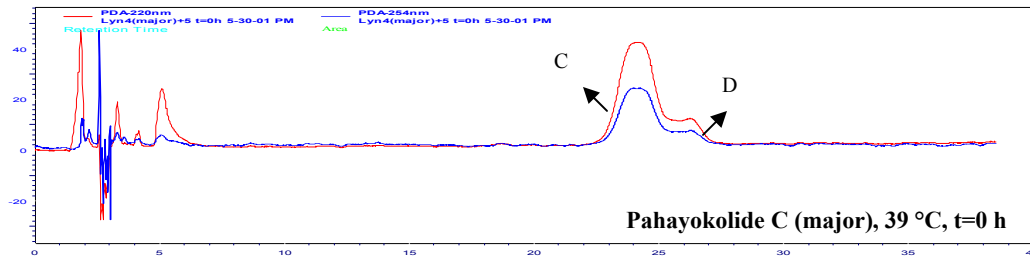
Figure 1 - 10. HPLC/PDA chromatogram of pahayokolides A-D

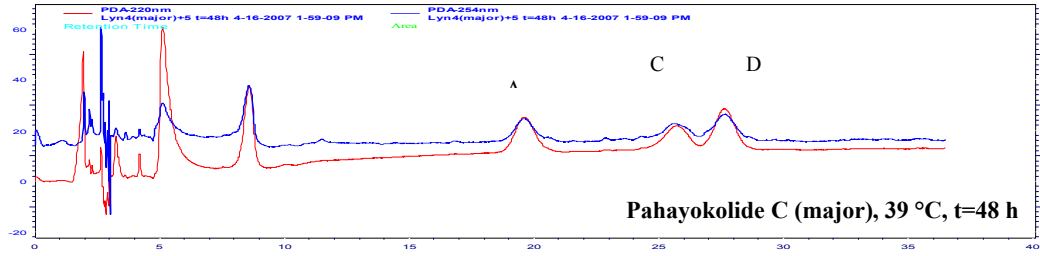
the same amino acids present in pahayokolides A (Table 1-2). Signal doubling and broadening in the  $^1\text{H}$  and  $^{13}\text{C}$  NMR spectra of pahayokolides A indicated the presence of isomeric species or slowly interconverting of conformers in solution. Upon heating pure



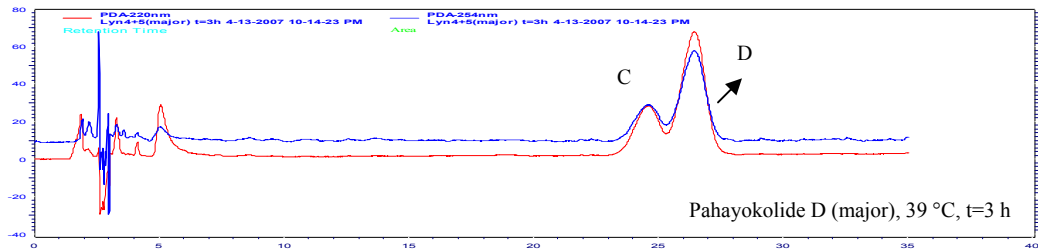
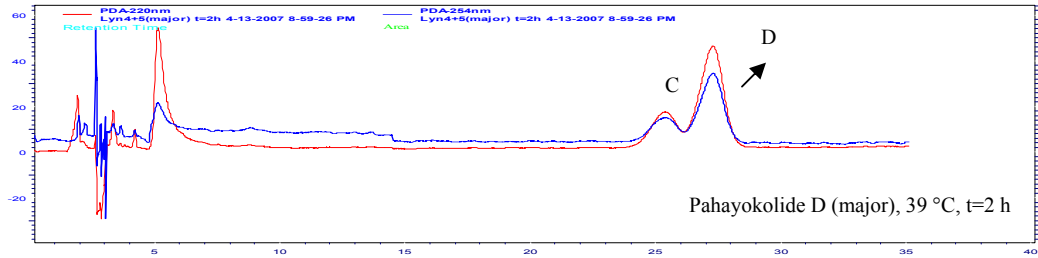
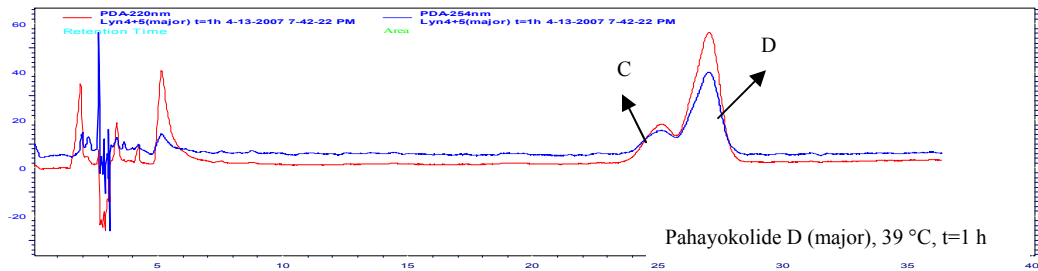
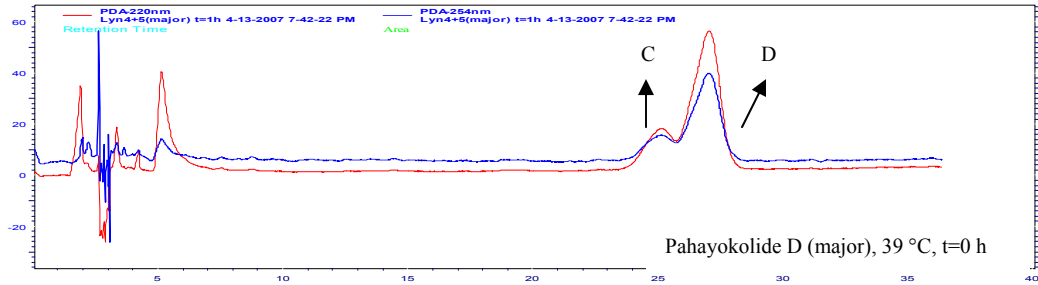


**Figure 1 - 11. HPLC/PDA chromatogram of pahayokolide A heated at 39 °C and monitored at 220 nm (red) and 254 nm (blue) at 0, 1, 2, 3, and 48 hours.**

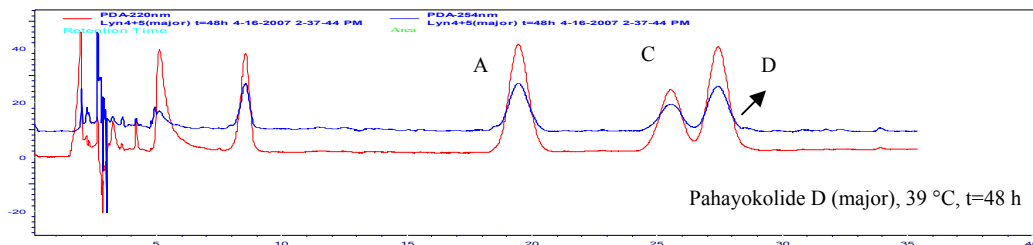




**Figure 1 - 12.** HPLC/PDA chromatogram of a complex with pahayokolide C as the major component heated at 39 °C and monitored at 220 nm (red) and 254 nm (blue) at 0, 1, 2, 3, and 48 hours.







**Figure 1 - 13. HPLC/PDA chromatogram of a complex with pahayokolide D as the major component heated at 39 °C and monitored at 220 nm (red) and 254 nm (blue) at 0, 1, 2, 3, and 48 hours.**

pahayokolides A at 39°C, equilibrium between pahayokolides A, C and D was reached after 48 hours (Fig. 1-11). The same equilibrium mixture of pahayokolides A, C and D was observed when a mixture of pahayokolides C and D was heated at 39°C for 48 hours (Fig. 1-12 and 1-13). All these observations allowed us to conclude that pahayokolides C and D were two slowly interconverting conformers of pahayokolide A.

### 1.3 Materials and Methods

#### 1.3.1 Culture conditions of *Lyngbya* sp. strain 15-2

*Lyngbya* sp. strain 15-2 collected from the Florida Everglades was cultured in BG11 medium supplemented with 0.19 M Na<sub>2</sub>CO<sub>3</sub> buffered with 2-morpholinoethanesulfonic acid (MES) at pH 7.2 at 20~22 °C under continuous white light for 40 days.

#### 1.3.2 Purification of pahayokolides A-D

Samples of biomass were harvested from cultures of *Lyngbya* sp. strain 15-2 and lyophilized. The freeze-dried biomass was pulverized in liquid nitrogen, and extracted in 70% MeOH (in water) at room temperature for two days. Following solvent removal, water-soluble constituents were separated by RP<sub>18</sub> solid-phase extraction using a stepwise

gradient solvent system from 20% MeOH in water to 100% MeOH (Berry et al., 2004). The bioactive 80% and 100% MeOH fractions were then separately chromatographed on preparative HPLC (Alltech Alltima C<sub>18</sub>, 10 μm, 250 × 22 mm) using 75% MeOH in H<sub>2</sub>O and detected at 220 nm. Pahayokolides A-D were obtained. The purity of pahayokolides was checked on analytical HPLC (Alltech Apollo C<sub>18</sub>, 5 μm, 250 mm × 4.6 mm) using 70% MeOH in H<sub>2</sub>O and detected at 220 and 254 nm.

### 1.3.3 Preparation of FDLA derivatives

To 40 μL of 2 μM each standard amino acid solution was added 20 μL of 1 M sodium bicarbonate, and 80 μL of 1% *L*- or *D*-FDLA in acetone. These solutions were vortexed and incubated at 40 °C for 60 min. These reactions were quenched by the addition of 10 μL of 2 M HCl and stored at 4 °C. According to the Marfey's method, the derivative of *N*-methyl-*D*-leucine with *L*-FDLA could be substituted with the derivative of *N*-methyl-*L*-leucine with *D*-FDLA, as the enantiomers will have the same retention time on C<sub>18</sub> HPLC column.

A 100 μg of pahayokolide A or B or a mixture of C and D was hydrolyzed at 110 °C for 14 h with 500 μL of 4M HCl. This solution was divided into two portions and dried with N<sub>2</sub> gas. Each portion was derivatized with *L*- or *D*-FDLA at 40 °C for 60 min. These reactions were quenched by the addition of 10 μL of 2 M HCl and stored at 4 °C.

### 1.3.4 HPLC/PDA conditions

The separation of the *L*- and *D,L*-FDLA derivatives of pahayokolide A was performed on analytical HPLC (Alltech Apollo C<sub>18</sub>, 5 $\mu$ m , 250 mm  $\times$ 4.6 mm) using a step-wise gradient solvent system (CH<sub>3</sub>CN: 0.1M TFA=4:6 in 24min, CH<sub>3</sub>CN: 0.1M TFA=1:1 in 25~34min, CH<sub>3</sub>CN: 0.1M TFA=7:3, after 35min ) at a flow rate of 0.4 mL/min. The FDLA derivatives were detected by PDA UV detector at 340nm.

### 1.3.5 ESI LC/MS Conditions

All mass spectra were carried out with a single quadrupole mass spectrometer (ThermoQuest Finnigan Navigator) in ESI-mode. The sample probe was set at 400 °C and 4 kV with an entrance cone voltage of 10V.

### 1.3.6 Spike-in experiments

The *L*-FDLA derivative of pahayokolide A hydrolysate was spiked with the *L*-FDLA derivatives of *N*-methyl-*L*-leucine, the *D*-FDLA derivatives of *N*-methyl-*L*-leucine, the *L*-FDLA derivatives of *L*-homophenylalanine and the *L*-FDLA derivatives of *D*-phenylalanine, separately.

## 1.4 Conclusions

Determination of absolute configuration is an important part of structure elucidation. Many biochemical processes depend on chirality. Proteins contain principally *L*-amino acids. However, certain microorganisms make use of both the *L* form and *D* form of amino acids. Based on Marfey's, pahayokolide A, a cyclic cyanobacterial peptide isolated from *Lyngbya* sp. strain 15-2, only contained the amino acids present in its given structure and its absolute configuration is as follows: *L*-proline (2), *L*-phenylalanine, *L*-serine, *L*-threonine, *D*-homophenylalanine, *D*-glutamine, *N*-methyl-*L*-leucine and *D*-Athmu. Pahayokolides A, C and D showed the same absolute configuration as pahayokolide B. It was also confirmed that a pendant *N*-acetyl-*N*-methyl leucine moiety in pahayokolide A was absent in pahayokolides B and pahayokolides C-D were conformers of pahayokolide A.

## **2. Biosynthetic Pathway for the Construction of the Pahayokolide A**

### **2.1 Introduction**

The biosynthesis of cyanobacterial peptides has attracted considerable attention not only because cyanobacterial peptides represent a potential source for pharmaceutical agents, but also because of the risk of exposure to humans as well as animals. Cyanobacterial peptide toxins can be synthesized via both ribosomal and nonribosomal pathways. In the ribosomal system, one of twenty proteinogenic amino acids is specifically activated by the aminoacyl-tRNA synthetase (aaRS) specific for that tRNA as aminoacyl-tRNAs and peptide bond formation is directed by the ribosome. Subsequently, a peptide bond is formed by the ribosome. Distinct from the traditional ribosomal mechanism, some microorganisms, such as bacteria and fungi, assemble small peptides (mostly less than 20 amino acids residues) by a nucleic acid-independent pathway. In these “non-ribosomal” peptides, the incorporation of a wider range of amino acid substrates, for which there exists no specific tRNA, suggests a nonribosomal origin of biosynthesis (Sieber et al., 2005). In the 1970s Lipmann et al. found *Bacillus* sp. contained multienzyme complexes which were responsible for the biosynthesis of the peptide antibiotics gramicidin S and tyrocidine A, and were similar to fatty acid synthetases (Lipmann et al., 1971). In the following years a variety of nonribosomal

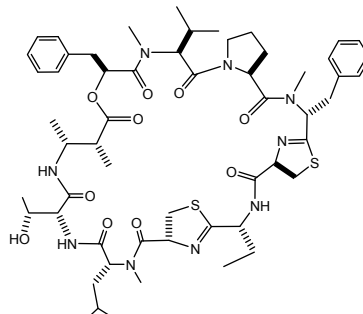
**Figure 2 - 1 Linear, cyclic and branched cyanobacterial peptides and their bioactivities**

**Grassypeptolide**

*Lyngbya confervoides*, Florida Keys

Grassypeptolide inhibited cancer cell growth ( $IC_{50}$  = 1.0 ~ 4.2  $\mu$ M.)

(Kwan et al., 2008)

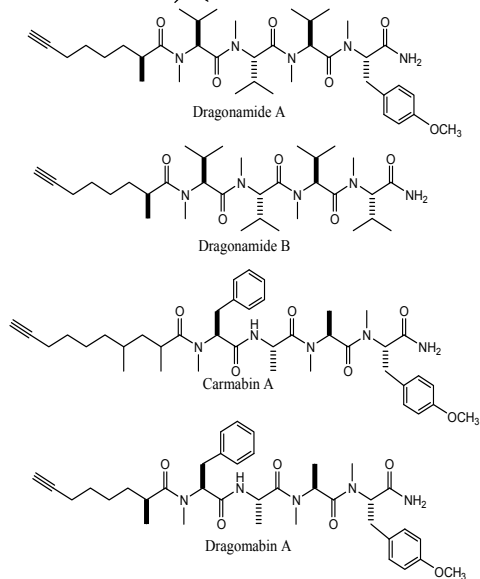


**Carmabin A, Dragomabin, Dragonamide A and B**

*Lyngbya majuscula*, Isla Bastimentos, Bocas del Drago and Crawl Cay in Bocas del Toro, Panama

Carmabin A, dragomabin, and dragonamide A showed good antimalarial activity ( $IC_{50}$  = 4.3, 6.0, and 7.7  $\mu$ M, respectively), whereas the nonaromatic analogue, dragonamide B, was inactive.

(McPhail et al., 2007)



**Microcolins A and B**

*Lyngbya majuscula*, a fore reef slope, Pta. Bobos, La Blanquilla, Venezuela

Microcolins A and B exhibit potent immunosuppressive activity

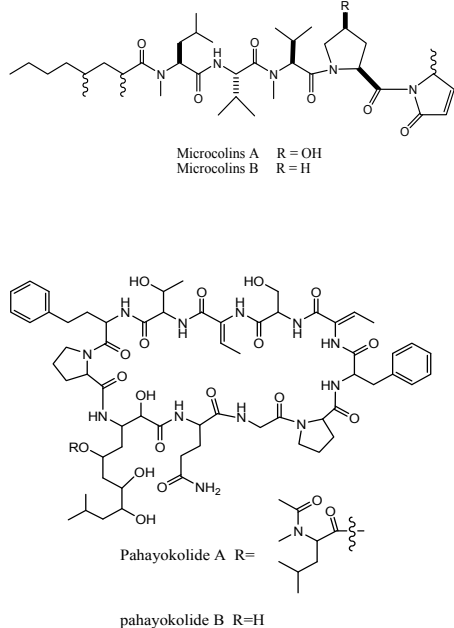
(Koehn et al., 1992)

**Pahayokolides A and B**

*Lyngbya sp.* 15-2, the Florida Everglades

Pahayokolides A and B inhibited a number of cancer cell lines over a range of conc. ( $IC_{50}$  = 2.13 ~ 44.57  $\mu$ M) depending on the cell-type, marginally toxic to brine shrimp only at the highest conc. (1 mg/mL), acutely toxic to zebrafish embryos ( $IC_{50}$  = 2.15  $\mu$ M).

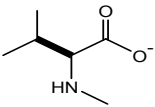
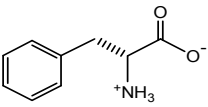
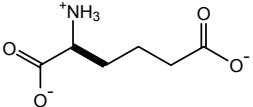
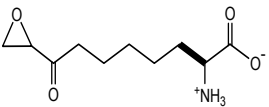
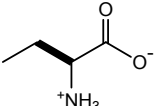
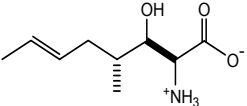
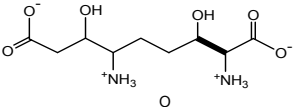
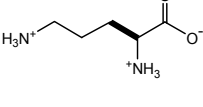
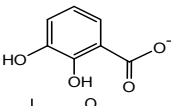
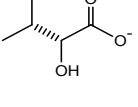
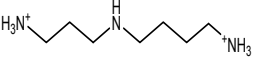
(An et al., 2007; Berry, 2004)



peptides have been isolated from microorganisms, including cyanobacteria, originating from aquatic and terrestrial habits. Nonribosomal peptides can be more complex than ribosomal peptides, and include macrocycles and/or branches (Fig. 2-1). Macrocycles can be formed through peptide or ester bonds. In contrast to the ribosomal peptides, the cyanobacterial peptides, synthesized by the non-ribosomal pathway may contain a high proportion of non-proteinogenic amino acids, such as pseudo, hydroxyl, *N*-methylated, and *D*-amino acids (Table 2-1) (Marahiel et al., 1997; Stachelhaus et al., 1995a). Some bioactive peptides have been used widely in medicine, and biological research, including antibiotic (vancomycin), immunosuppressive (cyclosporin A), antiviral (luzopeptin A) and antitumor (echinomycin and triostin A) activities.

In spite of their diverse structures and broad spectrum of bioactivities, non-ribosomal peptides share a common mode of biosynthesis. This biosynthetic process is performed by multienzyme complexes termed non-ribosomal peptide synthetases (NRPSs) (Marahiel et al., 1997). Over the past three decades, significant progress has been made in characterizing the architecture of NRPSs and understanding the molecular principles of nonribosomal peptide biosynthesis in microorganisms (Schwarzer et al., 2003; Mootz, 2002; Marahiel et al., 1997). Sequence analysis of genes which encode bacterial and fungal NRPSs has provided insight into their molecular architecture (Marahiel et al., 1997). The NRPS acts as the template for the non-ribosomal biosynthesis of polypeptides. The NRPS is organized as modules, which are independent enzymatic units in an assembly line. One module of an NRPS is responsible for one reaction cycle, and incorporates one amino acid into the peptide product (Stachelhaus et al., 1995). It is

**Table 2 - 1. Representative Examples of Non-proteinogenic Constituents of Peptide Antibiotics**

Name	Structure	Abbreviation	System(s)	Organism(s)
<u>Modified, proteinogenic amino acids</u>				
<i>N</i> -methyl aa (e.g. <i>N</i> -methyl valine)		MeVal	Cyclosporine enniatin	<i>Tolypocladium niveum</i>
<i>D</i> -aa (e.g. <i>D</i> -phenylalanine)		<i>D</i> -Phe	e.g. Bacitracin gramicidin S tyrocidine	<i>Bacillus licheniformis</i> <i>Bacillus brevis</i> <i>Bacillus brevis</i>
<u>Non-proteinogenic amino acids</u>				
$\delta$ -( <i>L</i> -amino adipic acid)		Aad	ACV-tripeptide (precursor of penicillin and cephalosporin)	<i>Penicillium chrysoogenum</i> <i>Aspergillus nidulans</i> <i>Streptomyces clavuligerus</i>
2-amino-9,10-epoxy-8-oxodecanoic acid		Aeo	HC-toxin	<i>Cochliobolus carbonum</i>
<i>L</i> - $\alpha$ -amino butyric acid		Abu	cyclosporin	<i>Tolypocladium niveum</i>
(4 <i>R</i> )-4[( <i>E</i> )-2-butenyl-4-methyl- <i>L</i> -threonine]		Bmt	cyclosporin	<i>Tolypocladium niveum</i>
2,6-diamino-7-hydroxy-azelaic acid		Dha	edeine	<i>Bacillus brevis</i>
ornithine		Orn	e.g. Bacitracin gramicidin S tyrocidine	<i>Bacillus licheniformis</i> <i>Bacillus brevis</i> <i>Bacillus brevis</i>
<u>carboxy acids</u>				
2,3-dihydroxy benzoic acid		DHB	enterobactin	<i>Escherichia coli</i>
<i>D</i> - $\alpha$ -hydroxyisovaleric acid		Hiv	Enniatin B	<i>Fusarium scirpi</i>
<u>amines</u>				
spermidine		Sperm	edeine	<i>Bacillus brevis</i>



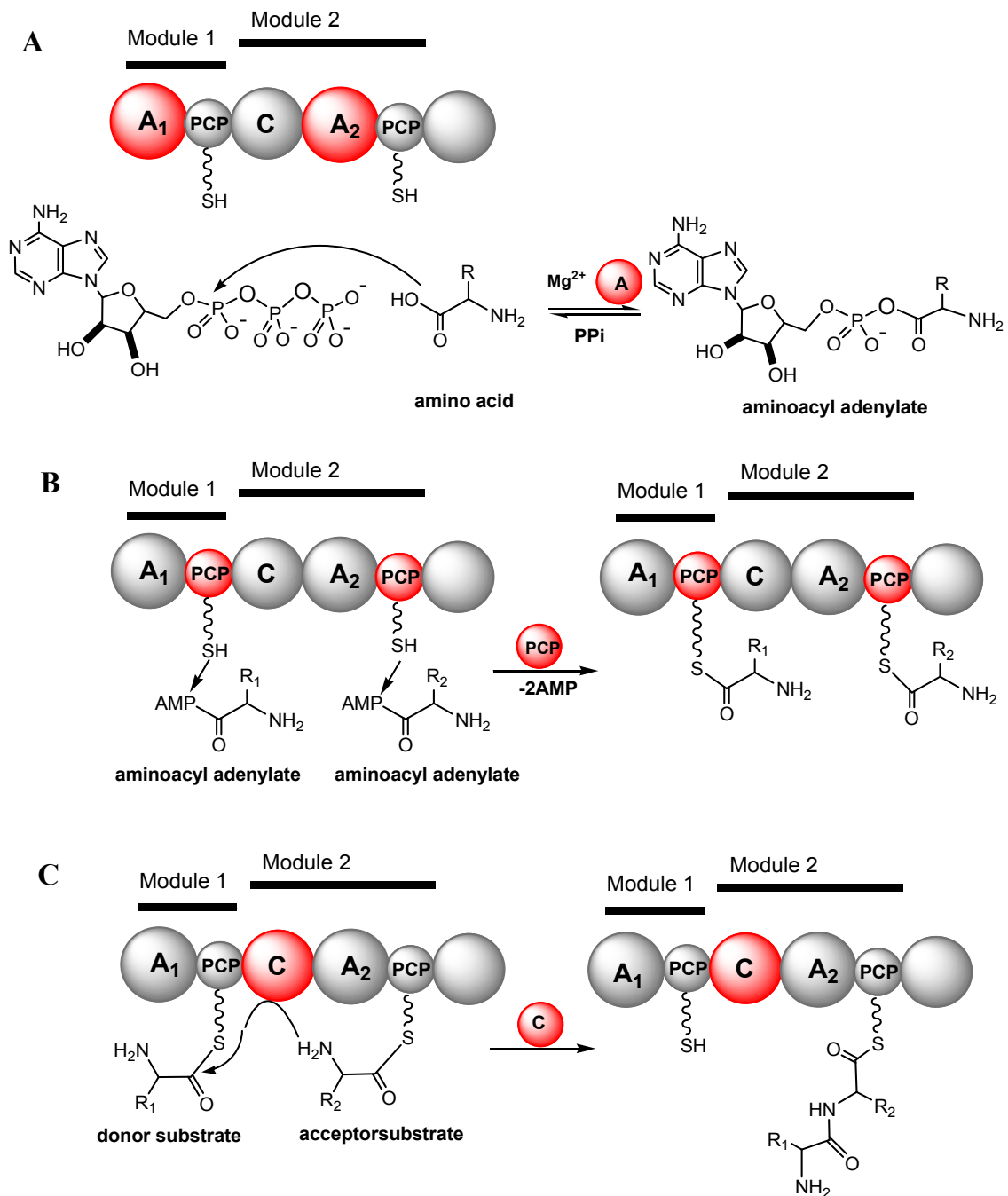
possible to predict the final peptide product on the basis of the number and order of these modules in the NRPS (Walsh et al., 2001).

A typical NRPS module is further subdivided into a set of distinct domains, each domain catalyzing a specific reaction. In a single reaction cycle, the biosynthetic steps, including substrate recognition, activation as adenylate, covalent intermediate fixation as an enzyme-bound thioester, and peptide-bond formation, can be assigned to different domains which are connected by flexible linkers (Sieber et al., 2005). A minimal elongation module consists of three core domains: a 55 kDa adenylation domain (A domain) responsible for substrate selection and activation through ATP hydrolysis (Turgay et al., 1992; Stachelhaus et al., 1995a), a 10 kDa downstream peptidyl carrier protein domain (PCP domain) for the covalent fixation as a thioester (Stachelhaus et al., 1996), and a 50 kDa condensation domain (C domain), located upstream of the A domain (Stachelhaus et al., 1998). The C domain catalyzes the peptide-bond formation between an activated amino acid and a peptidyl-bound intermediate of two adjacent modules (Fig. 2-2).

## **2.1.1 Functional domains of non-ribosomal peptide synthetases**

### *2.1.1.1 Activation by the adenylation domain*

Non-ribosomal peptide synthesis is initiated by the A domain which recognizes and activates the specific amino acid from a substrate pool, by forming the aminoacyl adenylate from ATP. Amino acid building blocks are activated by A domains to afford



**Figure 2 - 2. Domain-catalyzed reactions.**

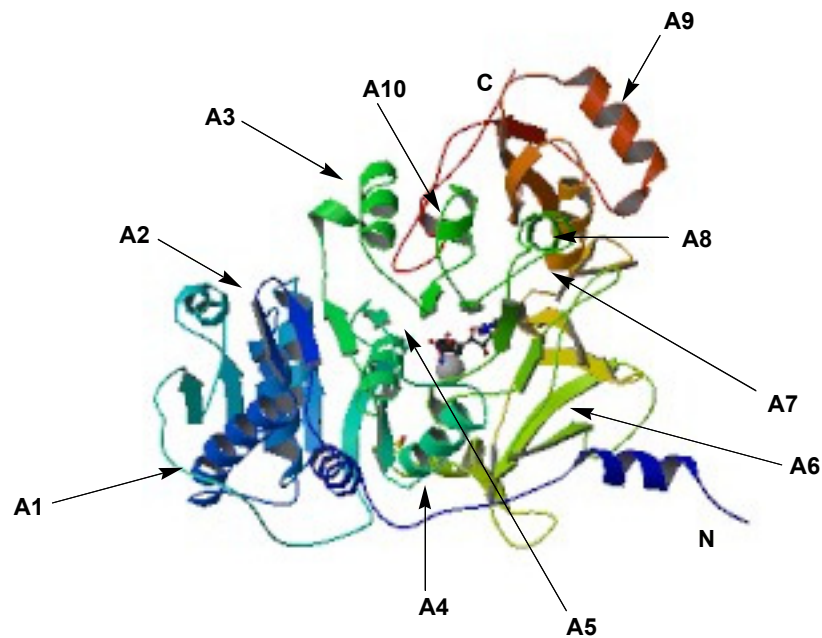
Domains in action are indicated in red. (A) Recognition and activation of a dedicated amino acid with ATP by the A domain. (B) Covalent attachment of the activated aminoacyl adenylate onto the free thiol group of the PCP-bound ppa (4'-phosphopantetheine) cofactor. (C) Peptide elongation by the C domain which catalyzes an attack of the nucleophilic amine of the acceptor substrate onto the electrophilic thioester of the donor substrate. (Stachelhaus, 1998).

aminoacyl adenylates at the expense of  $Mg^{2+}$ -ATP with release of PPi (pyrophosphate) (Figure 2-2A). The A domain belongs to the family of adenylate-forming enzymes which also contains aryl-AMP ligases, firefly luciferase, aminoacyl tRNA synthetases and the acyl-CoA ligases. The adenylate-forming enzymes catalyze the adenylation of their carboxy substrates. In the ribosomal pathways of peptide synthesis, aminoacyl tRNA synthetases also activate the amino acid substrate as aminoacyl adenylates. Despite synthetases are significantly different in their primary structures (Conti et al., 1996; Onest et al., 1995). However, in spite of their low sequence identity of 16%, firefly luciferase and the phenylalanine-activating A domain (PheA) of GrsA fold nearly identically (Baldwin, et al., 1996; Conti, et al., 1996).

**Table 2 - 2. Highly Conserved Core Motifs of the A Domain of Peptide Synthetases**

Domain	Core motifs	Consensus sequence
Adenylation	A1	L(TS)YxEL
	A2	LKAGxAYL(VL)P(LI)D
	A3	LAYxxYTSG(ST)TGxPKG
	A4	FDxS
	A5	NxTEYGP
	A6 (core 3)	GELxLxGxG(VL)ARGYL
	A7 (core 4)	Y(RK)TGDL
	A8 (core 5)	GRxDxQVKIRGxRIELGEIE
	A9	LPxYM(IV)P
	A10	NGK(VL)DR

Sequence alignments of A domains of all known NRPS show 30-60% sequence identity (Turgay, 1992). The expression and crystallization of PheA of GrsA provided insight into the 3D structure of the A domain (Marahiel et al., 1997). In the crystal structure containing the bound *L*-phenylalanine, AMP and Mg<sup>2+</sup>, PheA consists of two subdomains showing the large N-termini of ~400 residues and the small C-terminal subdomain of ~100 residues, which is connected with a short hinge (Fig. 2-3). Most of the residues responsible for substrate recognition are located in the large N-terminal subdomain. The complete loss of activity in the deletion studies



**Figure 2 - 3 Ribbon diagram of the phenylalanine-activating A domain (PheA) of gramicidin S synthetases I (GrsA).**  
The substrates, phenylalanine and AMP are drawn using a space-filling representation. The 10 core motifs (A1-A10) are indicated.

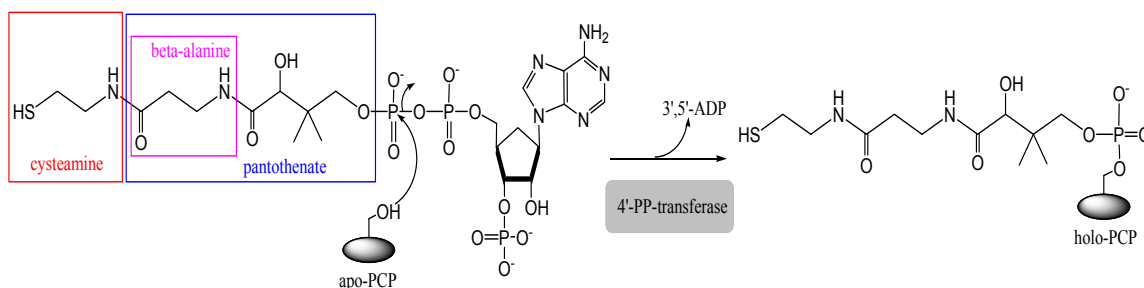
suggested that the small C-terminal subdomain is necessary for the activity of the protein (Marahiel et al., 1997). The active site is located at the interface between the two

subdomains and rotation closes the cleft between them. The substrate enters the active site through a channel-like entrance. Each A domain is composed of 10 highly conserved core motifs (A1-A10, Table 2-2) which work as 'functional anchors' for ATP binding, hydrolysis and adenylation of the substrate carboxy moiety (Sieber et al., 2005). Most of the 10 core motifs are located around the active site where the substrates are bound (Fig. 2-3).

In contrast to aminoacyl tRNA synthetases, A domains of NRPSs exhibit varying degrees of substrate selection at defined positions (Döhren et al., 1997; Kleinkauf et al., 1996). This NRPS strategy readily increases the diversity of the respective peptide products. On the other hand, no significant evolutionary pressure for accurate substrate recognition by NRPSs is expected. In the biosynthesis of cyclosporine A, enniatin, surfactin, and tyrocidine, the specificity of amino acid incorporation is a function of the availability of amino acids in the growth media (Marahiel et al., 1997). The binding pocket is located in the unconserved region between core motifs A3 and A6 (Cosmina et al., 1993). Within this region, eight amino acids control substrate specificity and can be considered the 'codons' of non-ribosomal peptide synthesis. As determined from the crystal structure of PheA of GrsA, the hydrophobic binding pocket is closed at the bottom by the indole ring of Trp239, bordered on one side by Ala236, Ile330 and Cys331, and by Ala322, Ala301, Ile299 and Thr278 on the other side. These eight amino acids located between the core motifs A3 and A6 provide a structural basis of substrate recognition in the A domains of NRPSs and could be used to predict from the primary sequence the specificity of biochemically uncharacterized A domains.

### 2.1.1.2 Intermediate transport by the peptidyl carrier protein

The PCP, also called the thiolation domain (T domain) and located downstream of the A domain within the NRPS assembly line, is responsible for transporting the intermediates to the catalytic centers (Fig. 2-2B). Before reacting with aminoacyl adenylate, the apo-PCP requires catalytic activation to its active holoform by posttranslational modification. Catalyzed by a 4'-phosphopantetheinyl transferases (4'-pp-transferases), the 4'-PP moiety of coenzyme A is transferred to the side chain of a highly conserved serine residue at the C-terminal region of the PCP (Fig. 2-4) (Marahiel et al., 1997). The activity of recombinant holo-PCP with the adjacent A domain can be tested by an aminoacylation assay with ATP and radioactive amino acids in vitro (Linne et al., 2004).

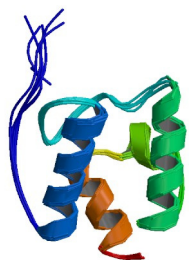


**Figure 2 - 4 Posttranslational phosphopantetheinylation**

The solution structure of the third PCP of the *B. brevis* tyrocidine synthetase was solved in 2000 (Weber et al., 2000). The PCP is a distorted four-helix bundle containing an extended loop between the first two helices (Fig. 2-5). The cofactor binds with the conserved serine residue of the PCP, which is located at the interface between the loop and the second helix (Marahiel et al., 1997). There is no evidence for the existence of a

binding pocket-like structure for the loaded substrate in the PCP, which is consistent with the observation that the PCP lacks substrate selectivity (Weber et al., 2001).

Acyl carrier proteins (ACP) from fatty acid and PKSs (polyketide synthetases) function similarly to PCPs. PCPs and ACPs fold into almost identical structures, however, these two proteins only share sequence homologies in the immediate neighborhood of the invariant serine residue. The most obvious difference between PCPs and ACPs is that the PCP surface [pI (isoelectric point) 6~7] is much more polar than the ACP surface (pI ~3.8) (Weber et al., 2001).



**Figure 2 - 5 Structure of the PCP domain derived from the third module of the *B. brevis* tyrocidine synthetase.**

### *2.1.1.3 Peptide elongation by the condensation domain*

After activation and covalent binding of the amino acid substrate by the A-PCP initiation module, the peptide bond is formed with aminoacyl substrates binding to PCPs of the downstream module (C-A-PCP). The condensation is catalyzed by the C domain. The free amino group on the downstream aminoacyl-S-PCP attacks the upstream peptidyl-S-PCP (Fig. 2-2C). C domains show low sidechain selectivity for the incoming peptidyl-S-PCP but high sidechain specificity for the incoming aminoacyl-S-PCP

nucleophile (Weber et al., 2001). Both A domains and C domains work as substrate filter in non-ribosomal peptide synthesis.

C domains share only moderate homology with each other. The C domains present at the N-terminus of the amino acid-accepting synthetase are even less conserved than internal C domains which are present between two adjacent A-PCP activating modules. The core sequences given in the Table 2-3 are more conserved for the internal C domains (de Crécy-Lagard et al.; 1995). One highly conserved core motif C3 (HHXXXDG) in the C domain was also found in acyltransferases such as chloramphenicol acetyltransferase (CAT), NRPS epimerization and heterocyclization domains (Keating et al., 2002; de Crécy-Lagard et al.; 1995). Mutational experiments proved the importance of the second histidine in the condensation reaction.

**Table 2 - 3. Highly Conserved Core Motifs of the C Domain of Peptide Synthetases**

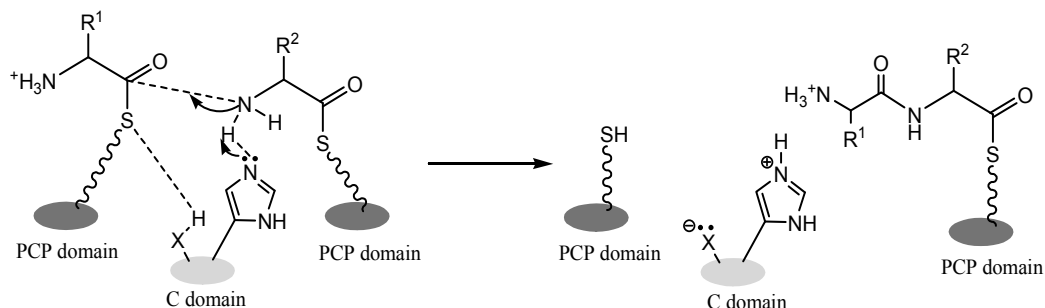
Domain	Core motifs	Consensus sequence
Condensation	C1	SxAQxR(LM)(WY)XI
	C2	RHExLRTxF
	C3 (His)	MHHxISDG(WV)S
	C4	YxD(FY)AVW
	C5	(IV)GxFVNT(QL)(CA)Xr
	C6	(HN)QD(YV)PFE
	C7	RDxSRNPL

The C domain structure was unavailable until Keating and his co-workers solved the crystal structure of VibH (Fig. 2-6) (Keating et al., 2002). VibH of the *Vibrio cholera* vibriobactin synthetase is a monomer with two pseudodimeric domains containing an  $\alpha \beta$



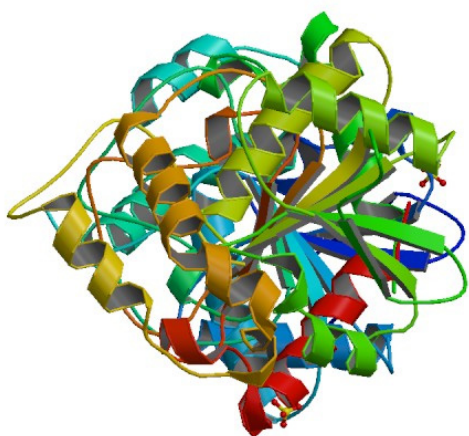


**Figure 2 - 6 Crystal structure of VibH.**  
A stand alone C domain of the *V. cholerae* vibriobactin synthetase.



**Figure 2 - 7 Suggested mechanism for the condensation/elongation reaction in peptide synthesis.**

$\alpha$  sandwich. The conserved HHXXXDG motif is present in a loop at the interface between the two domains. His126 can be reached from two different directions of the enzyme. As a base, His126 activates the free amino group of the downstream acceptor for nucleophilic attack on the upstream thioester. Two amino acid residues attached as thioesters to adjacent PCP domains via the cofactor 4'-PP and His126 of the C domain are shown (Fig. 2-7). A nucleophilic attack by the incoming amino group on the thioester activated carboxyl group of the preceding amino acid is proposed.



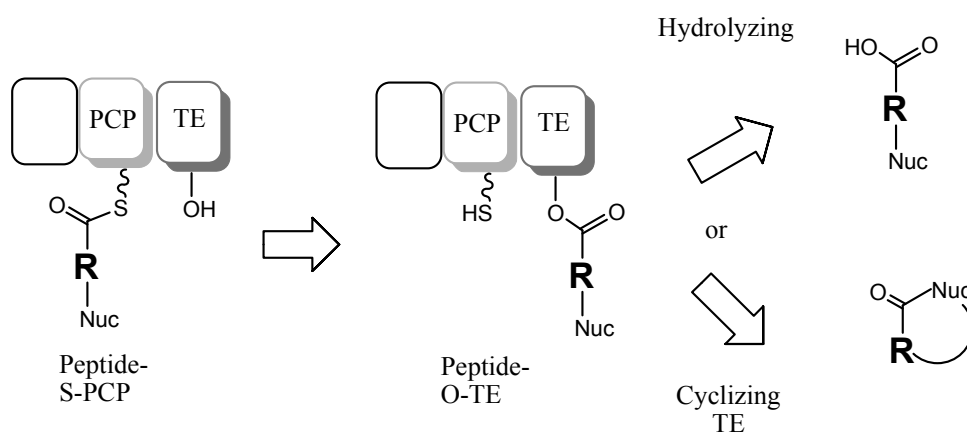
**Figure 2 - 8. Crystal structure of Srf TE.**  
( $\alpha,\beta$ -hydrolase fold).

### 2.1.2 Peptide release

Upon completion, the mature peptide must be cleaved from the NRPS to make the way for the next peptide product. A C-terminal thioesterase (TE) domain of about 250 aa catalyzes the deacylation of the most C-terminal PCP and cleaves the final product. In some NRPSs of fungal origin, the TE domain is replaced by a special C domain. Cyclosporin synthetase, the 1.7-MDa polypeptide responsible for assembly of the fungal cyclic undecapeptide immunosuppressant drug, is such an example (Weber et al., 1994). In the mechanism of TE catalysis, an ester bond is formed between the conserved serine residue of the TE and the C-terminal amino acid of the peptide product.

The crystal structure of the surfactin TE (Srf TE) domain represents an obvious  $\alpha/\beta$ -hydrolase fold which is characteristic of a large family of proteases, lipases, and esterases (Fig. 2-8A, Bruner, et al, 2002). The low homology (10-15% identity) in primary sequence of TE domains reflects the broad variety of cyclization strategies. On the other hand, the rigid fold of these enzymes ensures the precise alignment for catalytic

action. In the structure of Srf TE (Fig. 2-8B), the three essential amino acids (Ser80, His207, and Asp107) for the enzymatic activity of the catalytic triad were identified by mutational analysis (Tseng, et al., 2002). With the exception of two positive charged side chains of Lys111 and Arg120, the cavity is bound mainly by hydrophobic residues. Deacylation of the linear intermediate releases a linear peptide by hydrolysis or a cyclic product by an intramolecular nucleophile reaction (Fig. 2-9). The oxyanion hole encodes



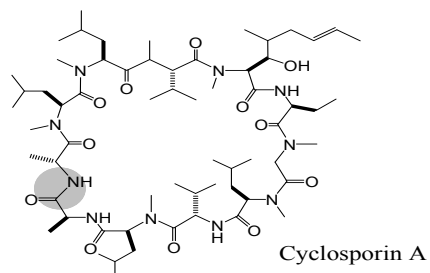
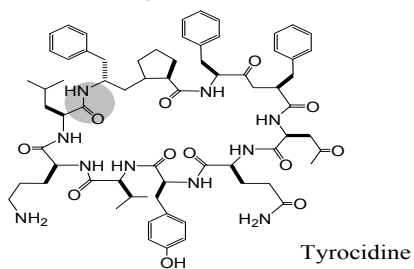
**Figure 2 - 9 Mechanism of thioesterase domain catalysis.**

a peptide-O-TE acyl-enzyme intermediate is formed by transfer of the peptidyl chain from the phosphopantethiene of the terminal peptidyl carrier protein (PCP) to the active site serine of the TE domain. For hydrolyzing TE domains, the intermediate is captured by water, generating the linear peptide; for cyclizing TE domains, an intramolecular nucleophile captures the intermediate, resulting in a cyclic product.

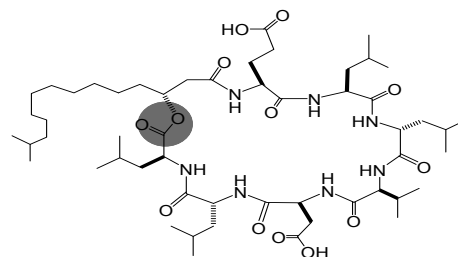
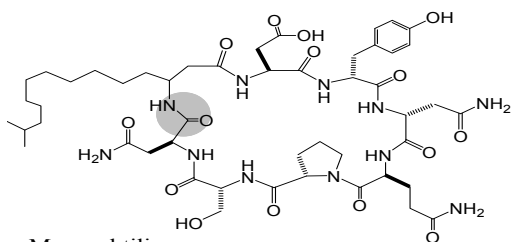
a critical stage of the two different paths. In the release reaction, the negatively charged tetrahedral intermediate is stabilized by hydrogen bonds to the amide bonds of the two amino acids with positively charged side chains. Hydrolytic cleavage is favored by a glycine residue adjacent to one of the two stabilizing amides, leaving only enough space for a water molecule to enter with ease during a critical step of catalysis. In contrast, a highly conserved proline is at the same position in the cyclizing TE domain, imparting

rigidity and favoring the intramolecular pathway. The cyclizing TE domain provides a source of structural diversity as a variety of groups can serve as the nucleophile in the cyclization reaction including the N-terminal amino group (resulting in head-to-tail cyclization) as in the peptide antibiotics tyrocidine and cyclosporine A, a side chain nucleophile (resulting in a lipo or amino acid branched cyclic molecule) as in the antibiotics bacitracin and daptomycin, and the  $\beta$ -hydroxyl group of a  $\beta$ -hydroxy fatty acid as in surfactin A (Fig. 2-10). Many polypeptides, polyketides, and their hybrids have macrocyclic structures, including the immunosuppressant cyclosporine, the antibiotics erythromycin and daptomycin, and the anticancer agent epothilone. A macrocyclic structure is a desirable feature for biologically or pharmaceutically active compounds (Stephan et al., 2003). Macrocyclic compounds exhibit greater cell permeability and resistance to cellular breakdown, and more favorable conformational orientation compared with their linear counterparts (Singh, et al., 2008). Because macrocyclic structures are less conformationally flexible than their linear analogues, the former are readily constrained to a particular biologically active conformation (Kohli, 2001). The conformational constraint that results from cyclization promotes the macrocycles to present themselves favorably to the binding region of potential molecular targets and thus influence the targeted cellular pathway. Peptidomimetic macrocycles exhibit a 2 to 5 kcal/mol savings in binding energy when interacting with proteases versus their linear counterparts because of their significant conformational restriction (Singh, et al., 2008).

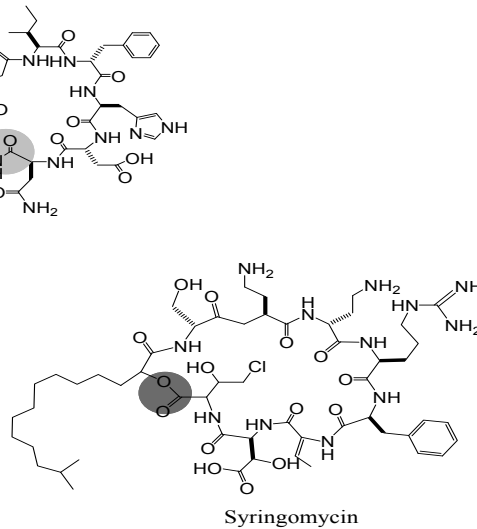
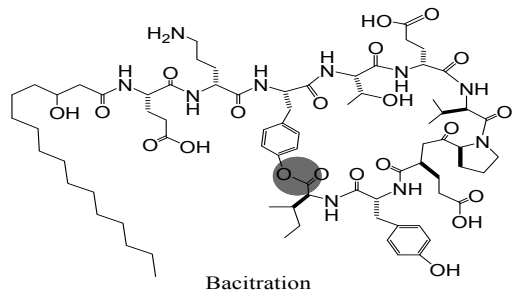
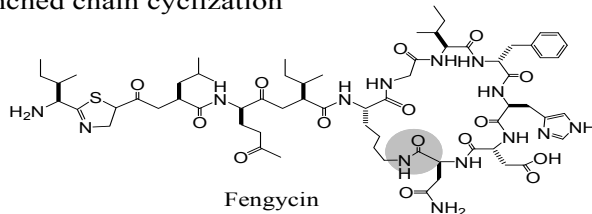
Head-to-tail cyclization



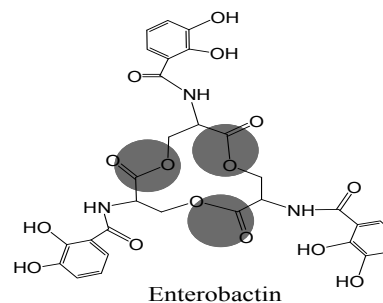
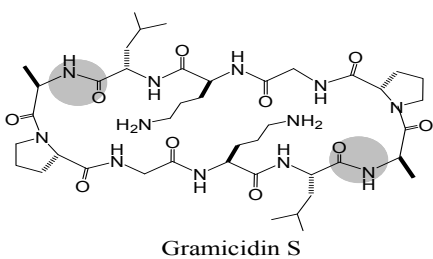
Lipo branched chain cyclization



Amino acid branched chain cyclization



Oligomerization



● Macrolactam

● Macrolactone

Figure 2 - 10. Naturally occurring macrolactones and macrolactams.

### 2.1.3 Modifying domains

In addition to the core domains, several modifying domains, such as epimerization (E), cyclization (Cy) and N-methylation (N-Mt) domains, are occasionally present in non-ribosomal peptide assembly lines and enrich the structural diversity of NRPS products (Konz, 1999).

#### 2.1.3.1 Epimerization domain

One common structural feature of NRPs is the inclusion of amino acids of *D*-configuration. Two strategies have been discovered for incorporation of *D*-amino acids into NRPs (Marahiel et al., 1997). The most common route involves epimerization of the C $\alpha$ -carbon of the PCP-tethered aminoacyl substrate by an integrated 450-amino-acid-long E domain to afford a *D/L* equilibrium (Pfeifer et al., 1995; Stachelhaus et al, 2000). The enantio-selective donor site of the downstream condensation domain ensures the specific incorporation of only the desired *D*-amino acid into the growing peptide chain. A secondary strategy of *D*-amino acid incorporation is often observed in fungal NRPSs. For example, the A domain of the cyclosporine synthetase incorporates *D*-Ala provided by an external racemase (Hoffmann et al., 1994).

The E domain is substrate specific and racemizes noncognate amino acids with lower efficiency. At this time a crystal structure of an E domain has not been acquired. However, alignment studies show the similarity of the E domain to the C domain. C and E domains share a HHXXXDG conserved motif (C3 and E2), in which the second Histidine

seems to be involved in catalysis (Table 2-4). Moreover, E domains play a crucial role in NRPS protein-protein recognition, mainly in bacterial systems (Sieber et al., 2005).

**Table 2 - 4. Highly Conserved Core Motifs of the Epimerization Domain of Peptide Synthetases**

Domain	Core motifs	Consensus sequence
Epimerization	E1	PIQ <sub>x</sub> WF
	E2 (His)	HH <sub>x</sub> ISDG(WV)S
	E3	DxLLxAxG
	E4	EGHGRE
	E5	RTV <sub>x</sub> GWFT <sub>xx</sub> YP(YV)PFE
	E6	P <sub>xx</sub> G <sub>x</sub> GYG
	E7	FNYLG(QR)

#### 2.1.3.2 Cyclization domain

Cy domains replace C domains in modules where oxazoline or thiazoline are formed and perform both intermolecular condensation and intramolecular heterocyclization reactions (Marshall et al., 2002; Keating et al., 2000). In Cy domains, the conserved Asp residues in the DXXXXD motif are critical for both condensation and heterocyclization (Keating et al., 2000). 4,5-dihydrooxazole, the heterocycle in enterobactin is formed by two Cy domains (Fig. 2-12).

### 2.1.3.3 *N*-methyltransferase domain

*N*-methylation significantly contributes to biological activities of non-ribosomal peptides and their stability against proteolytic cleavage. *N*-methylation in NRPSs is catalyzed by the 420 amino acids comprising *N*-Mt domain which is inserted into the accompanying A domain. The *N*-Mt domain contains at least three core motifs (M1-M3, Table 2-5). The glycine-rich sequence VL(ED)GxGxG (M1) exhibits significant similarity to the common S-adenosylmethionine (SAM) binding site of a heterologous class of co-substrate-dependent methyltransferases (Marahiel et al., 1997). Recombinant protein and deletion studies reveal that *N*-methylation takes place at the thioester stage before peptide bond formation (Haese et al., 1994; Pieper et al., 1995; Billich et al., 1987). With the help of the *N*-Mt domain, the S-methyl group of SAM is transferred to the  $\alpha$ -amino group of the thioesterified amino acid and S-adenosylhomocysteine is released as a reaction byproduct.

**Table 2 - 5. Highly Conserved Core Motifs of the *N*-methylation Domain of Peptide Synthetases**

Domain	Core motifs	Consensus sequence
N-methylation	M1	VL(DE)GxGxG
	M2	NEL <sub>S</sub> YRY <sub>X</sub> AV
	M3	VE <sub>X</sub> S <sub>X</sub> ARQ <sub>X</sub> G <sub>X</sub> LD

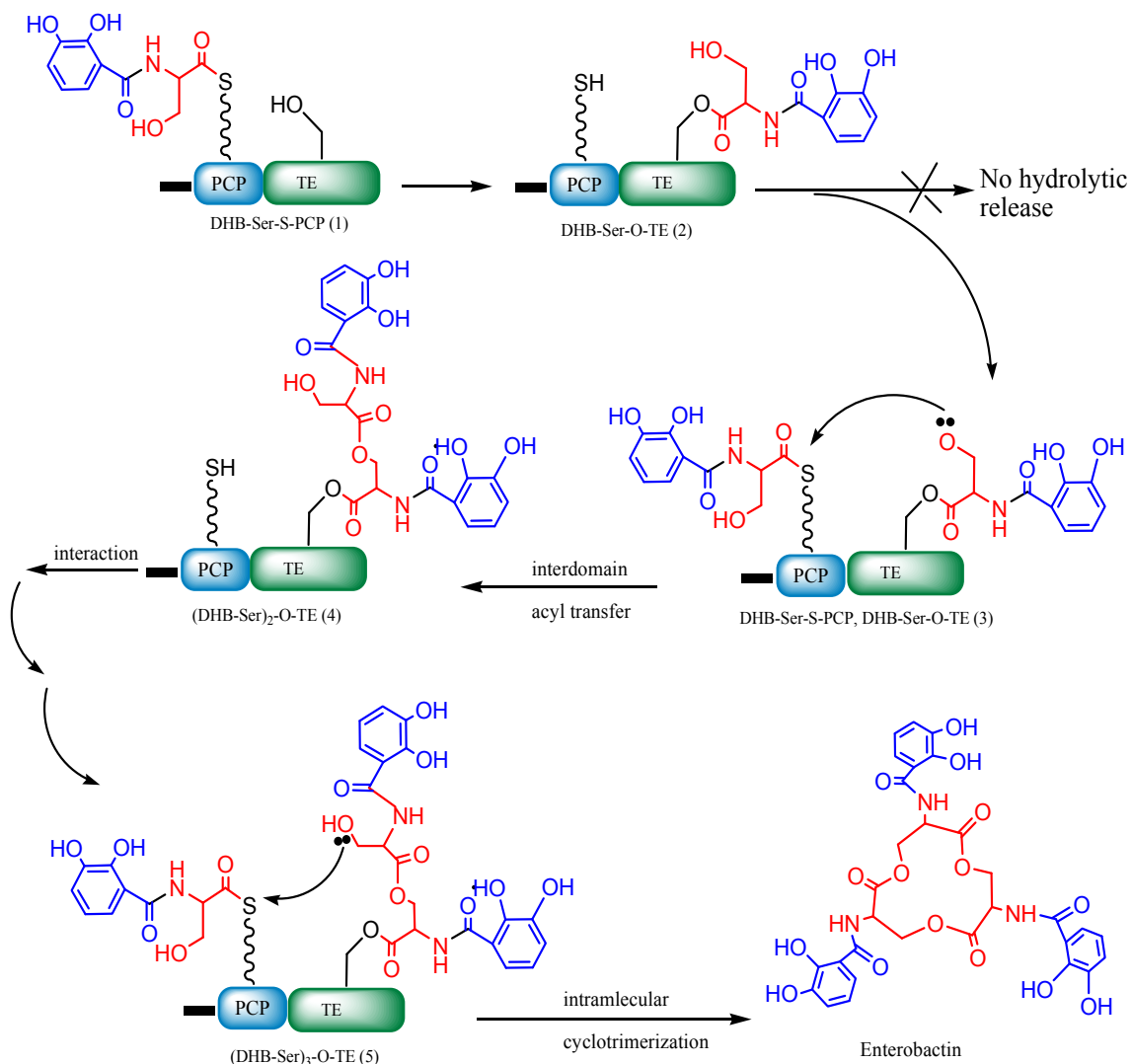
### 2.1.4 Deviations from NRPS co-linearity rule

The co-linearity rule refers to the 1:1 correspondence between the number of modules present and amino acids in the final product with the order of the amino acids in



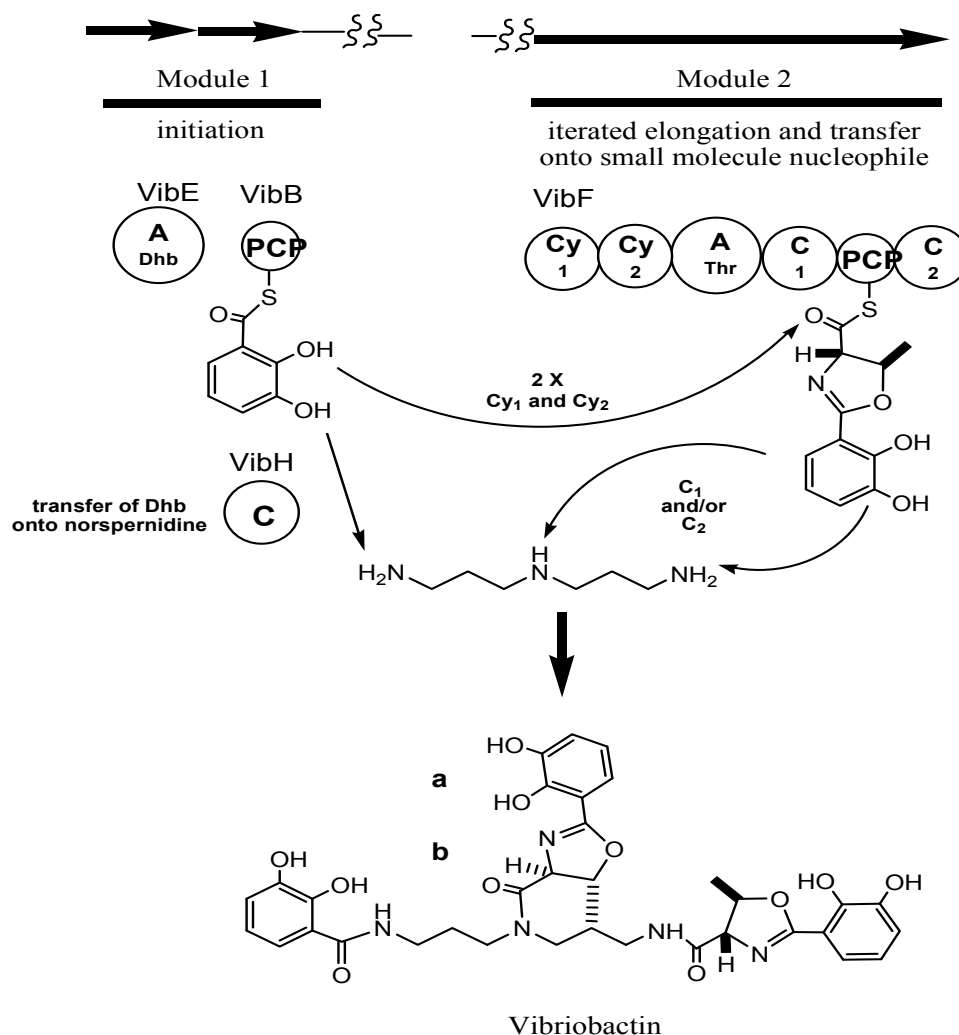
the peptide determined by the order of the modules in the synthetase. In most cases the peptide products are co-linear with the number and organization of the modules within NRPSs. The starting module only contains A and PCP domains. The typical elongation module is organized as C-A-PCP. Usually a TE domain in the terminal module hydrolyzes the full peptide from the immediate upstream PCP domain. In the case of an iterative NRPSs use their modules iteratively and produce peptides containing repeated units (Fig. 2-11). Iterative and linear NRPSs cannot be distinguished by their primary sequences. The iterative ester-bond-forming step has not yet been demonstrated experimentally. Known iterative enterobacterin and gramicidin S NRPSs show that the terminal TE domain controls the multimeration process (Shaw-Reid et al, 1999). It is presumed that in most PKS and NRPS enzymes, the rate of acyl chain release from the covalent acyl-O-TE domain would be variable. For cyclization catalyzing TE domains hydrolysis must be suppressed, and the acyl-O-TE variant must be stable enough to wait for capture by a hydroxyl or amide sidechain internal to the natural product acyl chain. In EntF the mono DHB (dihydroxybenzoate)-Ser-O-TE (Fig. 2-11, 2) must be relatively stable and serve as a waiting room for interdomain acyl transfers of covalent intermediates. The free phosphopantetheinylated PCP domain of EntF is reacylated by seryl-AMP from the adenylation domain of EntF and the secondary DHB-Ser chain in thioester linkage on PCP (Fig. 2-11, 3) is produced through the condensation domain. At this point, catalysis of ester-bond formation could occur by attack of the sidechain of the seryl moiety on the DHB-Ser-O-TE (Fig. 2-11, 2) onto the carbonyl of the thioester linkage of DHB-Ser-S-PCP (Fig. 2-11, 3) to effect ester synthesis, chain translocation and generation of the dimeric DHB-Ser-O-DHB-Ser-O-TE (Fig. 2-11, 4). The trimeric

(DHB-Ser)<sub>3</sub>-O-TE (Fig. 2-11, 5) would result from iteration of the ‘waiting room’ proposal. The final product is released through cyclization. However, the alignment of iterative and common TE domains failed to explain such an iterative function.



**Figure 2 - 11. Enterobactin NRPS as an example for iterative NRPS.**

Proposed mechanism for elongation, lactonization and termination of enterobactin by EntF TE domain. non-ribosomal peptide consisting of  $n$  amino acids, the core domains within its linear NRPS are arranged in the order A-PCP-(C-A-PCP) <sub>$n-1$</sub> -TE. However, some exceptions, such as iterative and nonlinear NRPSs, don't follow the co-linearity rule (Mootz, 2002). The TE domain serves as a ‘waiting room’ for interdomain acyl transfers of covalent intermediates (species 1-5) from the upstream peptidyl carrier protein and it is the putative location for intermolecular cyclotrimerization to release enterobactin.



**Figure 2 - 12 Vibriobactin NRPS as an example for nonlinear NRPS**

In addition to iterative NRPSs, the putative nonlinear NRPSs are not rare. Nonlinear peptide products result from unusual internal cyclization or branch-point syntheses. Vibriobactin synthetase (vibF) contains six-domain (Cy1-Cy2-A-C1-PCP-C2) (Fig. 2-12). The free, soluble, small norspermidine molecule is used as nucleophiles for the carbonyl groups of the intermediates of DHB and threonine. The C domain of vibH transfers the first DHB from DHB-S-vibB to the amine group of a norspermidine molecule. VibF only contains one A domain and one PCP to build two Dhb-threonyl-S-

PCPs for each final product, but two Cy domains and two C domains. It is believed that each C domain catalyze transfer one Dhb-Oxa on the norspermidine acceptor because recognition of the middle and terminal amine might require a dedicated catalyst. Cy2 is capable of and limited to the condensation (amide bond formation) step of the three-step heterocyclization process, while Cy1 is capable of and limited to the final processing (cyclization/dehydration) steps to the completed heterocycle (Marshall et al., 2002).

### **2.1.5 Polyketide synthetases associated with nonribosomal peptide synthesis**

Polyketide synthetases (PKS) are multimodular enzyme complexes which share a similar mode of product assembly as NRPSs and also possess a modular arrangement (Katz et al., 1993). Instead of assembling amino acids, PKSs link acetate and propanate as the primary building block. PKSs can be divided into three classes (Gokhale, et al., 2001). Type I PKSs are organized into modules with each module responsible for one cycle of chain elongation. Type I PKSs produce nonaromatic polyketides and currently represent all cyanobacterial PKSs to date. Type II are only found in bacteria and typically involved in the biosynthesis of aromatic polyketides. The individual domain of type II PKSs are used iteratively for each cycle of chain elongation. Type III PKSs are homodimeric condensation enzymes which are responsible for biosynthesis of flavonoids and chalcone compounds (Barrios-Llerena et al., 2007). Typically, a module of the type I PKS contains at least an acyltransferase (AT) domain for extender unit loading, an acyl carrier protein (ACP) domain for holding the growing polyketide chain as a thiol ester, and a  $\beta$ -ketoacyl synthetase (KS) domain catalyzing the decarboxylative condensation

between the extender unit and the growing polyketide chain. Both of the PCP for PKSs and the ACP for NRPSs are post-translationally modified by 4'-pp-transferases. The growing acyl chain can be modified by  $\beta$ -ketoreductase (KR), enoyl reductase (ER), dehydratase (DH) or methyl transferase (MT) domains before the final polyketide product is released from the assembly enzyme via the thioesterase (TE) domain (Du et al., 2001). Because of the similar molecular architecture of NRPSs and Type I PKSs, many cyanobacterial secondary metabolites are, in fact, fusions of NRPs and polyketides. In essence, this occurs when Type I PKS modules follow NRPS modules, and vice versa.

### **2.1.6 Biosynthesis of non-ribosomal peptides in cyanobacteria**

Over the last three decades, a large number of new compounds have been isolated from over 4000 strains of marine and freshwater cyanobacteria (Burja et al., 2001). Cyanobacterial secondary metabolites possess a wide spectrum of toxicities or biological activities and various chemical structures, including nitrogen-containing lipids, novel cyclic and linear lipopeptides, fatty acids and other chemicals (Gerwick, et al., 2001; Burja et al., 2001; Mayer et al., 2003). Among these secondary metabolites, peptides represent a large subclass of bioactive natural products with unique characteristics such as non-proteinogenic amino acids, *D*-amino acids, heterocyclic elements, and glycosylated as well as N-methylated residues. Cyanobacteria frequently produce peptides by NRPSs or NRPS/PKS (polyketide synthetase) hybrids. The main structure of the resulting peptide or hybrid of peptide/polyketide product depends on the number and order of modules in NRPS or NRPS/PKS system.

Since the first cyanobacterial NRPS cluster which coded for microcystin biosynthesis was cloned and characterized from *Microcystis aeruginosa* (Dittmann et al., 2001), a number of NRPS or hybrid NRPS/PKS gene clusters have been identified and sequenced from cyanobacterial genomes (Welker et al., 2006). Only a few NRPS-containing biosynthetic gene clusters have been characterized from the marine cyanobacterium *Lyngbya majuscula*, including the gene sets of barbamide (Chang et al., 2002), curacin A (Chang et al., 2004), lyngbyatoxins (Edwards et al., 2004a), jamaicamides (Edwards et al., 2004b), hectochlorin (Ramaswamy et al., 2007). No biosynthetic gene cluster has been isolated from a freshwater *Lyngbya*. Indeed only a few peptides have been isolated from freshwater *Lyngbya*. The cyclic peptides, pahayokolides A and B are isolated from a freshwater *Lyngbya* sp. from the Florida Everglades (An et al., 2007). Both pahayokolides A and B contain a cyclic undecapeptide and an unusual  $\beta$ -amino acid, Athmu (3-amino-2,5,7,8-tetrahydroxy-10-methylundecanoic acid). Although  $\beta$ -amino acids are common in cyanobacterial peptides, only the pahayokolides, lyngbyazothrins A-D and schizotrin A contain Athmu (Zainuddin et al., 2009; Pergament et al., 1994).

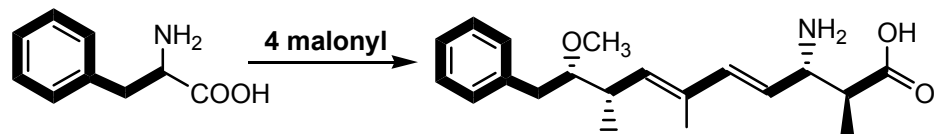
Many microbes are not amenable to cultivation and require time-consuming empirical optimization of incubation conditions for mass production of desired secondary metabolites for clinical and commercial use (Shuler, 1994). Therefore, a fast, simple system for heterologous production of natural products is much desired. In chemical synthesis of bioactive cyanobacterial peptides, the use of protecting groups and chiral catalysts in regio- and stereoselective reactions is common. Compared with natural NRPS or NRPS/PKS strategies, chemical synthesis is difficult to achieve in satisfying yields.

NRPSs represent the enzymatic assembly line for the biosynthesis of pharmacologically relevant peptides. Understanding the architecture of NRPSs and how they incorporate a diverse group of precursors into complex peptides is important, not only for revealing the mechanisms of new enzymatic reactions, but also for engineered production of new peptides with better therapeutic properties. Recombinate NRPS and NRPS/PKS facilitate the sustainable bioproduction of peptides. Herein we report the cloning, partial sequencing and characterization of the Pahayokolide A biosynthetic genes of *Lyngbya* sp. strain 15-2. This study provides a better knowledge of these metabolic products and underlying biological mechanisms, and ultimately to contribute to the redesigning of natural products in order to obtain new bioactive compounds for drug discovery.

## **2.2 Results and Discussion**

### **2.2.1 The biosynthetic origin of the Athmu moiety in pahayokolides A**

$\beta$ -Amino acids are not uncommon in cyanobacterial peptides and include Adda (3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-4,6-decadienoic acid) in microcystins (Carmichael, 1994), and nodularin (Annala et al., 1996; Drakenberg, 1996), Ahoa (3-amino-2,5-dihydroxy-8-phenyloctanoic acid) in nostophycin (Fujii et al., 1999), Ahda (3-amino-2,5,9-trihydroxy-10-phenyldecanoic acid) in scytonemin A (Helms et al., 1988). The biosynthetic precursors and pathway for the Adda moiety of microcystin-LR is the only cyanobacterial  $\beta$ -amino acid whose pathway has been elucidated. The Adda moiety of



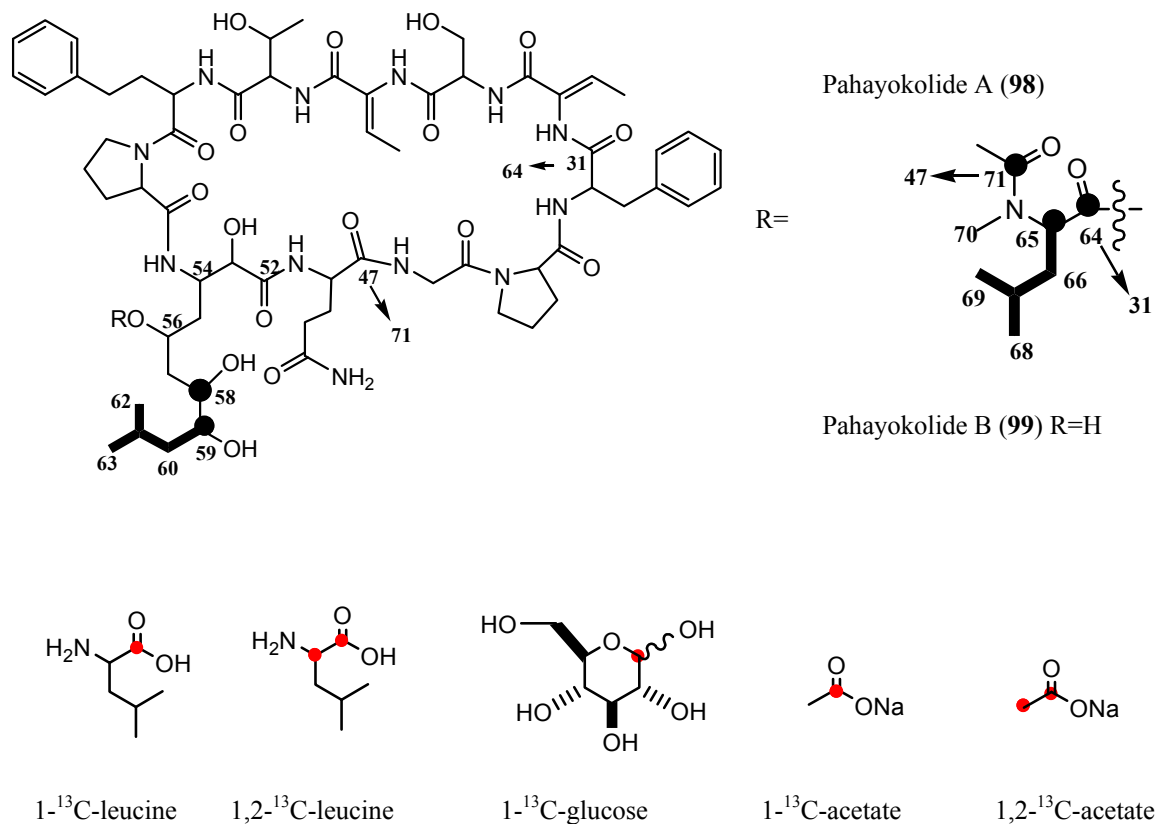
**Figure 2 - 13. Proposed biosynthesis of Adda**

the microcystins has been shown to be of mixed peptide and polyketide origin. Stable isotope feeding experiments demonstrated that the biosynthesis of the Adda sidechain begins with a phenylalanine starter unit which is extended by four rounds of condensation with malonate (Fig. 2-13) (Moore et al., 1991). Thus four PKS (polyketide synthetase) modules are involved in the biosynthesis of the polyketide amino acid Adda (Tillett, et al., 2000). Inspection of the Athmu residue of the pahayokolides suggests a polyketide origin as well. We anticipated that the starter unit for the biosynthesis of the Athmu sidechain could either be valine (which would require subsequent incorporation of an additional seven carbons) or leucine or  $\alpha$ -ketoisovalerate, which could be further extended by three rounds of condensation with malonate units (Fig. 2-22). We anticipated that the later pathway would be more likely than the former. Mehner, et al. (2008), recently proposed a similar polyketide pathway for the  $\alpha$ -hydroxy- $\beta$ -amino acids Adha, Aoha and Atpoa.

Laboratory cultures of *Lyngbya* sp. strain were incubated with  $^{13}\text{C}$  labeled substrates (Fig. 2-14) and harvested. Pahayokolide A was isolated by standard methods (Berry et al., 2004). Isolated pahayokolide A was analyzed by  $^1\text{H}$  and  $^{13}\text{C}$  NMR to identify sites of  $^{13}\text{C}$  incorporation. Assignments of  $^{13}\text{C}$  NMR signals and the results of isotope incorporation into the Athmu and *N*-methyl-*N*-acetyl leucine subunits of



pahayokolide A (**98**) are presented in Table 2-6 and Fig 2-14. These values were calculated using the procedure described by Gerwick for barbamide and only the Athmu carbons-describe (Williamson, et al., 1999).



**Figure 2 - 14. The structure of pahayokolide A-B and  $^{13}\text{C}$  labeled substrates.**  
Labeled  $^{13}\text{C}$  atoms are in red.

**Table 2 - 6. Isotopic enrichments based on <sup>13</sup>C NMR for pahayokolide A**

position	δ <sub>C</sub>	[1- <sup>13</sup> C]-Leu	[1,2- <sup>13</sup> C]-Leu	[1- <sup>13</sup> C]-Acetate
Athmu				
52	172.8, s	0.76	0.51	1.02
53	72.1, d	0.70	0.95	1.41
54	48.8, d	0.56	1.36	0.72
55	35.0, t	0.34	0.39	0.93
56	69.8, d	0.36	0.72	1.40
57	36.9, t	0.94	0.58	0.77
58	71.1, d	<b>2.02</b>	<b>3.98</b>	<b>1.94</b>
59	72.6, d	0.69	<b>3.29</b>	0.79
60	41.1, t	1.42	1.03	1.49
61	24.3, d	1.33	1.22	0.97
62	23.4, q	1.81	1.76	0.82
63	21.6, q	1.64	1.59	0.88
<i>N</i> -Me-Leu				
64	172.4, s	2.26	<b>2.44</b>	<b>2.03</b>
65	55.4, d	0.35	<b>2.71</b>	1.53
66	36.8, t	1.5	0.64	1.09
67	24.8, d	1.32	1.48	0.93
68	22.9, q	1.25	0.93	0.61
69	21.2, q	1.19	1.10	0.86
70	32.8, q	0.71	0.66	0.71
<i>N</i> -Acyl				
71	174.3, s	0.98	0.73	<b>3.80</b>
72	21.7, q	1.09	1.35	1.00

<sup>a</sup> Significant enrichments are shown in bold.

The  $^{13}\text{C}$  NMR spectrum ( $d_6$  DMSO/  $\text{D}_2\text{O}$ ) of pahayokolide A (**98**) derived from  $[1-^{13}\text{C}]$  leucine fed cultures showed significant enrichment at C-58 (Fig. 2-15) and C-64 (Fig. 2-16) while doubly labeled  $[1, 2-^{13}\text{C}]$  leucine showed enrichment at C-58, C-59, C-64 and C-65 as well as the anticipated  $^{13}\text{C}$ - $^{13}\text{C}$  coupling (Fig. 2-17 and 2-18). Pahayokolide A (**98**) derived from  $[1-^{13}\text{C}]$  acetate fed cultures showed significant enrichment (nearly fourfold) at C-71 and more modest enrichment at C-58 and C-64 (Fig. 2-19 and 2-20). The experiment reported here demonstrated that our original NMR assignments of C-71 and C-47, C-31 and C-64 were switched. The corrected assignments are C-71, 174.3 ppm; C-47, 173.3 ppm; C-31, 172.6 ppm; C-64, 172.4 ppm. The incorporation of  $[1-^{13}\text{C}]$  acetate at C-58 and C-64 occurs via the condensation of acetate with  $\alpha$ -ketoisovalerate during the biosynthesis of leucine from valine (Fig. 2-21).

No enrichment of C-52 through C-57 was observed by feeding *Lyngbya* sp. Strain 15-2 with  $[1-^{13}\text{C}]$  acetate indicating that the Athmu moiety is not directly derived from acetate. The results of our  $[1-^{13}\text{C}]$  acetate feeding experiments are not consistent with our original hypothesis that the Athmu moiety is derived from three units of acetate via condensation by a polyketide synthetase. The failure to observe incorporation of  $[1-^{13}\text{C}]$  acetate at C-52, C-54 and C-56 was surprising; however enrichment at C-71, C-58, and C-64 serves as an internal control indicating that the  $[1-^{13}\text{C}]$  acetate feeding experiment was successful. Clearly the Athmu moiety is not constructed directly from acetate (Fig. 2-22). Alternatively, we suggest that these  $\alpha$ -hydroxy- $\beta$ -amino acids may arise from incorporation of an amino sugar such as kanosamine, ezoaminuroic acid and desosamine (Fig. 2-23). Amino sugars are part of the peptidoglycan cell wall of gram negative

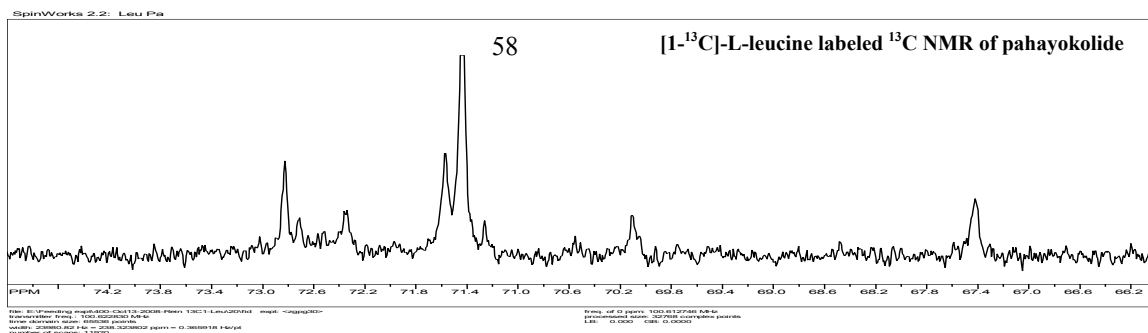
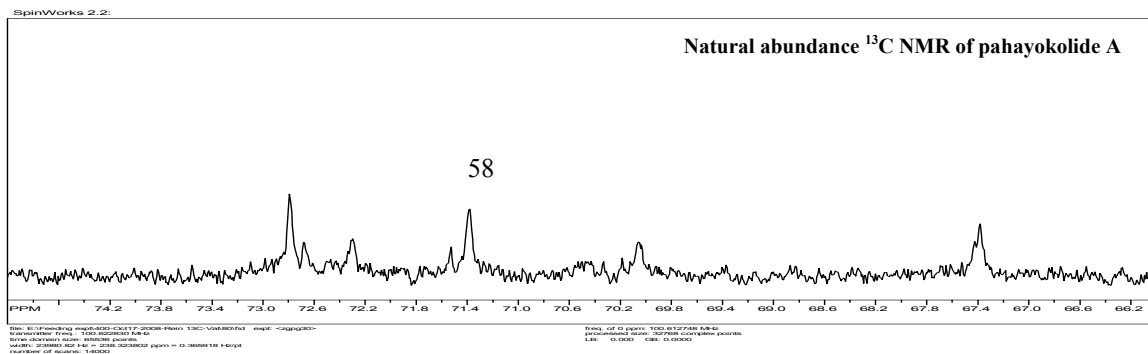


Figure 2 - 15. Natural abundance (top) and [1- $^{13}\text{C}$ ]-L-leucine labeled (bottom)  $^{13}\text{C}$  NMR of pahayokolides A (98) (66~76 ppm)

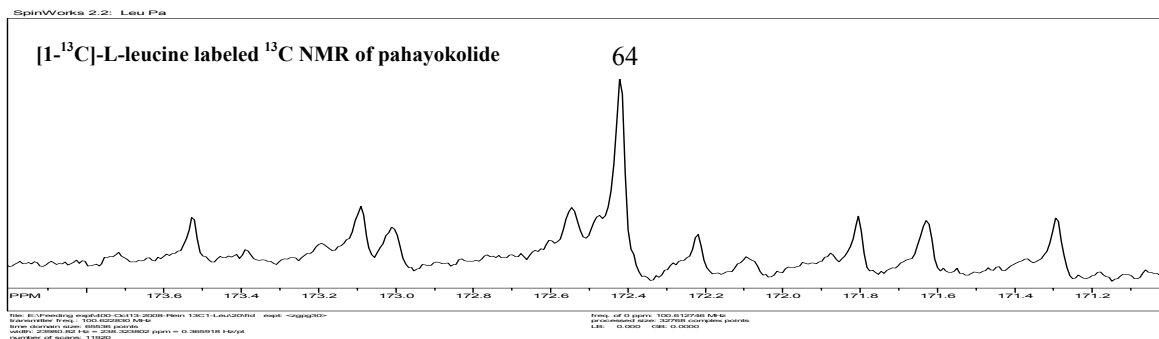
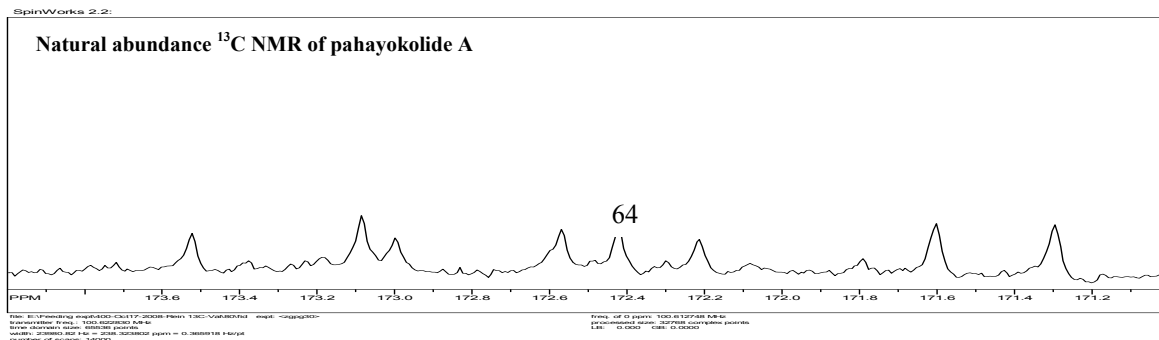
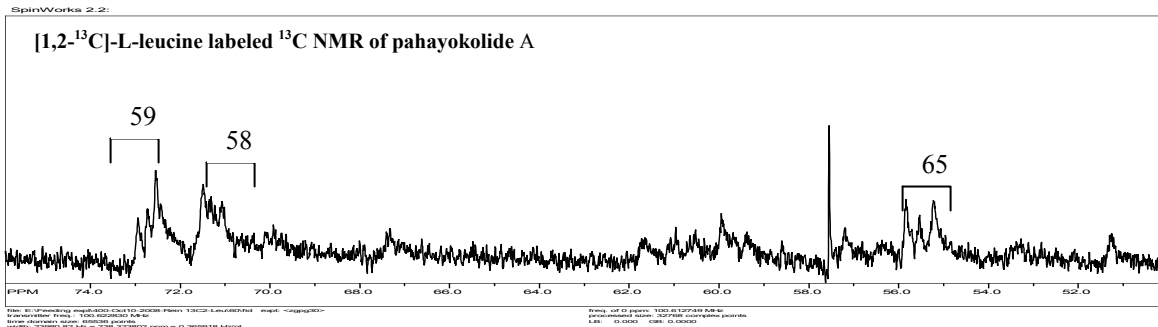
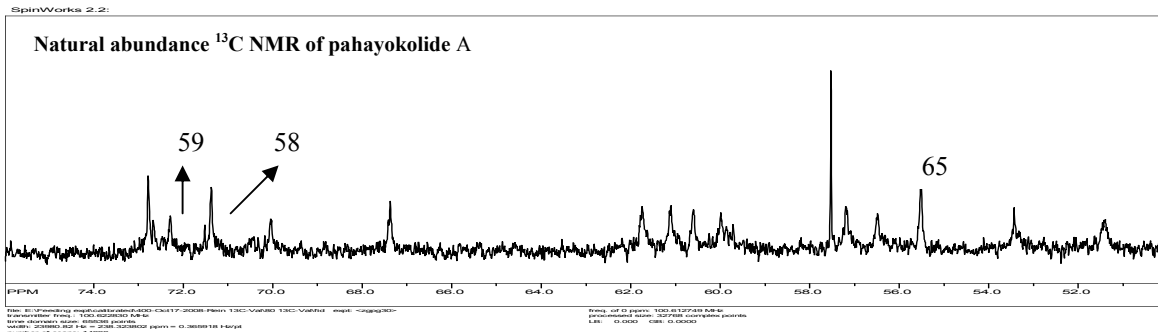
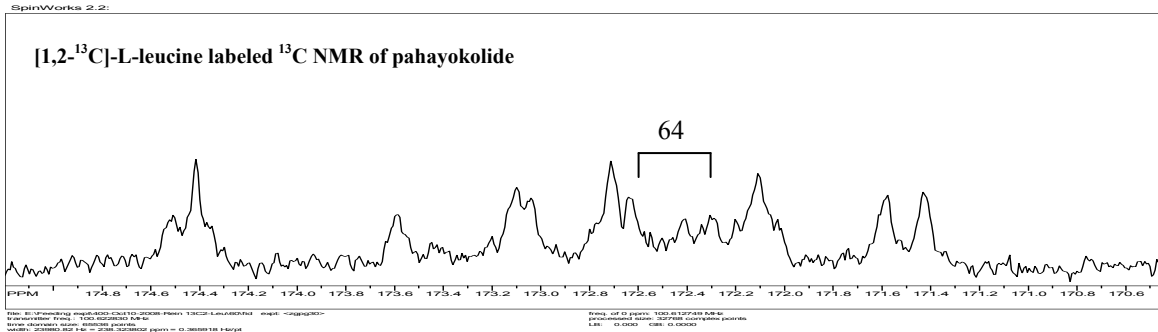
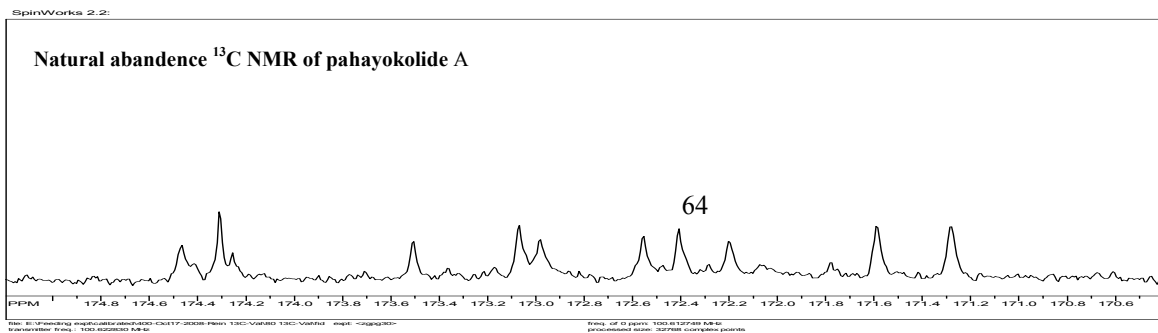


Figure 2 - 16. Natural abundance (top) and [1- $^{13}\text{C}$ ]-L-leucine labeled (bottom)  $^{13}\text{C}$  NMR of pahayokolides A (98) (170~175 ppm)



**Figure 2 - 17. Natural abundance (top) and [1,2-<sup>13</sup>C]-L-leucine labeled (bottom) <sup>13</sup>C NMR of pahayokolide A (98) (66~76 ppm)**



**Figure 2 - 18 Natural abundance (top) and [1,2-<sup>13</sup>C]-L-leucine labeled (bottom) <sup>13</sup>C NMR of pahayokolides A (98) (170~175 ppm)**

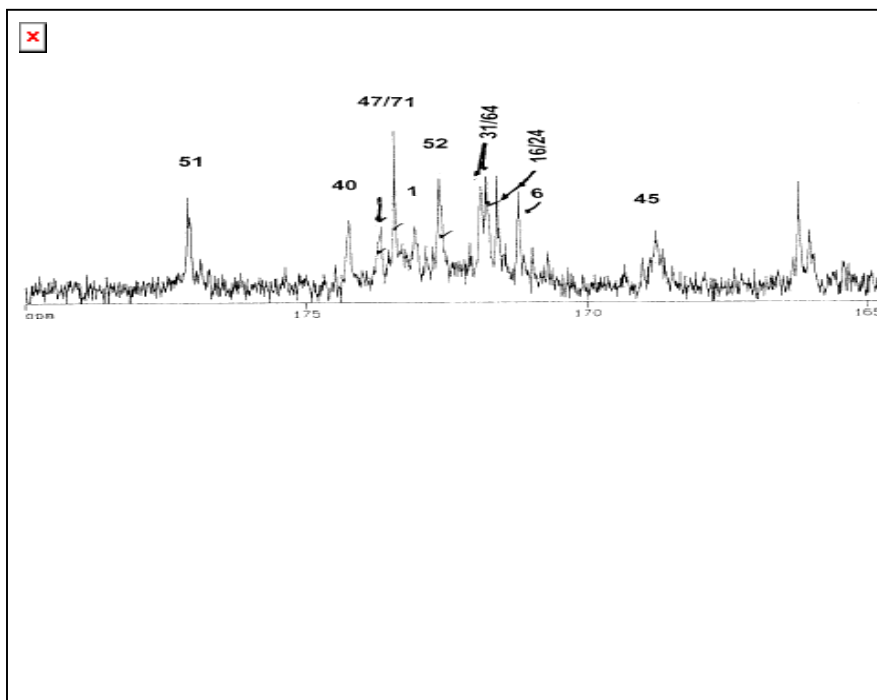


Figure 2 - 19. Natural abundance (top) and [1-<sup>13</sup>C]-acetate labeled (bottom) <sup>13</sup>C NMR of pahayokolides A (98) (165~180 ppm)

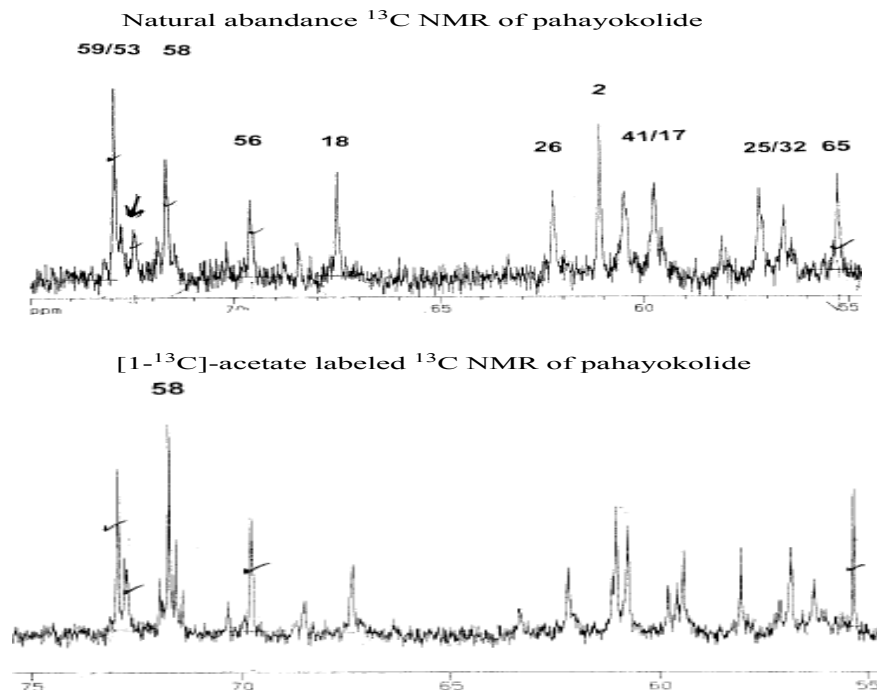


Figure 2 - 20. Natural abundance (top) and [1-<sup>13</sup>C]-acetate labeled (bottom) <sup>13</sup>C NMR of pahayokolides A (98) (55~75 ppm)

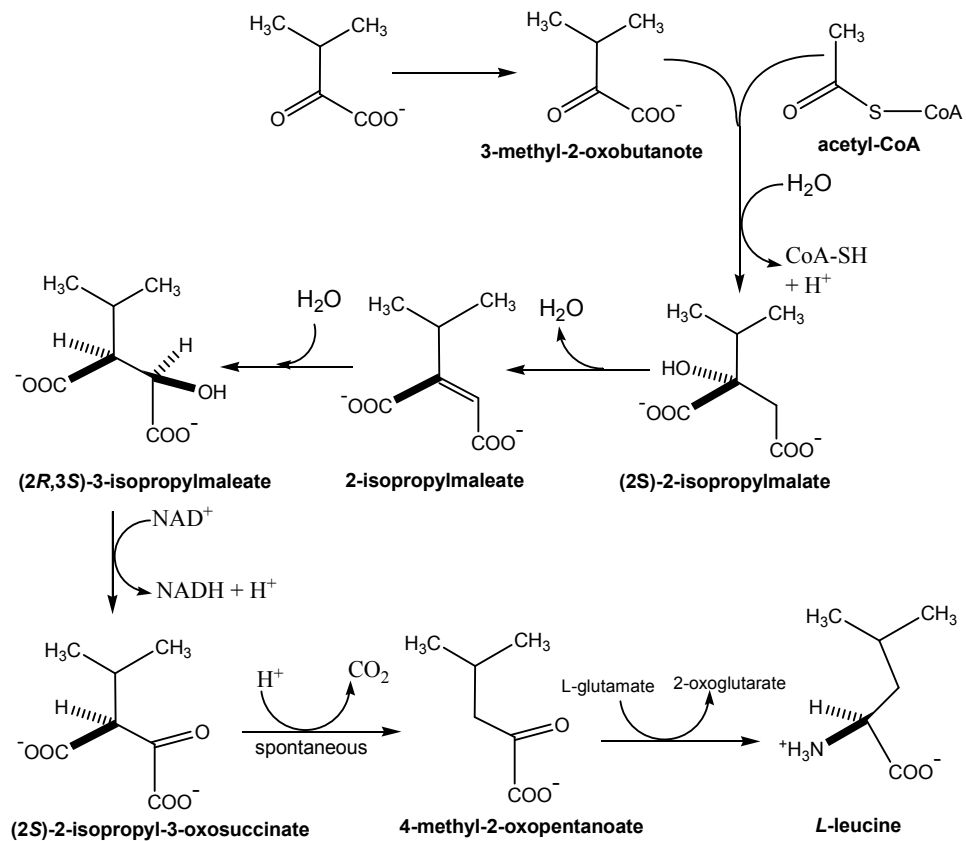


Figure 2 - 21. Biosynthesis of *L*-leucine from valine

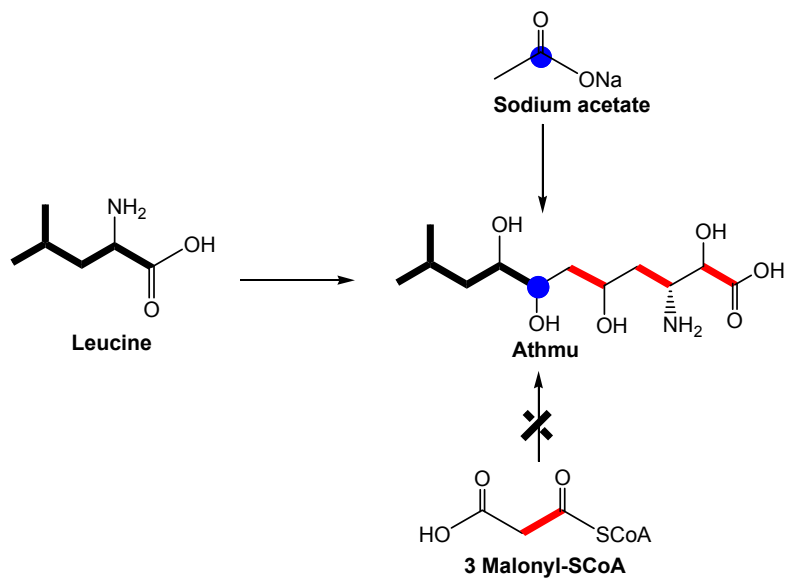
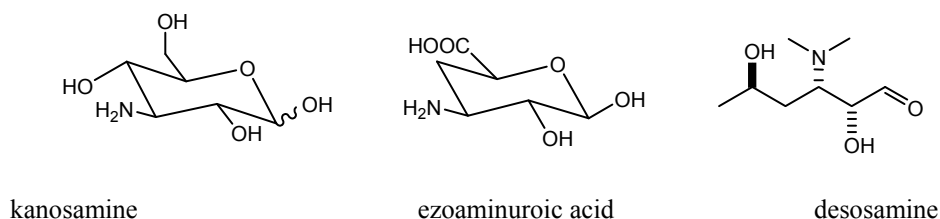


Figure 2 - 22. Proposed biosynthesis of Athmu

bacteria including cyanobacteria. Feeding of [1-<sup>13</sup>C] glucose to *Lyngbya* sp. failed to yield C-52 labeled pahayokolide A. However, it is possible that glucose is too far upstream of the Athmu precursor to show enrichment.



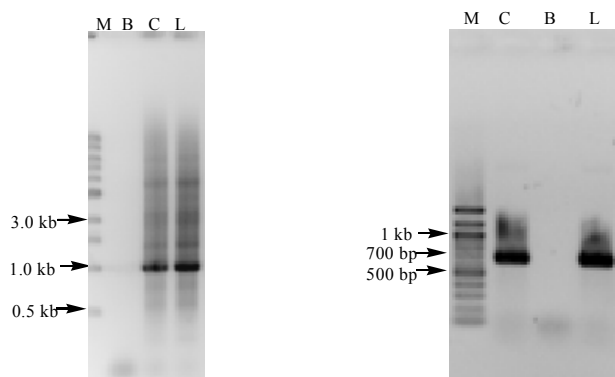
**Figure 2 - 23. Structure of three amino sugars**

### **2.2.2 Identification putative peptide and polyketide synthetases in *Lyngbya* sp. strain 15-2**

The presence of non-proteinogenic amino acids, including *D*-homophenylalanine, *D*-glutamine, Dhb and Athmu in the structure of pahayokolides A indicated that NRPSs were involved in the biosynthesis of pahayokolide A. Our feeding experiment indicated the Athmu moiety does not arise directly from intact acetate units assembled by PKSs. It is very possible that a PKS is not involved in the biosynthesis of pahayokolide A. To identify the NRPS and questioned PKS genes in *Lyngbya* sp. strain 15-2, degenerate PCR primers targeted to conserved sequence motifs of known NRPS and PKS genes were used with chromosomal DNA of *Lyngbya* sp. strain 15-2. The use of degenerate primers targeting the A3 and A8 core motifs of NRPSs and the  $\beta$ -ketosynthase (KS) domain of PKSs resulted in products of the expected size (1 kb and 0.7 kb, respectively) (Fig. 2-24). The 1 kb and 0.7 kb gene fragments amplified from *Lyngbya* sp. strain 15-2 were



subcloned in pCR 2.1 TOPO TA (Invitrogen) and sequenced. The resulting sequences were searched against NCBI's nonredundant database by using the BLASTX or BLASTP algorithms (Altschul et al., 1997).



**Figure 2 - 24. PCR of genomic DNA from *Lyngbya* sp. Strain 15-2 using degenerate NRPS and PKS primers.**

M, marker; B, blank control; C, positive control; L, *Lyngbya* DNA.

One of the 20 NRPS clones analyzed (pLN2) showed high homology to the phenylalanine-activating domain from *ociB* belonging to the NRPS complex responsible for oscillapeptin E production (gb|ABI26078.1|; 47% identity, 139 of 293 aligned amino acids) in *Planktothrix agardhii* NIVA-CYA 11. One of the 20 NRPS clones (pLN6) showed significant homology to a threonine-activating module from *mcnB* belonging to the NRPS complex responsible for cyanopeptolin production (gb|AAZ03551.1|; 74% identity, 228 of 307 aligned amino acids) in *Microcystis* sp. NIVA-CYA 172/5. One of the 20 NRPS clones (pLN8) showed significant homology to a serine-activating module from NosA encoded by a gene belonging to the NRPS complex responsible for nostopeptide A production (gb|AAF15891.2|AF204805\_1; 73% identity, 210 of 287 aligned amino acids) in *Nostoc* sp. GSV224. The deduced amino acid sequences of pLN2,

pLN6 and pLN8 were used to predict their substrate specificities by the method of Stachelhaus et al. (1999) and Challis et al. (2000). The primary sequences of the phenylalanine-activating A domain of gramicidin S and the translated sequences of pLN2, pLN6 and pLN8 were aligned by using Jotun-Hein method (Hein, 1994), and the signature amino acids conferring substrate recognition were compared to a database of known NRPS A domains (<http://www-ab.informatik.unituebingen.de/toolbox/index.php?view=domainpred>). The substrate specificities of pLN2, pLN6 and pLN8 were predicted

**Table 2 - 7. Activated substrates and the binding pocket amino acid residues of A domains in pLN2, pLN6 and pLN8**

	Binding pocket of A domain								Proposed substrates
	235	236	239	278	299	301	322	330	
pLN2	D	A	F	V	L	A	A	V	Phe
pLN6	D	F	W	N	I	G	M	V	Thr
pLN8	D	V	W	H	I	S	L	I	Ser

to activate the amino acid substrates phenylalanine (pLY2), threonine (pLY6), and serine (pLY2) (Table 2-7). Phenylalanine, threonine, and serine are present in the structure of pahayokolide A. Sequence analysis of the 20 PKS clones (pLP1-20) indicated that all the PKS PCR products contained a single, identical KS domain and showed high homology to the KS domain from *BarE* belonging to the PKS complex responsible for barbamide production ([gb|AAN32979.1](http://gb|AAN32979.1)); 72% identity, 181 of 251 aligned amino acids) in *Lyngbya majuscula*. The conserved active site motif (GPSVNVQTACSTSLVTVH) and catalytic cysteine residue for the KS domain were observed. The result conflicted with the logic rule that the long side chain of Athmu was assembled by three rounds of ketide

extensions if PKSs were involved. Therefore, it is not unreasonable to assume that Athmu may arise from the incorporation of a larger single unit, such as an amino sugar, rather than acetate units by PKSs.

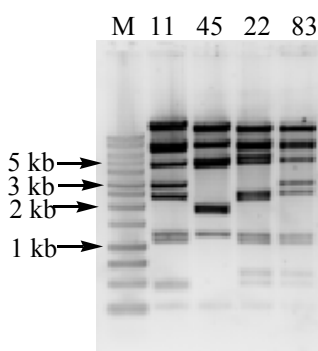
**Table 2 - 8. Partial sequencing of Fos11, 45, 22, 83**

Fosmid and modules	Proposed domains	Binding pocket of A domain	Proposed substrates
Fos11	C A <sub>Phe</sub> PCP	DAFVLA AV	phenylalanine
	C A <sub>Phe/Leu</sub> PCP	DAWLLGAV	Phenylalanine/leucine
	C A <sub>?</sub> PCP	DAFVLSSF	No prediction
Fos45	C A <sub>Phe/Leu</sub> PCP	DAWLLGAV	Phenylalanine/leucine
Fos22	C A <sub>Phe/Leu</sub> PCP	DAWLLGAV	Phenylalanine/leucine
Fos83	C A <sub>Phe/Leu</sub> PCP	DAWLLGAV	Phenylalanine/leucine

### 2.2.3 Cloning and partial characterization of the NRPS genes for pahayokolide A production in *Lyngbya* sp. strain 15-2

Chemical analysis of *Lyngbya* sp. Strain 15-2 has yielded only pahayokolide A, its cyclic fragment (pahayokolides B), and its conformers (pahayokolides C-D), as major metabolites (An et al., 2007). However, it remained possible that additional NRPS biosynthetic pathways are present in the genome whose products have yet to be characterized. Therefore, a strategy was devised to probe *Lyngbya* sp. strain 15-2 broadly for fosmids containing phenylalanine specific A domains of NRPS, and then focus on specific structural characteristics to identify the pathway of interest.

The fosmid library was constructed following the manufacture's protocol (Epicentre). About 12,760 colonies (60-fold coverage of the genome) were screened by colony PCR using the specific NRPS primers (Babcock et al., 2007). Four fosmids (Fos11, Fos45, Fos22, and Fos83) were sequenced partially and yielded three NRPS module products (Table 2-8). Fos11, the largest fosmid containing all the three NRPS modules, was chosen for full sequencing (Fig. 2-25).



**Figure 2 - 25. NotI and Bsu36I restriction map of four NRPS-positive fosmids of *Lyngbya* sp. strain 15-2.**

M, maker; 11, Fos11; 45, Fos45; 22, Fos22; 83, Fos83.

A contiguous 45 kb region was partially sequenced (37 kb) to reveal 14 putative opening reading frames (ORFs). The cluster was predicted to encode a NRPS assembly line responsible for biosynthesis, transport and degradation of a peptide in *Lyngbya* sp. strain 15-2. The 14 ORFs were transcribed in the same direction except ORF2, and showed a high degree of similarity to a number of bacterial NRPSs or associated proteins. Part of the Pah biosynthetase, including PahA, PahB, PahC and PahI, was analyzed in detail. As shown in Table 2-9, PahA is a monomodular NRPS harboring the first module and a C domain with the domain structure of A-PCP-C. The first module starts with an A domain, which suggested that it might be the initial module and activate one amino acid. PahB consisted of A-PCP-C-A-PCP, which was assumed to activate two amino acids.

The C domain of PahA became a module with the first two domains (A-PCP) of PahB. PahI was composed of three modules with the domain structure of (C-A-PCP)<sub>2</sub>-C-A-PCP-Te, which might be responsible for activating three amino acids, and releasing mature peptide during peptide biosynthesis. A fragment of 8 kb between PahC and PahI remained unsequenced.

**Table 2 - 9. Deduced functions of encoded proteins in the pah gene cluster of *Lyngbya* sp. strain 15-2**

Protein	Length (aa)	Proposed functions	Sequence similarity	I/S	Binding pocket of A domain	substrate	
						Proposed & identity	product
PahA	1019	A PCP C	Glu racemase [ <i>Bacillus thuringiensis</i> serovar <i>huazhongensis</i> BGSC 4BD1]	37%, 58%	DTWTVGAIEK	Arg 70%	D-Gln
PahB	1572						
PahB1		A PCP	NcpA [ <i>Nostoc</i> sp. ATCC53789]	45%, 64%	DACXLAAVCK	Tyr 70%	Pro
PahB2		C A PCP	Phe module of SubC [ <i>Bacillus subtilis</i> ]	47%, 64%	DAFVLAAVCK	Phe 80%	Phe
PahC							
PahC1		C A PCP	D-Phe module of BacC [ <i>Bacillus licheniformis</i> ]	47%, 64%	DAWVLAACK	Phe 80%	Gly
PahC2		C ?					
PahI							
PahI-4		? A PCP			????AVCK		
PahI-3		C A PCP	D-Phe module of BacC [ <i>Bacillus licheniformis</i> ]	49%, 67%	DAWVLAAVCK	Phe 80%	D-Homophe
PahI-2		C A PCP	NosD [ <i>Nostoc</i> sp. GSV224]	38%, 56%	DAXVLAAVCK	Tyr 70%	Pro
PahI-1		C A PCP TE	Phe module of SubC [ <i>Bacillus subtilis</i> ]	47%, 64%	DATKYGAVNK	Asp 80%	Athmu

aa, amino acid; I, identity; S, similarity.

In order to elucidate predicted substrates of A domains in PahA, PahB, and PahC PahI, primary sequences between core motifs A3 and A10 of A domains were aligned to corresponding sequence of PheA, and substrates of the binding-pockets were determined by their proximity to conserved sequences (Conti, et al., 1997; Stachelhaus, et al., 1999). As shown in Table 2-9 and Fig. 2-26, on the basis of known nonribosomal specificity-conferring code (Stachelhaus, et al., 1999), the specified substrates of the A domain in PahA were deduced to be Arg which was different from the glutamine in pahayokolide A. The binding pocket of A domains of glutamine and arginine should be similar because their hydrophobicity is close. The specificity of A domains of PahB-1 and PahI-2 was the tyrosine residue, which was discordant with the corresponding amino acid proline in the product respectively. The discordance can also be explained by their close hydrophobicity (Tyr, 1.3; Pro, -1.6). The specified substrate of the A domain of PahB-2 and PahI-3 was consistent with corresponding amino acid in the product. Most of the bacterial NRPS gene clusters are organized in a linear order that parallels the order of the amino acids in the resultant peptides. In spection of the *pah* gene cluster (Table 2-9) showed that the Pah modules were not organized according to the collinear rule for pahayokolide biosynthesis. One structural variation that deviated from the collinear rule in the *pah* gene cluster was that PahC-1 functioned before PahB-1. PahC-1 was supposed to activate glycine, but its specificity was phenylalanine. However, the binding-pockets of available A domains activating glycine showed weak homology (Challis et al., 2000). The predicted specificity of PahI-1 disagreed with the corresponding amino acid (Athmu) in pahayokolide A. The inharmony is reasonable because the database of nonribosomal specificity-conferring code does not include Athmu. The lack of PKS domains in PahI-1



ATCC 53789. The three conserved residues in the catalytic triad, serine84, aspartate111 and histidine201 were observed. The conserved proline29 was crucial for driving the mode of catalysis to cyclization instead of hydrolysis, and preferred producing cyclic products. It is consistent with the structure of pahayokolides. As the proline residue can shield the hydroxyl group of serine84 and affects the characteristics of the oxyanion hole. A glycine at this position will prefer a nucleophilic attack of a water molecule on the acyl-enzyme intermediate and form a linear product (Bruner et al., 2002).

Usually one typical NRPS module spans across about 3 kb. A 8 kb DNA fragment can only bear three NRPS modules. On the basis of the pahayokolide A structure and the deduced functions of the individual NRPS modules, the unsequenced fragment of 8 kb between PahC and PahI was predicted to encode four or five NRPS modules. These NRPS modules were proposed to assemble one *N*-acetyl-*N*-methyl-*L*-leucine, one threonine, one serine, and two Dhb residues into pahayokolide A. The module activating *N*-acetyl-*N*-methyl-*L*-leucine does not need to be in the 8 kb. Dhb arises from the dehydration of threonine. Only three NRPS modules are needed if one of these three NRPS modules activate two or three threonine iteratively. Therefore, the 8 kb of unsequenced DNA fragment is repetitive and offers a challenge for sequencing.

Antibiotic production genes have been found nearly in all cases to be clustered in one region of the bacterial chromosome, consisting of structural, resistance, and regulatory genes. The *pah* cluster also contains 10 ORFs that may be involved in pahayokolides resistance, transport and degradation (Table 2-10). The 2436-bp ORF10, which is located 42 bp downstream and transcribed in the same direction of *pahI*, encodes a putative 92.9-kDa protein with high homology to TonB-dependent receptor



plug. TonB-dependent outer membrane proteins take up molecules that are too large to pass through porins, using energy provided by the membrane-spanning protein TonB. Many siderophores are peptidic molecules that are synthesized by NRPS systems in Gram negative bacteria. Siderophores bind iron extracellularly, and then are taken up by

**Table 2 - 10. Deduced functions of the opening frame reading frames in the *pah* gene cluster**

Protein	Amino acids	Proposed function	Sequence similarity	Identity/ Similarity
ORF1	410	hypothetical protein ALPR1_04198	<i>Algoriphagus</i> sp. PR1	35%, 53%
ORF2	282	$\alpha/\beta$ hydrolase family protein	<i>Algoriphagus</i> sp. PR1	66%, 82%
ORF3	401	penicillin-binding protein, $\beta$ -lactamase class C	<i>Spirosoma linguale</i> DSM 74	40%, 56%
ORF4	369	efflux transporter, RND family, MFP subunit	<i>Microscilla marina</i> ATCC 23134	29%, 52%
ORF5	418	lipoprotein releasing system, permease component	<i>Algoriphagus</i> sp. PR1	78%, 90%
ORF6	226	ABC-type antimicrobial peptide transport system, ATPase component	<i>Spirosoma linguale</i> DSM 74	62%, 82%
ORF7	737	Signal Transduction Histidine Kinase	<i>Algoriphagus</i> sp. PR1	45%, 63%
ORF8	165	response regulator	<i>Algoriphagus</i> sp. PR1	67%, 84%
ORF9	183	two-component system response regulator protein	<i>Pedobacter</i> sp. BAL39	34%, 60%
ORF10	812	TonB-dependent receptor plug	<i>Dyadobacter fermentans</i> DSM 18053	36%, 57%

the cell through a TonB-dependent membrane transport system (Faraldo-Gomez et al., 2003). The 675-bp ORF6, which is located 5713 bp upstream and transcribed in the same direction of *pahA*, encodes a putative 25.7-kDa protein with high homology to ATP

banding cassette (ABC) transporters. It suggested that ORF6 may be involved in exporting pahayokolides from the cyanobacterium. Located 9382 bp downstream from *pahA* is ORF3 (1203 bp) with high homology to  $\beta$ -lactamases. The peptidase ORF3 may be involved in the hydrolysis of the peptide. It is supported by evidence that homologous MlrB (51%/31% similarity/identity) is involved in microcystin degradation in the bacterium *Sphingomonas linguale* DSM74.

## **2.3 Materials and Methods**

### **2.3.1 Culture conditions of *Lyngbya* sp. strain 15-2**

*Lyngbya* sp. strain 15-2 collected from the Florida Everglades was cultured in BG11 medium supplemented with 0.19 M Na<sub>2</sub>CO<sub>3</sub> buffered with 2-morpholinoethanesulfonic acid (MES) at pH 7.2 at 20~22 °C under continuous white light. After one month *Lyngbya* sp. strain 15-2 was fed with <sup>13</sup>C labeled substrates (50mg/tray each time) for three times on the first, fourth, and sixth days. Samples of biomass were harvested on the eighth day, dried with paper towels to remove medium, stored at -80 °C in an ultracold freezer.

### **2.3.2 Stable isotope incorporation experiments**

Samples of biomass were harvested from cultures of *Lyngbya* sp. strain 15-2 and lyophilized. The freeze-dried biomass was pulverized in liquid nitrogen, and extracted in 70% MeOH (in water) at room temperature for two days. Following solvent removal, water-soluble constituents were separated by RP C<sub>18</sub> solid-phase extraction using a stepwise gradient solvent system from 20% MeOH in water to 100% MeOH (Berry et al., 2004). The bioactive 80% and 100% MeOH fractions were then separately chromatographed on prep-HPLC (Alltech Alltima C<sub>18</sub>, 10 μm, 250 × 22 mm) using 75% MeOH in H<sub>2</sub>O and monitored at 220 nm. Pahayokolides A-D were obtained. The purity of pahayokolides was checked by analytical HPLC (Alltech Apollo C<sub>18</sub>, 5 μm, 250 mm × 4.6 mm) using 70% MeOH in H<sub>2</sub>O and monitored at 220 and 254 nm.

### **2.3.3 NMR experiment**

NMR spectroscopic data were acquired in DMSO-d<sub>6</sub>-D<sub>2</sub>O (3:7) on a Bruker 400 or 600 MHz AVANCE spectrometers. The NMR data were processed with Topspin/XWIN-NMR, and analyzed using Sparky software (An et al., 2007).

### **2.3.4 DNA manipulations**

High molecular weight (HMW) genomic DNA for the construction of *Lyngbya* sp. strain 15-2 genomic library was prepared from flash frozen, pulverized filaments of *Lyngbya* sp. strain 15-2 using the combination method of phenol/chloroform extraction

and Genomic DNA Purification Kit (Qiagen). DNA from *Lyngbya* sp. strain 15-2 for PCR amplification was obtained using FastDNA SPIN for Soil Kit (Bio 101 Systems). For PCR amplification of A domains of NRPSs and KS domain of PKSs, degenerate primers were used (Table 2-11). Plasmid preparation was carried out by using commercial kits (Qiagen).

**Table 2 - 11. primer sequences**

primer	Comments	Sequence (5' to 3')
NRPS2upper	Degenerate NRPS forward	GCN GCN GGN GCI TAY GTI CC
NRPS2lower	Degenerate NRPS reverse	CCN CKD ATY TTN ACY TG
KS4U	Degenerate PKS forward	MGI GAR GCI YTI CAR ATG GAY CCI CAR CAR MG
KS5L	Degenerate PKS reverse	GGR TCN CCI ARY TGI GTI CCI GTI CCR TGI GC
PheLN02f	Phenylalanine specific NRPS forward	TCA AAC CCC GTA CCT CAA AG
PheLN02r	Phenylalanine specific NRPS reverse	CCG ATG GTG ATT TGG CTA GT
ThrLN06f	Threonine specific NRPS forward	ACC CAA CCT ATC CCC AAG AC
ThrLN06r	Threonine specific NRPS reverse	TAA ATC GCC TTC TGT TTC GG
SerLN08f	Sernine specific NRPS forward	CGA ACA TAA CGC CCA AAT CT
SerLN08r	Sernine specific NRPS reverse	GAA GGC ACA AAA TGC AAG GT

### 2.3.5 Construction of the fosmid library

The *Lyngbya* sp. strain 15-2 fosmid library was constructed using the CopyControl Fosmid library production Kit (Epicentre) following the manufacturer's protocol. The only exception was that the HWM genomic DNA was not sheared before being end-

repaired. About 12760 colonies (~ 60-fold coverage of the genome) were obtained. The average size of insert DNA was 40 kb.

### 2.3.6 Screening of the fosmid library

The fosmid library was split into 12 pools (Lyn FG1 to Lyn FG12) and screened by PCR with NRPS specific primers. Among the six NRPS positive pools, Lyn FG1 were subdivided further into 10 subpools and screened by PCR with NRPS and specific primers (Figure 2-27). Ninety-six single colonies were picked up from each NRPS

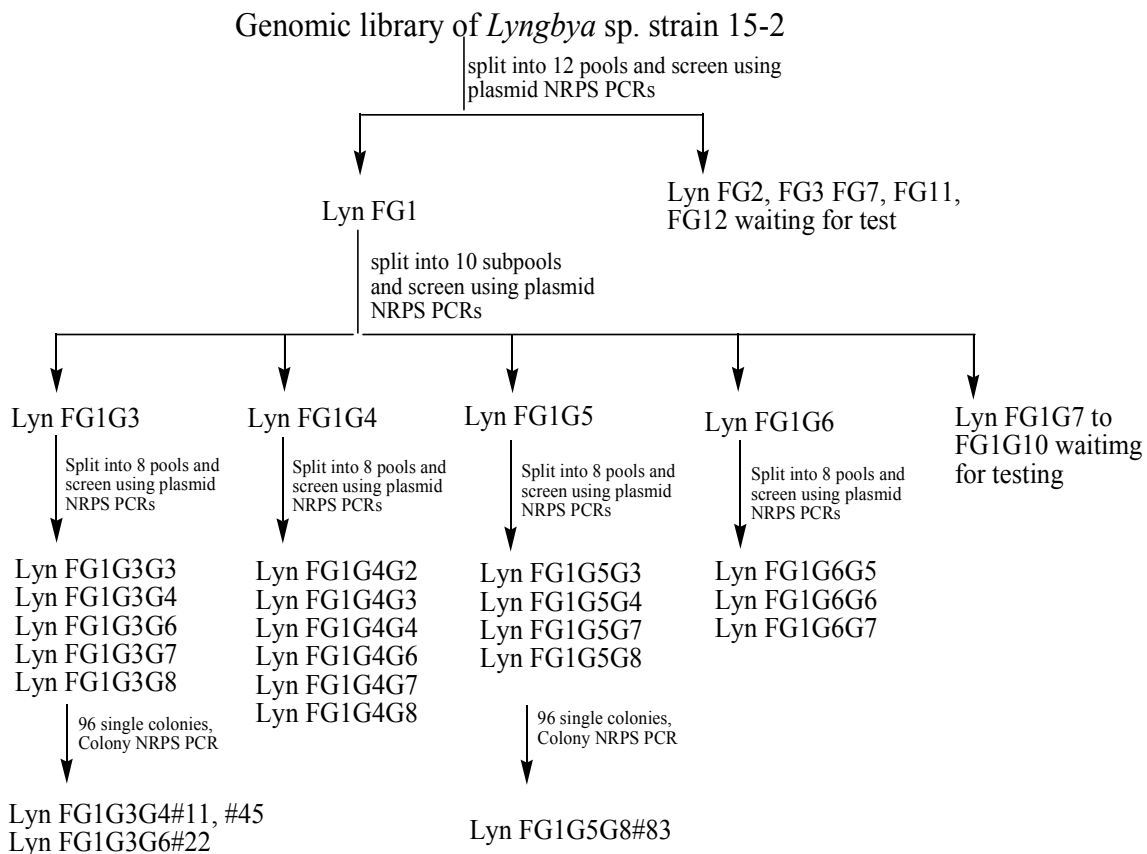
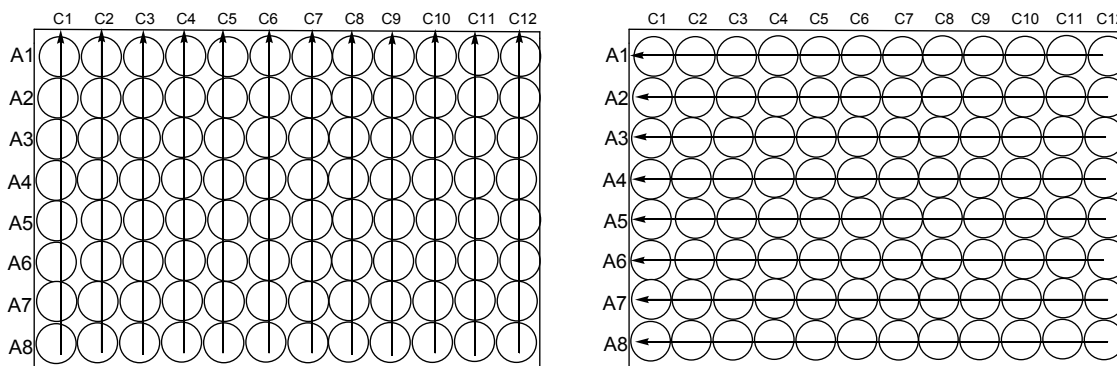


Figure 2 - 27. Screening of the genomic library of *Lyngbya* sp. strain 15-2

positive subgroup of Lyn FG1G1 to Lyn FG1G6. Each group of 96 single colonies was cultured and combined across 8 rows and down 12 columns. Colony PCR was then performed on the 20 pools of cells with NRPS specific primers (Fig. 2-28). Finally, four NRPS positive fosmids (Lyn FG1G3G4#11, #45, Lyn FG1G3G6#22 and Lyn FG1G5G8#83) were sent to Eurofins MWG Operon for DNA sequencing.



**Figure 2 - 28. Preparation of 20 pools of cells for the genomic library screening.**  
 Left, 12 pools from combining across 8 rows; right, 8 pools from combining down 12 columns

### 2.3.7 Sequence analysis

Database searches for genes and proteins for sequence homology with the putative NRPS and PKS were performed by using the NCBI (National Center for Biology Information) BLAST service.

## 2.4 Conclusions

The biosynthesis of the Athmu sidechain arises from leucine or  $\alpha$ -ketoisovalerate.

Our prediction that this subunit is further extended by three rounds of condensation with acetate was based on the biosynthetic pathway for the Adda sidechain of microcystin-LR. However, stable isotope incorporation experiments using both [1-<sup>13</sup>C] acetate and [1,2-<sup>13</sup>C] acetate demonstrated that this was not correct. Incorporation of the acetate label at other positions which may be predicted from primary metabolic pathways serves as an internal control experiment, and validates the lack of incorporation into the Athmu moiety.

Putative peptide and polyketide synthetases in *Lyngbya* sp. strain 15-2 were identified by using a PCR method and degenerate primers derived from conserved core sequences of known NRPSs and PKSs. The subcloned NRPS fragments (pLY) obtained by degenerate PCR showed a high level of homology to the phenylalanine (pLY2), threonine (pLY6) and serine-activating (pLY8) A domains. Substrate-binding pocket analysis suggested that the deduced A domains of pLY2, pLY6 and pLY8 indeed activates phenylalanine, threonine and serine, respectively. All the subcloned PKS fragments (pLP) obtained by degenerate PCR showed a high level of homology to the KS domain from *BarE*. Identification of one unique KS domain conflicted with the logic rule that the long side chain of Athmu was assembled by three rounds of ketide extensions if PKSs were involved. However, it does not rule out the possibility of incorporation of an unusual subunit through a Claisen type condensation similar to that used for the construction of polyketides.

The genomic library of *Lyngbya* sp. strain 15-2 was constructed and screened by colony PCR. Partial sequence showed that the genetic architecture of the *pah* biosynthetic system is comprised of thirteen genes in two transcriptional units. Besides ten genes

responsible for peptide resistance, transport and degradation, six NRPS modules were identified. Although the predicted substrate specificities of NRPS did not agree with the structure of pahayokolide A, the disagreement could be explained. However, no PKS gene was found in the *pah* gene cluster.

## **Significance and Future Work**

*Lyngbya* continues to be a source of novel natural products, many of which are bioactive. Pahayokolide A represents a bioactive branched-cyclic polypeptide possessing a usual  $\beta$ -amino acid and *D*-amino acids. Pahayokolide A shows a spectrum of biological activities consistent with their function as a defense metabolite, including cytotoxicity, ichthyotoxicity and antimicrobial activity. This study was successful in determining the absolute configuration of pahayokolides A-D using the advance Marfey's method and figuring out the biosynthetic origin of the Athmu moiety using a classic precursor feeding and NMR detection approach. Great insight was achieved by identifying putative peptide and polyketide synthetases in *Lyngbya* sp. strain 15-2 using PCR. Based on the structural features of pahayokolide A and biosynthetic finding, a set of gene probes was constructed, which led to the efficient identification of a fosmid containing portion of the pahayokolide A gene cluster from a total genomic library. Sequencing of the fosmid and assembly of the pathway revealed the *pah* genes encoding the enzymes catalyzing the biosynthesis of pahayokolide A. The future studies are required to complete the sequencing of 8 kb repetitive fragment in Fosmid11. To determine whether the clone peptide synthetase gene *pah* is required for pahayokolide synthesis, gene disruption of



*pah* in the genome of *Lyngbya* sp. strain 15-2 should be performed by homologous recombination. However, nobody has done homologous recombination in a *Lyngbya*, which must wait for the development of the appropriate molecular biological methods. This report broadens new peptide synthetases waiting for discovery from *Lyngbya* species. One potential application of *Lyngbya* NRPS is to use particular domains and enzymes for combinatorial biochemistry. This facilitate future attempt to produce novel bioactive peptides with improved therapeutic efficacy and lower toxicity.

## References

- Altschul, S. F.; Madden, T. L.; Schäffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. *J. Nucleic Acids Res.* **1997**, *25*, 3389–3402.
- An, T.; Krishnaswamy, T.; Kumar, S.; Wang, M.; Liu, L.; Lay, Jr. Jo.; Liyanage, R.; Berry, J.; Gantar, M.; Marks, V.; Gawley, R. E.; Rein, K. S. *J. Nat. Prod.* **2007**, *70*(5), 730-735.
- An, T.; Krishnaswamy, T.; Kumar, S.; Wang, M.; Liu, L.; Lay, Jr. Jo.; Liyanage, R.; Berry, J.; Gantar, M.; Marks, V.; Gawley, R. E.; Rein, K. S. *J. Nat. Prod.* **2007**, *70*(5), 730-735.
- Annala, A.; Lehtimäki, J.; Mattila, K.; Eriksson, J. E.; Sivonen, K.; Rantala, T. T.; Baldwin, T.O. *Structure* **1996**, *4*, 223-228.
- Babcock, D. A.; Wawrik, B.; Paul, J. H.; McGuinness L.; Kerkhof L. J. *J. Microbiol. Methods*, **2007**, *71*, 156-161.
- Barrios-Ilerena M. E.; Burja, A. M.; Wright, P. C. *J. Ind. Microbiol. Biotechnol.* **2007**, *34*, 443-456.
- Becker J. E.; Moore, R. E.; Moore, B. S. *Gene* **2004**, *325*, 35-42.
- Berry, J. P., Gantar, M., Gawley, R. E., Wang, M., Rein, K. S. *Comp. Biochem. Physiol.* **2004**, *139*, 231-238.
- Billich, A.; Zocher, R. *Biochemistry* **1987**, *26*, 8417-8423.
- Bonnard, I.; Rolland, M.; Francisco, C.; Banaigs, B. *Lett. Pept. Sci.* **1997**, *4*, 289-292.
- Bourne, D. G.; Riddles, P.; Jones, G. L.; Smith, W.; Blakeley, R. L. *Environ. Toxicol.* **2001**, *16*, 523-534.
- Bruner, S. D.; Weber, T.; Kohli, R. M.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T.; Stubbs, M. T. *Structure* **2002**, *10*, 301-310.
- Carmichael, W. W. *Adv. Biochem. Res.* **1997**, *27*, 211-256.
- Carmichael, W. W. *Sci. Am.* **1994**, *270*(1), 78-86.
- Carmichael, W. W. *Sci. Am.* **1994**, *270*, 64-72.
- Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211–224.
- Challis, G. L.; Ravel, J.; Townsend, C. A. *Chem. Biol.* **2000**, *7*, 211–224.

- Conti, E; Franks, N. P.; Brick, P. *Structure* **1996**, 4, 287-298.
- Cosmina, P.; Rodriguez, F.; de Ferra, F.; Oerego, M.; Venema, G.; van Sinderen, D. *Mol. Microbiol.* **1993**, 8, 821-831.
- Dahlmann, J.; Budakowskib, W. R.; Luckas, B. J. *Chromatogr. A* 2003, 994, 45-47.
- Daranas, A. H.; Norte, M.; Fernandez, J. J. *Toxicon* 2001, 39, 1101-1132.
- Davies-Coleman, M. T.; Dzeha, T. M.; Gray, C. A.; Hess, S.; Pannell, L. K.; Hendricks, D. T.; Arendse, C. E. *J. Nat. Prod.* 2003, 66(5): 712-715.
- De Crécy-Lagard, V.; Marlière, P.; Saurin, W. *Acad Sci. III* **1995**, 318, 927-936.
- Döhren, H.V.; Keller, U.; Vater, J.; Zocher, R. *Chem. Rev.* **1997**, 97, 2675–2705.
- Drakenberg, T. *J. Biol. Chem.* **1996**, 271, 16695-16702.
- Du, L.; Sánchez, C.; Shen, B. *Metab. Eng.* **2001**, 3, 78-95.
- Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. *Chem. Biol.* 2004, 11(6), 743-745.
- Falconer, I. R. *Environ. Toxicol.* 1999, 14, 5-12.
- Faraldo-Gomez, J. D.; Sansom, M. S. *Nat Rev Mol Cell Biol* **2003**, 4, 105–116.
- Fernandez-Suarez, M.; Muñoz, L.; Fernandez, R.; Riguera, R. *Tetrahedron: Asymmetry* 1997, 8, 1847-1854.
- Frankmölle, W. P.; Knübel, G.; Moore, R. E.; Patterson, G. M. L. *J. Antibiot.* 1992a, 45, 1458-1466.
- Frankmölle, W. P.; Larsen, L. K.; Caplan, F. R.; Patterson, G. M. L.; Knübel, G.; Levine, I. A.; Moore, R. E. *J. Antibiot.* 1992b, 45, 1451-1457.
- Fujii, K.; Ikai, Y.; Oka, H.; Suzuki, M.; Harada, K. *Anal. Chem.* 1997, 69, 5146-5151.
- Fujii, K.; Shimoya, T.; Ikai, Y.; Oka, H.; Harada, K. *Tetrahedron Lett.* 1998, 39, 2579-2582.
- Fujii, K.; Sivonen, K.; Kashiwagi, T.; Hirayama, K.; Harada K. *J. Org. Chem.* 1999, 4, 5777-5782.
- Fujii, K.; Sivonen, K.; Kashiwagi, T.; Hirayama, K.; Harada, K. *J. Org. Chem.* **1999**, 64, 5777-5782.
- Gerwick W.H., Roberts M.A., Proteau P.J., Chen J.L. *J. Appl. Phycol.*, 1994, 6, 143-149.

- Gerwick W.H., Tan L.T., Siachitta N. Nitrogen-containing metabolites from marine cyanobacteria. In *The Alkaloids*, vol 57. Edited by Cordell G. San Diego: Academic Press, 2001, 75-184.
- Gerwick, W. H.; Jiang, Z. D.; Agarwal, S. K.; Farmer, B. T. *Tetrahedron* 1992, 48, 2313-2324.
- Gokhale, R. S.; Tuteja, D. Biochemistry of polyketide synthases. In: *Biotechnology* (Eds. Reid and Rehm), WILEY-VCH Verlag, GmbH, Germany, **2001**, 10, 342-372.
- Gunasekera, S. P.; Ritson-Williams, R.; Paul, V. J. *J. Nat. Prod.* 2008b, 71(12), 2060-2063.
- Gunasekera, S. P.; Ross, C.; Paul, V. J.; Matthew, S.; Luesch, H. *J. Nat. Prod.* 2008a, 71(5), 887-890.
- Gutiérrez, M.; Suyama, T. L.; Engene, N.; Wingerd, J. S.; Matainaho, T. and Gerwick, W. H. *J. Nat. Prod.* 2008, 71 (6), 1099–1103.
- Haese, A.; Pieper, R.; von Ostrowski, T.; Zocher, R. *J. Mol. Biol.* **1994**, 243, 116-122.
- Han, B., McPhail, K.L., Ligresti, A., Di Marzo, V., Gerwick, W.H. *J. Nat. Prod.* 2003, 66, 1364–1368.
- Han, B.; Goeger, D.; Maier, C. S.; Gerwick, W. H. *J. Org. Chem.*, 2005, 70(8), 3133-3139.
- Han, B.; Gross, H.; Goeger, D. E.; Mooberry, S. L.; Gerwick, W. H. *J. Nat. Prod.* 2006, 69(4), 572-575.
- Harrigan, G. G.; Luesch, H.; Moore, R. E.; Paul, V.J. *Spec. Publ.-R.Soc. Chem.* 2000, 257, 126-139.
- Harrigan, G. G.; Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Nagle, D. G.; Paul, V. J. *J. Nat. Prod.* 1999, 62, 655-658.
- Hawkins, C. J.; Lavin, M. F.; Marshall, K. A.; van den Brenk, A. L.; Watters, D. J. *J. Med. Chem.* 1990, 33, 1634-1638.
- Hein, J. *J. Theor. Biol.* **1994**, 167, 169-174.
- Helms, G. L.; Moore, R. E.; Niemczura, W. P.; Patterson, G. M. L. *J. Org. Chem.* 1988, 53, 1296-1307.
- Helms, G. L.; Moore, R. E.; Niemczura, W. P.; Patterson, G. M. L.; Tomer, K. B.; Gross, M. L. *J. Org. Chem.* **1988**, 53, 1298-1307.

- Hoffman, D.; Hevel, J. M.; Moore, R. E.; Moore, B. S. *Gene* 2003, 311, 169-178.
- Hoffmann, K.; Schneider-Scherzer, E.; Kleinkauf, H.; Zocher, R. *J. Biol. Chem.* **1994**, **269**, 12710-12714.
- Hofheinz, W.; Oberhänsli, W. E. *Helv. Chim. Acta* 1977, 60, 660–669.
- Jiménez, J. I.; Scheuer, P.J. *J. Nat. Prod.* 2001, 64(2), 200-203.
- Jochimsen, E. M.; Carmichael, W. W.; An, J. S.; Cardo, D. M.; Cookson, S. T.; Holmes, C. E.; Antunes, M. B.; de Melo Filho, D. A.; Lyra, T. M.; Barreto, V. S.; Azevedo, S. M.; Jarvis W. R. N. *Engl. J. Med.* 1998, 338, 873-878.
- Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. *Nat. Struct. Biol.* **2002**, 9, 522-526.
- Keating, T. A.; Miller, D. A.; Walsh, C. T. *Biochemistry*, **2000**, 4729-4739.
- Kiyonaga, F., Yoshitomo, I., Hisao, O., Makoto, S. & Ken-ichi, H. *Anal. Chem.* 1997, 69, 5146-5151.
- Kleinkauf, H.; Döhren, H. V. *Eur. J. Biochem.* **1996**, 236, 335–351.
- Kohli, R. M.; Trauger, J. W.; Schwarzer, D.; Marahiel, M. A.; Walsh, C. T. *Biochemistry* **2001**, 40, 7099-7108.
- Konz, D.; Marahiel, M. A. *Chem. Biol.* **1999**, 6, 39-48.
- Kwan, J. C.; Eksioğlu, E. A.; Liu, C. Paul, V. J.; Luesch, H. *J. Med. Chem.* 2009, 52, 5732-5747.
- Kwan, J. C.; Rocca, J. R.; Abboud, K. A.; Paul, V, J.; Luesch, H. *Org. Lett.*, 2008, 10(5), 789-792.
- Linington, R. G.; Edwards, D. J.; Shuman, C. F.; McPhail, K. L.; Matainaho, T.; Gerwick, W. H. *J. Nat. Prod.* 2008, 71, 22-27.
- Linne, U.; Marahiel, M. A. *Methods Enzymol.* **2004**, 388, 293–315.
- Lipmann, F.; Gevers, W.; Kleinkauf, H.; Roskoski, R. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1971**, 35, 1-34.
- Luesch, H.; Pangilinan, R.; Yoshida, W.Y.; Moore, R. E.; Paul, V. J. *J. Nat. Prod.* 2001b, 64(3), 304-307.
- Luesch, H.; Williams, P. G.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. *J. Nat. Prod.* 2002a, 65(7), 996-1000.

- Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. *Bioorg. Med. Chem.* 2002c, 10(6), 1973-1978.
- Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. *J. Nat. Prod.*, **2000b**, 63(10), 1437-1439.
- Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J. *Tetrahedron* 2002b, 58(39), 7959-7966.
- Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Corbett, T. H. *J. Am. Chem. Soc.* 2001a, 123, 5418-5423.
- Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Paul, V. J.; Mooberry, S. L. *J. Nat. Prod.* 2000a, 63(5), 611-615.
- MacMillan, J. B.; Ernst-Russell, M. A.; de Ropp, J. S.; Molinski, T. F. *J. Org. Chem.* 2002, 67(23), 8210-8215.
- Magarvey, N. A.; Beck, Z. Q.; Golakoti, T.; Ding, Y.; Huber, U.; Hemscheidt, T. K.; Abelson, D.; Moore, R. E.; Sherman, D. H. *ACS Chem. Biol.* 2006, 1, 766-779.
- Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem Rev.* **1997**, 97(7), 2651-2673.
- Marquez, B. L.; Watts, K. S.; Yokochi, A.; Roberts, M. A.; Verdier-Pinard, P.; Jimenez, J. I.; Hamel, E.; Scheuer, P. J. and Gerwick W. H. *J. Nat. Prod.* 2002, 65 (6), 866-871.
- Marshall, C. G.; Hillson, N. J.; Walsh, C. T. *Biochemistry*, **2002**, 41, 244-250.
- Matern, U.; Schleberger, C.; Jelakovic, S.; Weckesser, J.; Schulz, E. G. *Chem. Biol.* 2003, 10, 997-1001.
- Matsumori, N.; Kaneno, D.; Murata, M.; Nakamura, H.; Tachibana, K. *J. Org. Chem.* 1999, 64, 866-876.
- Matthew, S.; Paul, V. J.; Luesch, H. *Phytochemistry* 2009a, Oct 6 [Article in press].
- Matthew, S.; Paul, V. J.; Luesch, H. *Planta Med.* 2009b, 75, 528-533.
- Matthew, S.; Ross, C.; Paul, V. J.; Luesch, H. *Tetrahedron* 2008a, 64, 4081-4089.
- Matthew, S.; Ross, C.; Rocca, J. R.; Paul, V. J.; and Luesch, H. *J. Nat. Prod.* 2007, 70 (1), 124-127.
- Matthew, S.; Schupp, P. J.; Luesch, H. *J. Nat. Prod.* 2008b, 71(6), 1113-1116.
- Mayer M.S. and Gustafson K.R. *Int. J. Cancer* 2003, 105, 291-299.

- McPhail, K. L.; Correa, J.; Linington, R. G.; Gonzalez, J.; Ortega-Barría, E.; Capson, T. L.; Gerwick, W. H. *J. Nat. Prod.* 2007, 70(6), 984-988.
- Mehner, C.; Mueller, D.; Krick, A.; Kehraus, S.; Löeser, R.; Güetschow, M.; Maier, A.; Fiebig, H.-F.; Brun, R.; Köenig, G. M. *Eur. J. Org. Chem.* 2008, 10, 1732-1739.
- Mehner, C.; Mueller, D.; Krick, A.; Kehraus, S.; Löeser, R.; Güetschow, M.; Maier, A.; Fiebig, H.-F.; Brun, R.; Köenig, G. M. *Eur. J. Org. Chem.* **2008**, 10, 1732-1739.
- Miller, J. C.; Miller, L. D.; Olson, C.; Gillette, K. J. *Natl. Cancer Inst.* 1969, 43(6), 1297-1305.
- Moore, R. E. *J. Ind. Microbiol.* 1996, 16, 134-143.
- Moore, R. E.; Chen, J. L.; Moore, B. S.; Patterson, G. M. L.; Carmichael, W. J. *Am. Chem. Soc.* **1991**, 113, 5083-5084.
- Mootz, H. D.; Schwarzer, D.; Marahiel, M. A. *ChemBioChem* **2002**, 3, 490-504.
- Murakami, M.; Suzuki, S.; Itou, Y.; Kodani, S.; Ishida, K. *J. Nat. Prod.* 2000, 63, 1280-1282.
- Nakao, Y.; Yoshida, W. Y.; Szabo, C. M.; Baker, B. J.; Scheuer, P. J. *J. Org. Chem.* 1998, 63, 3272-3280.
- Namikoshi, M.; Rinehart, K. L. *J. Ind. Microbiol.* 1996, 17, 373-384.
- Nogle, L. M.; Gerwick, W. H. *J. Nat. Prod.* 2002, 65(1), 21-24.
- Nogle, L. M.; Marquez, B. L.; Gerwick, W. H. *Org. Lett.* 2003, 5(1), 3-6.
- Nogle, L. M.; Okino, T.; Gerwick, W. H. *J. Nat. Prod.* 2001a, 64(7), 983-985.
- Nogle, L. M.; Williamson, R. T.; Gerwick, W. H. *J. Nat. Prod.* 2001b, 64, 716-719.
- Onest, S.; Miller, A. D.; Brick, P. *Structure* **1995**, 3, 163-176.
- Patočka, J. *Acta Medica* 2001, 44 (2), 69-75.
- Pergament, I.; Carmeli, S. *Tetrahedron Lett.* 1994, 35, 8473-8476.
- Pettit, G. R.; Kamano, Y.; Herald, C. L.; Dufresne, C.; Cerny, R. L.; Herald, D. L.; Schmidt, J. M.; Kizu, H. *J. Am. Chem. Soc.* 1989, 111, 5015-5017.
- Pfeifer, E.; Pavela-Vrancic, M.; Döhren, H.; Kleinkauf, H. *Biochemistry* **1995**, 34, 7450-7459.
- Pieper, R.; Haese, A.; Schröder, W.; Zocher, R. *Eur. J. Biochem* **1995** 230, 119-126.

- Phillipson, D. W.; Milgram, K. E.; Yanovsky, A. I.; Rusnak, L. S.; Haggerty, D. A.; Farrell, W. P.; Greig, M. J.; Xiong, X.; Proefke, M. L. *J. Comb. Chem.* **2002**, 4(6), 591-599.
- Ramaswamy, A. V.; Sorrels, C. M.; Gerwick, W.H. *J Nat Prod.* **2007**, 70(12), 1977-1986.
- Rinehart, K. L.; Sakai, R.; Kishore, V.; Sullins, D. W.; Li, K. M. *J. Org. Chem.* 1992, 57, 3007-3013.
- Schwarzer, D.; Finking, R.; Marahiel, M. A. *Nat. Prod. Rep.* **2003**, 20, 275–287.
- Shaw-Reid, C. A.; Kelleher, N. L.; Losey, H. C.; Gehring, A. M.; Berg, C.; Walsh, C. T. *Chem Biol.* **1999**, 6(6), 385-400.
- Shuler, M. L. *Ann. NY Acad. Sci.* **1994**, 745, 455–461.
- Sieber, S. A.; Marahiel, M. A. *Chem. Rev.* **2005**, 105, 715-738.
- Simmons, T. L.; Coates, R. C.; Clark, B. R.; Engene, N.; Gonzalez, D.; Esquenazi, E.; Dorrestein, P. C.; Gerwick, W. H. *Proc. Natl. Acad. Sci. U.S.A.* 2008, 105, 4587-4594.
- Simmons, T. L.; Nogle, L. M.; Media, J.; Valeriote, F. A.; Mooberry, S. L.; Gerwick, W. H. *J. Nat. Prod.* 2009, 72(6), 1011-1016.
- Singh, E, K; Sellers, R, P; Alexander, L, D; McAlpine, S, R. *Curr Opin Drug Discov Devel.* **2008**, 11, 544-552.
- Sitachitta, N.; Williamson, R. T.; Gerwick, W. H. *J. Nat. Prod.* 2000, 63, 197-200.
- Sone, H.; Kondo, T.; Kiryu, M.; Ishiwata, H.; Ojika, M.; Yamada, K. *J. Org. Chem.* 1996, 60, 4774-4781.
- Stachelhaus T.; Huser A.; Marahiel M. A. *Chem. Biol.*, **1996**, 3, 913–921.
- Stachelhaus, T.; Marahiel, M. A. *FEMS Micro. Lett.*, **1995a**, 125, 3-14.
- Stachelhaus, T.; Marahiel, M. A. *J. Biol. Chem.*, **1995b**, 270, 6163-6169.
- Stachelhaus, T.; Mootz, H. D.; Bergendahl V.; Marahiel M. A. *J. Biol. Chem.*, **1998**, 273, 22773–22781.
- Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, 6, 493–505.
- Stachelhaus, T.; Walsh, C. T. *Biochemistry* **2000**, 39, 5775-5787.
- Stanley, S. M. *Earth system history* (Edition: 2), Published by Macmillan, 2004, P263.



- Stephan A. S.; Marahiel, M. A. *J. Bacteriol.* **2003**, 185(24), 7036–7043.
- Tan, L. T. *Phytochemistry* 2007, 68, 954-979.
- Tan, L. T.; Chang, Y. Y.; Ashootosh, T. *Phytochemistry* 2008, 69(10), 2067-2069.
- Tan, L. T.; Sitachitta, N. and Gerwick, W. H. *J. Nat. Prod.*, 2003, 66 (6), 764-771.
- Taori, K.; Matthew, S.; Rocca, J. R.; Paul V. J.; Luesch, H. *J. Nat. Prod.* 2007, 70(10), 1593-1600.
- Taori, K.; Paul, V. J.; Luesch, H. *J. Nat. Prod.* 2008, 71(9), 1625-1629.
- Teruya, T.; Sasaki, H.; Fukazawa, H.; Suenaga, K. *Org. Lett.* 2009, 11, 2421-2424.
- Tillet, D.; Dittmann, E.; Erhard, M.; von Döhren, H.; Börner, T.; Neilan, B. A. *Chem. Biol.* 2000, 7, 753-764.
- Tillett, D., Dittmann, E., Erhard, M.; von Döhren, H.; Börner, T.; Neilan, B. A. *Chem. Biol.* **2000**, 7, 753-764.
- Trimurtulu, G.; Ohtani, I.; Patterson, G. M. L.; Moore, R. E.; Corbett, T. H.; Valeriote, F. A.; Dechik, L. J. *Am. Chem. Soc.* 1994, 116, 4729-4737.
- Tripathi, A.; Puddick, J.; Prinsep, M. R.; Lee, P. P. F.; Tan, L. T. *J. Nat. Prod.*, 2009, 72 (1), 29-32.
- Tseng, C. C.; Bruner, S. D.; Kohli, R. M.; Marahiel, M. A.; Walsh, C. T.; Sieber, S. A. *Biochemistry* **2002**, 41, 13350-13359.
- Turgay, K.; Krause, M.; Marahiel, M. A. *Mol. Microbiol.* **1992**, 6, 2743–2744.
- Walsh, C. T.; Chen, H.; Keating, T. A.; Hubbard, B. K.; Losey, H. C.; Luo, L.; Marshall, C. G. ; Miller D. A. and Patel, H. M. *Curr. Opin. Chem. Biol.* **2001**, 5, 525-534.
- Wan, F.; Erickson, K. L. *J. Nat. Prod.* 2001, 64, 143-146.
- Watts, R. E.; Tse, M. L.; Khosla, C. *Biochemistry* 2007, 46, 3385–3393.
- Weber, G.; Schörgendorfer, K.; Schneider-Scherzer, E.; Leitner, E. *Curr. Genet.* **1994**, 26, 120-125.
- Weber, T.; Baumgartner, R.; Renner, C.; Marahiel, M.; Holak, T. A. *Structure* **2000**, 8, 407-418.
- Weber, T.; Marahiel, M. *Structure* **2001**, 9, R3-9.
- Welker, M.; von Döhren, H. *FEMS Microbiol. Rev.* 2006, 30, 530-563.

- Westaway, E.G. *Am. J. Epidemiol.* 1966, 84, 439-456.
- Williams, P. G.; Luesch, H.; Yoshida, W. Y.; Moore, R. E. and Paul, V. J. *J. Nat. Prod.* 2003c, 66 (5), 595-598.
- Williams, P. G.; Moore, R. E.; Paul, V. J. *J. Nat. Prod.* 2003a, 66(10), 1356-1363.
- Williams, P. G.; Yoshida, W. Y.; Moore, E. R. and Paul, V. J. *J. Nat. Prod.* 2002, 65 (1), 29-31.
- Williams, P. G.; Yoshida, W. Y.; Quon, M. K.; Moore, R. E.; Paul, V. J. *J. Nat. Prod.* 2003b, 66(12), 1545-1549.
- Williams, P. G.; Yoshida, W. Y.; Quon, M. K.; Moore, R. E.; Paul, V. J. *J. Nat. Prod.* 66(5), 2003d, 651-654.
- Williamson, R. T.; Sitachitta, N.; Gerwick, W. H. *Tetra. Lett.* **1999**, 48 (2), 5175-5178.
- Wipf, P.; Fritch, P. C.; Geib, S. J.; Seffler, A. M. *J. Am. Chem. Soc.* 1998, 120, 4105-4112.
- Zainuddin, N. E.; Jansen, R.; Nimtz, M.; Wray, V.; Preisitsch, M.; Lalk, M.; Mundt, S. *J. Nat. Prod.* 2009, 72 (8), 1373-1378.
- Zainuddin, N. E.; Jansen, R.; Nimtz, M.; Wray, V.; Preisitsch, M.; Lalk, M.; Mundt, S. *J. Nat. Prod.* **2009**, 72 (8), 1373-1378.

## VITA

Li Liu

### EDUCATION

- 2004-2009      Doctoral Candidate in Chemistry and Biochemistry  
Florida International University, Miami, Florida
- 1998-2001      Master of Science in Marine Chemistry  
Institute of Oceanology, Chinese Academy of Science, P. R. China
- 1994-1998      Bachelor of Science in Chemistry  
Shandong Normal University, Ji'nan, Shandong, P.R. China

### ACADEMIC HONORS AND AWARDS

- 2008-2009      Dissertation Year Fellowship, Florida International University
- 2006              Presidential Fellowship, Florida International University
- 2000-2001      Liu-Yongling Fellowships, Chinese Academy of Sciences

### PUBLICATIONS AND PRESENTATIONS

Li Liu, Minglei Wang, and Kathleen S. Rein. Absolute configuration of pahayokolides A and the biosynthetic origin of the 3-amino-2,5,7,8-tetrahydroxy-10-methylundecanoic acid moiety (Submitted, *Tetrahedron Lett.*)

Roberto Perez, Li Liu, Jose Lopez, Tianying An and Kathleen S. Rein. Diverse Bacterial PKS Sequences Derived From Okadaic Acid-Producing Dinoflagellates. *Mar. Drugs*, **2008**, 6, 164-179.

Tianying An, Thallapuram Krishnaswamy Suresh Kumar, Minglei Wang, Li Liu, Jackson O. Lay, Jr., Rohana Liyanage, John Berry, Miroslav Gantar, Vered Marks, Robert E. Gawley, and Kathleen S. Rein. Structures of pahayokolides A and B, cyclic peptides from a *Lyngbya* sp. *J. Nat. Prod.*, **2007**, 70(5), 730–735.

Li Liu, Minglei Wang, Kathleen S. Rein. The biosynthetic origin of the Athmu moiety of pahayokolide A. *Mycotoxins & Phycotoxins*, June 21-26, 2009, Colby-Sawyer College, New London, New Hampshire. (Poster)

Li Liu, Kathleen S. Rein. Absolute configuration of pahayokolides A and B. Mycotoxins & Phycotoxins, June 17-22, 2007, Colby College, Waterville, Maine. (Poster)

Li Liu, Kathleen S. Rein. Absolute configuration of pahayokolide A. FLACS, Vol. LX, May 10-12, 2007, special issue, pp 40, Orlando, Florida. (Presenter)