

UNIVERSITY OF HAWAII LIBRARY

**INVESTIGATIONS INTO THE MICROCYSTIN GENE CLUSTER  
FROM *HAPALOSIPHON HIBERNICUS* BZ-3-1 AND  
*PLANKTOTHRIX AGARDHII* CYA 126/8**

A THESIS SUBMITTED TO THE GRADUATE DIVISION OF THE  
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE

MASTER OF SCIENCE

IN

CHEMISTRY

DECEMBER 2006

By

**Benjamin Philmus**

Dissertation Committee:

Thomas K. Hemscheidt, Chairperson  
Marcus A. Tius  
Sean M. Callahan


We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Chemistry.

THESIS COMMITTEE



---

Chairperson



## Acknowledgments

First and foremost I would like to thank my advisor Dr. Thomas Hemscheidt and Dr. Guntram Christiansen. Thank you to Dr. Hemscheidt for allowing me the freedom to work my preferred hours and the freedom to explore an entirely new area of research. Thanks to Dr. Guntram Christiansen for allowing me to work with him and being patient while teaching me all the techniques I would need for my research as well as thoughtful discussion and insight during my research. I also need to thank Hilary Okumura for her camaraderie, which made working in the lab fun.

## Abstract

*Hapalosiphon hibernicus* BZ-3-1 is a terrestrial cyanobacterium isolated from a soil sample from Maui, Hawaii. It was shown to produce microcystin-LA during screening for protein phosphatase inhibitors. Using a PCR approach involving degenerate primer mixes, which were derived from comparison of the three known gene clusters previously sequenced from *Microcystis aeruginosa*, *Planktothrix agardhii* and *Anabaena circinalis*. Using the primer mixes the genes *mcyA*, *-B*, *-C*, *-G*, *-H*, and *-J* were amplified in almost their entirety. The genes *mcyD* and *-E* did not give amplicons under similar conditions and therefore needed to be amplified in a slightly different manner. Fragments from the 5' and 3' end of each gene (1 kb) were amplified with the degenerate primer mixes and then the known fragments were connected together using the gathered sequence. After sequencing the genes, the organization of the gene cluster was established by connecting PCR. The resulting gene cluster yielded two operons, *mcyABC* and *mcyGDJEFIH*, which are transcribed in opposite directions from a putative bi-directional promoter. The *H. hibernicus* cluster shows the greatest similarity to the *Anabaena* cluster (*mcyA*, *-B*, *-D*, *-G*, *-J*) and the nodularin gene cluster from *Nodularia* (*mcy-E*, *F*, *-I*).

It was noticed that the *Planktothrix* cluster was the only one to contain *mcyT* and be deficient in *mcyI* and *-F*. Genetic studies were then undertaken to disrupt *mcyT* using homologous recombination. The DNA was introduced into the cyanobacterial cells via electroporation and resistance to chloramphenicol was used to select for the mutant. The mutant was analyzed for peptide production and it was discovered that microcystin biosynthesis had been reduced to 5-10% of that found in the wild type. During

characterization of the mutant peptide production a new anabeanopeptin derivative was found. Structure determination was accomplished using NMR and HR-MS techniques. Stereochemistry of the amino acid residues was determined by the advanced Marfey analysis using LC-MS.

## Table of Contents

<b>Acknowledgments</b> .....	<b>iii</b>
<b>Absract</b> .....	<b>iv</b>
<b>List of Tables</b> .....	<b>x</b>
<b>List of Figures</b> .....	<b>xi</b>
<b>List of Abbreviations</b> .....	<b>xii</b>
<b>1. Sequencing of the <i>mcy</i> gene cluster from <i>Hapalosiphon hibernicus</i></b> .....	<b>1</b>
<b>1.1 INTRODUCTION</b> .....	<b>1</b>
1.1.1 Natural products .....	<b>1</b>
1.1.2 Natural products from microorganisms .....	<b>4</b>
1.1.3 Microcystin.....	<b>5</b>
1.1.4 Polyketide synthetases and non-ribosomal peptide synthetases .....	<b>7</b>
1.1.5 Microcystin toxicity .....	<b>11</b>
1.1.6 <i>Hapalosiphon hibernicus</i> .....	<b>13</b>
1.1.7 Research objectives.....	<b>14</b>
<b>1.2 EXPERIMENTAL SECTION</b> .....	<b>15</b>
1.2.1 General .....	<b>15</b>
1.2.1.1 General laboratory procedures .....	<b>15</b>
1.2.1.2 Preparation of electro-competent <i>E. coli</i> cells .....	<b>15</b>
1.2.1.3 Transformation of electro-competent cells .....	<b>16</b>
1.2.1.4 Preparation of chemical competent cells .....	<b>16</b>
1.2.1.5 Transformation of chemically competent cells .....	<b>18</b>

1.2.1.6 Colony PCR using <i>Taq</i> polymerase from Qiagen .....	18
1.2.1.7 Plasmid preparation by hand .....	18
1.2.1.8 Cyanobacterial DNA isolation .....	19
1.2.2 Cyanobacterial culture .....	21
1.2.2.1 Z+G media .....	21
1.2.2.2 BG-11 media .....	21
1.2.2.3 Cultivation of alga .....	22
1.2.3 <i>E. coli</i> culturing .....	22
1.2.3.1 Antibiotics .....	22
1.2.3.2 Agar plate preparation .....	23
1.2.4 Recombinant DNA procedures .....	23
1.2.4.1 Ligation using T4 DNA ligase .....	23
1.2.4.2 Ligation into pDrive vectors .....	24
1.2.4.3 PCR using <i>Taq</i> DNA polymerase from Qiagen .....	24
1.2.4.4 PCR using Advantage 2 PCR kit .....	24
1.2.4.5 Gel electrophoresis for DNA analysis .....	25
1.2.5 DNA sequencing .....	25
1.2.6 Sequencing of the <i>mcy</i> cluster .....	25
1.2.6.1 Gene amplification of <i>mcyA</i> , - <i>B</i> , - <i>C</i> , and - <i>G</i> .....	25
1.2.6.2 Amplification of <i>mcyD</i> and - <i>E</i> .....	26
1.2.6.3 Amplification of <i>mcyH</i> and - <i>I</i> .....	28
1.2.6.4 Sequencing of <i>mcyA</i> , - <i>B</i> , - <i>C</i> , - <i>D</i> , - <i>E</i> , and - <i>G</i> .....	30
1.2.6.5 Orientation of the <i>mcy</i> gene cluster .....	31

1.2.6.6 Completing <i>mcyH</i> and <i>-C</i>	32
1.3 RESULTS AND DISCUSSION	35
1.3.1 Amplification/Identification of the <i>mcy</i> gene cluster	35
1.3.2 Organization of the microcystin gene cluster	36
1.3.3 Characterization of the microcystin genes	38
1.3.4 NRPS adenylation domain specificities	40
1.3.5 Phylogenetic analysis of adenylation domains	41
1.4 CONCLUSION	43
2 Genetic knockout of <i>mcyT</i> in <i>P. agardhii</i> CYA 126/8	44
2.1 INTRODUCTION	44
2.1.1 <i>McyI</i>	44
2.1.2 <i>McyT</i>	46
2.2 EXPERIMENTAL SECTION	48
2.2.1 Insertional mutagenesis of <i>mcyT</i>	48
2.2.1.1 Construction of the homologous recombination plasmid	48
2.2.1.2 Creating the chloramphenicol disrupted <i>mcyT</i> gene	48
2.2.1.3 Preparation of the transformation construct	49
2.2.1.4 Transformation of <i>P. agardhii</i> CYA 126/8	50
2.2.1.5 Confirmation of mutant identity	51
2.2.2 Peptide analysis	51
2.2.2.1 Peptide extraction of <i>P. agardhii</i> CYA126/8 $\Delta$ <i>mcyT</i>	51
2.2.2.2 HPLC analysis of CYA 126/8 $\Delta$ <i>mcyT</i> peptide content	52
2.2.2.3 LC-MS analysis of CYA 126/8 $\Delta$ <i>mcyT</i> peptide content	53



2.2.3 Anabaenopeptin I .....	53
2.2.3.1 Isolation of Anabaenopeptin I (14) .....	53
2.2.3.2 Anabaenopeptin (14) .....	53
2.2.4 <i>N</i> -Me homotyrosine synthesis .....	54
2.2.4.1 <i>DL-N</i> -Acetyl-homotyrosine .....	54
2.2.4.2 <i>L</i> and <i>D</i> -Homotyrosine .....	54
2.2.4.3 Homotyrosine methyl ester .....	55
2.2.4.4 <i>N,N</i> -Bn,Me homotyrosine methyl ester .....	55
2.2.4.5 <i>N</i> -Me homotyrosine methyl ester .....	56
2.2.4.6 <i>N</i> -Me homotyrosine .....	57
2.3 RESULTS AND DISCUSSION .....	58
2.3.1 Creation of CYA 126/8 $\Delta mcyT$ .....	58
2.3.2 Peptide analysis .....	60
2.3.3 Anabaenopeptin I .....	61
2.3.4 <i>N</i> -Me homotyrosine synthesis .....	64
2.3.5 Stereochemistry of Anabaenopeptin I (14) .....	65
2.4 CONCLUSION .....	67
3 Appendix .....	69
4 References .....	72

## List of Tables

### Table

1. Known microcystin variants.....	6
2. Degenerate primer sequences used in the amplification of <i>mcyA</i> , - <i>B</i> , - <i>C</i> , - <i>D</i> , - <i>E</i> , - <i>G</i> , - <i>H</i> , and - <i>J</i> . .....	29
3. Primers used during the orientation PCR reactions. ....	31
4. Primers used during the completion of <i>mcyC</i> and - <i>H</i> . .....	34
5. Percent homology of genes from <i>H. hibernicus</i> to known <i>mcy</i> genes.....	38
6. Amino acid specificity conferring sequences of adenylation domains as Stachelhaus et al. ....	40
7. NMR data for Anabaenopeptin I (14). .....	63

## List of Figures

1. Starter units for fatty acid synthetase and polyketide synthetase systems. ....	7
2. Conversion of <i>apo</i> protein into <i>holo</i> protein by nucleophilic attack of the conserved serine residue on coenzyme A. ....	9
3. Schematic diagram of the microcystin clusters from <i>M. aeruginosa</i> , <i>P. agardhii</i> , <i>A. circinalis</i> and the nodularin gene cluster from <i>N. spumigena</i> . ....	11
4. Biosynthesis of microcystin-LA. ....	12
5. Schematic diagram of the <i>mcy</i> cluster sequenced from <i>H. hibernicus</i> as well as the known clusters from <i>M. aeruginosa</i> , <i>P. agardhii</i> , <i>A. circinalis</i> and the nodularin gene cluster from <i>N. spumigena</i> . ....	37
6. (a) Evolutionary tree diagram of the adenylation domains found in the microcystin clusters and the nodularin cluster using GrsA-A1 as an out-group. (b) Evolutionary tree diagram of the condensation domains found in the microcystin clusters as well as the nodularin cluster. ....	42
7. Multialign of 3-D-phosphoglycerate dehydrogenases from <i>E. coli</i> W3110, <i>E. coli</i> K-12, <i>Yersinia pestis</i> , <i>Mycobacterium tuberculosis</i> , <i>Synechocystis sp. PCC 6803</i> , <i>Homo sapiens</i> and McyI from <i>H. hibernicus</i> BZ-3-1. ....	45
8. 2.0% agarose gel verifying the correct mutant had been created. PCR performed with primers <i>mcyT5'seq</i> and <i>mcyT3'seq</i> . ....	60
9. Synthesis of enantiomerically enriched <i>N</i> -Me homotyrosine. ....	65

## List of Abbreviations

4'-PP	4'-phosphopantetheinyl arm
A	adenylation
Ac	acetyl
AcCN	acetonitrile
ACP	acetyl carrier protein
Adda	(2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-(4E, 6E)- decadienoic acid
ADP	adenosine diphosphate
Amp	ampicillin
AT	acetyl transferase
ATP	adenosine triphosphate
C	condensation
Cm	chloramphenicol
CMT	carbon methyl transferase
CoA	Coenzyme A
dH <sub>2</sub> O	deionized water
DNA	deoxyribonucleic acid
Ep	epimerization
ESI	electrospray ionization
EtBr	ethidium bromide
EtOAc	ethyl acetate
EtOH	ethanol

FABMS	fast atom bombardment mass spectrometry
FDAA	1-fluoro-2-4-dinitrophenyl-5- <i>L</i> -alanine amide
HMBC	heteronuclear multiple bond correlation
HMQC	heteronuclear multiple quantum coherence
Hty	homotyrosine
IPTG	isopropyl-beta- <i>D</i> -thiogalactopyranoside
KR	ketoreductase
KS	ketosynthase
LB <sub>Amp</sub>	LB media containing 100 µg/mL ampicillin
LC-MS	liquid chromatography-mass spectrometry
Masp	methyl aspartate
MC	microcystin
Mdha	<i>N</i> -methyl dehydroalanine
MeOH	methanol
MOPS	3-( <i>N</i> -Morpholino)propanesulfonic acid
MSD	mass selective detector
Na <sub>2</sub> EDTA	ethylenediaminetetraacetate, disodium salt
NAD	nicotinamide adenine dinucleotide
NMR	nuclear magnetic resonance
NMT	<i>N</i> -methyl transferase
NRPS	nonribosomal peptide synthetase
ORF	open reading frame
PCR	polymerase chain reaction

<b>PKS</b>	<b>polyketide synthase</b>
<b>PP 1/PP 2a</b>	<b>protein phosphatase 1/ protein phosphatase 2a</b>
<b>RT</b>	<b>room temperature</b>
<b>SAM</b>	<b><i>S</i>-adenosyl methionine</b>
<b>T</b>	<b>thiolation</b>
<b>Te</b>	<b>thioesterase</b>
<b>Tm</b>	<b>melting temperature</b>
<b>TMS-Cl</b>	<b>chlorotrimethylsilane</b>
<b>TOCSY</b>	<b>total correlated spectroscopy</b>
<b>TOF</b>	<b>time of flight</b>
<b>XGal</b>	<b>5-bromo-4-chloro-3-indolyl-<math>\beta</math>-<i>D</i>-galactoside</b>

# **1. Sequencing of the *mcy* gene cluster from *Hapalosiphon hibernicus***

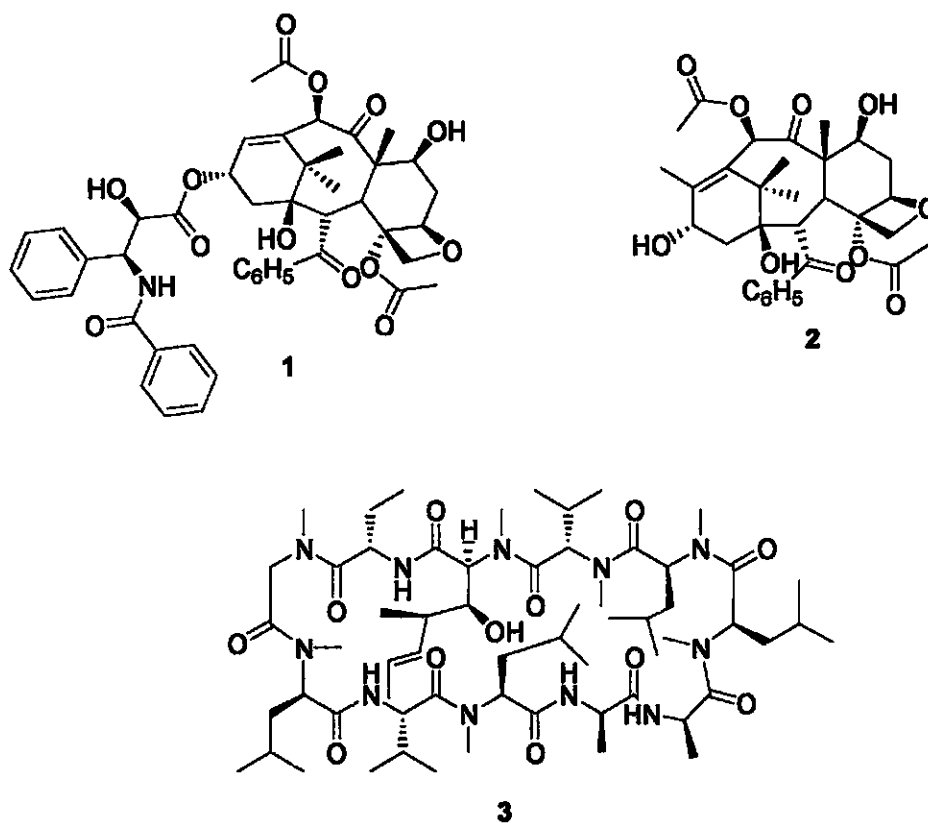
## **1.1 INTRODUCTION**

### **1.1.1 Natural products**

Throughout the history of human civilization, humans and natural products have had a close, intertwining relationship. From the ancient Incas chewing on coca leaves for the energizing effect of cocaine to the use of willow bark as a source of pain and fever relief, to the discovery of penicillin by Sir Alexander Fleming, nature has provided a tremendous number of bioactive metabolites that have influenced civilizations.

Currently 60% of anticancer and 70% of anti-infective antibiotics are natural products or semisynthetic derivatives.<sup>1</sup> For example the anticancer chemotherapeutic agent paclitaxel (1) is obtained in commercially relevant amounts by synthetic derivation of baccatin III (2), which is isolated from the needles of the English yew tree (*Taxus baccata*).<sup>2</sup> Cyclosporin (3), the immunosuppressant used to suppress graft rejections after organ transplants, is isolated from cultures of the fungus *Tolypocladium inflatum* Gams.<sup>3</sup>

Microorganisms have been a tremendous source of bioactive molecules; the actinomycetes are especially valuable in the search for and production of clinically useful drugs,<sup>4</sup> while more recently other groups of microorganisms such as cyanobacteria have been investigated. Marine invertebrates, particularly from the phylum *Porifera* (sponges) and *Coelenterata* (bryozoans, corals, etc.) have also been heavily explored for novel natural products. The natural products discovered from these macroscopic organisms have been hypothesized to originate, at least in part, from symbiotic microorganisms. With the recent advances in microbial culture and genetic techniques, additional evidence for the endosymbiosis production theory has been obtained. An example is swinholide A



(4), which was originally isolated from the marine sponge *Theonella swinhoei* by Kobayashi and co-workers.<sup>5</sup> Bewley *et al.* found swinholidide A to be localized within the symbiotic unicellular bacteria within this sponge, which were separated using centrifugation.<sup>6</sup> More recently swinholidide A and a derivative were isolated from a collection of marine cyanobacteria.<sup>7</sup>

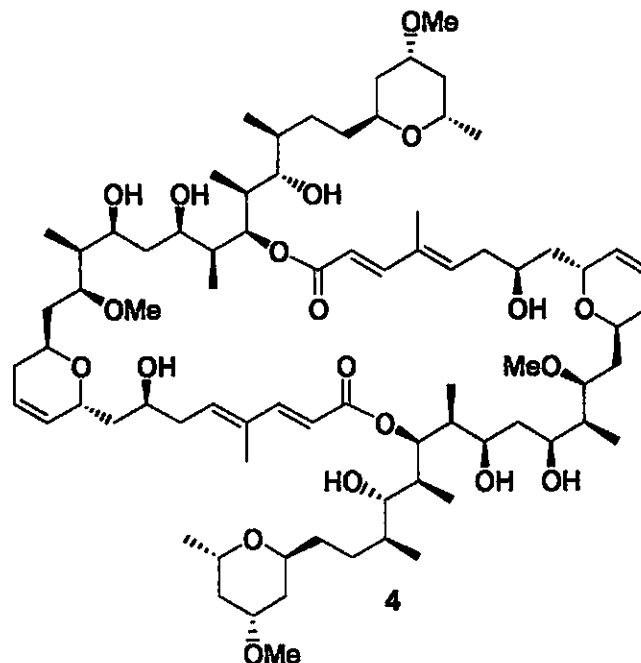
The isolation of bioactive compounds from microorganisms offers unique advantages over isolation from larger organisms. First, microorganisms typically contain higher percentages of metabolite per kilogram dry mass, and microorganisms can be easier to culture and maintain in the laboratory. However there are numerous problems associated with culturing of microorganisms, as it has been estimated from genetic data

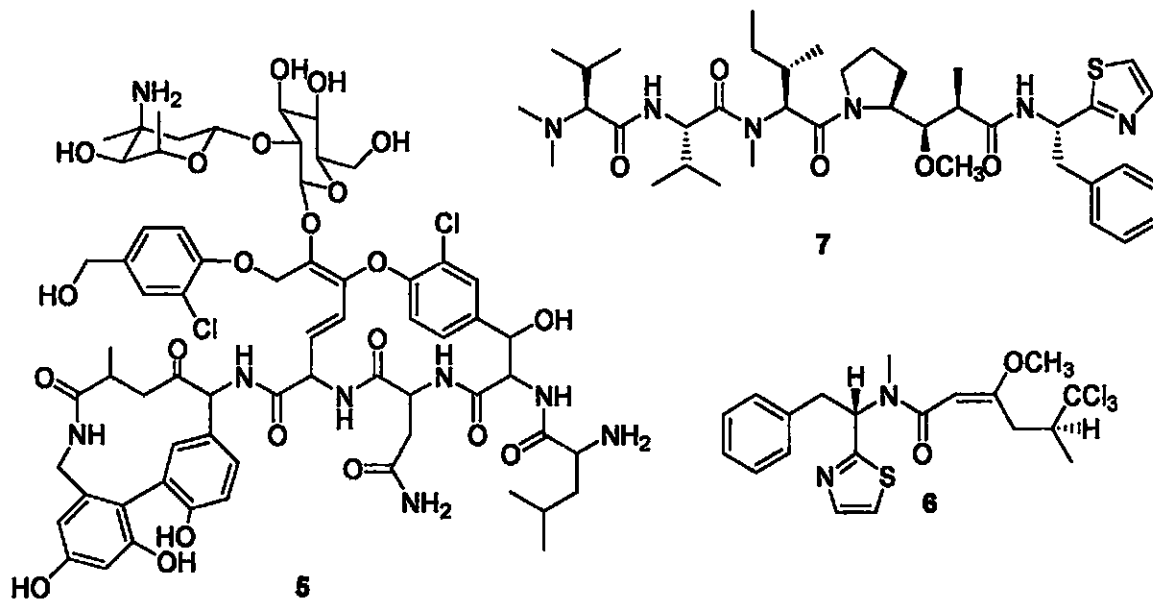


that fewer than 1% of known microorganisms have been cultured and examined in the lab. To make things more complicated some bacterial cultures have been known to only produce certain metabolites in the presence of a specific stimulus.<sup>8</sup>

The possibility for the discovery of novel structural skeletons from microbes is

immense considering the large unexplored microbial genetic diversity and the low percentage of microbes that have been cultured. To this date over 7,000 natural products have been isolated from the actinomycetes alone with most produced by the genus *Streptomyces* (80%) and a few other well-known soil dwelling genera.<sup>4</sup> The diversity of known structural backbones and biological activities of compounds isolated from microorganisms range from vancomycin (5), a cyclic peptide antibiotic isolated from *Amiclatopsis orientalis* to barbamide (6), a linear polyketide-peptide hybrid with potent molluscicidal activity<sup>9</sup> to the linear peptide dolastatin 10 (7), which disrupts microfilament depolymerization.<sup>10</sup> Both barbamide and dolastatin 10 were isolated from the marine cyanobacterium *Lyngbya majuscula*.



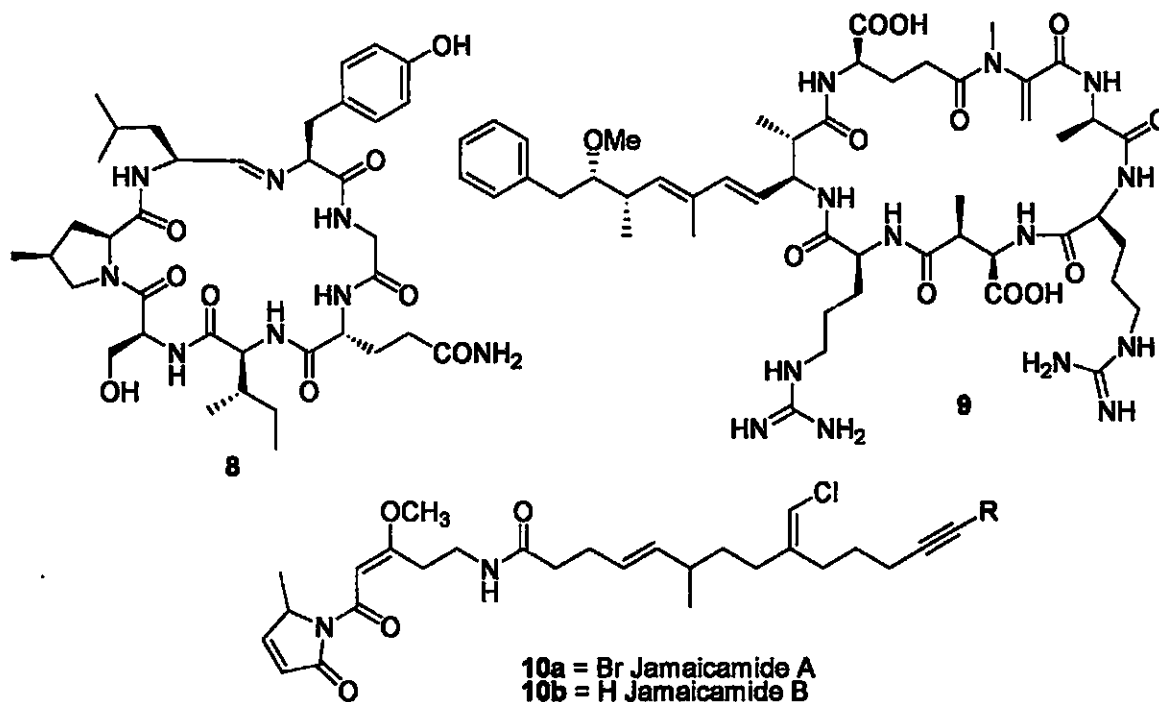


### 1.1.2 Natural products from microorganisms

Large multi-modular enzymes called polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) catalyze the formation of the two largest families of natural products including the compounds mentioned above. However *in silico* analysis of *Streptomyces coelicolor* and *S. avermitilis* showed that only 12% (14 of 118) of putative PKS and NRPS clusters had been correlated with the production of a specific natural product.<sup>11</sup> This data coupled with the estimate that 1% of microorganisms have been cultured suggests that a wide variety of unknown structures could/should exist.

Cyanobacteria have recently become a hot area of study because of their ability to produce prolific amounts of secondary metabolites, typically cyclic peptides and depsipeptides, which are produced by PKS and NRPS systems. As is the case with *S. coelicolor* and *S. avermitilis*, the sequenced genome of *Nostoc punctiforme* contains 22 putative PKS and NRPS clusters, however no natural product has been definitively associated with any of these clusters. To date, ten different cyanobacterial gene clusters

have been sequenced including the clusters responsible for nostocyclopeptide A1 (8), microcystin-RR (9), jamaicamide (10), and barbamide (6) biosynthesis.

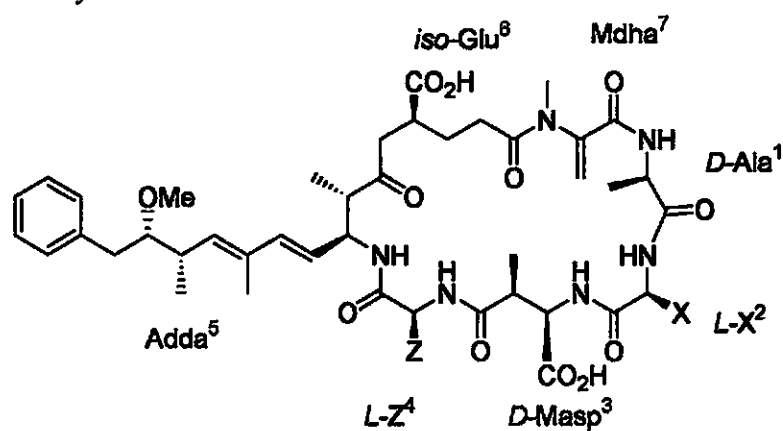


### 1.1.3 Microcystin

The microcystins (MCs) are a class of compounds with the general structure *cyclo*-(*D*-Ala-*L*-X-*D*-erythro- $\beta$ -Masp-*L*-Z-Adda-*D*-isoGlu-Mdha), where X and Z are variable amino acids, *D*-erythro- $\beta$ -Masp is (3*S*)-3-methyl aspartate, Mdha is *N*-methyl dehydroalanine and Adda is (2*S*,3*S*,8*S*,9*S*)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyl-(4*E*, 6*E*)-decadienoic acid. Over 70 variants have been described (Table 1) with the majority of the substitutions taking place at the X and Z variable positions. The naming of the microcystins is based on the above base formula with the variable position amino acids listed afterwards in one letter code. For example, microcystin-LA (MC-LA) has the structure *cyclo*-(*D*-Ala-*L*-Leu-*D*-erythro- $\beta$ -Masp-*L*-Ala-Adda-*D*-isoGlu-Mdha).

While some characterized variants can be viewed as artifacts of isolation, thereby reducing the actual number of naturally occurring variants, the structural diversity remains impressive. Interestingly, most strains that have been surveyed for microcystin production contain only a few of these multiple variants, albeit at different levels (i.e. *Planktothrix agardhii* CYA 126/8 produces (*D*-Asp<sup>3</sup>)-MC-RR (90%), (*D*-Asp<sup>3</sup>)-MC-LR (9%) and (*D*-Asp<sup>3</sup>)-MC-YR (1%)). The production of multiple peptide variants is not a unique occurrence found only in microcystin producing strains. Moore and co-workers have isolated 24 cryptophycin variants from the strain *Nostoc* sp. GSV 224.<sup>12</sup>

Table 1: Known microcystin variants



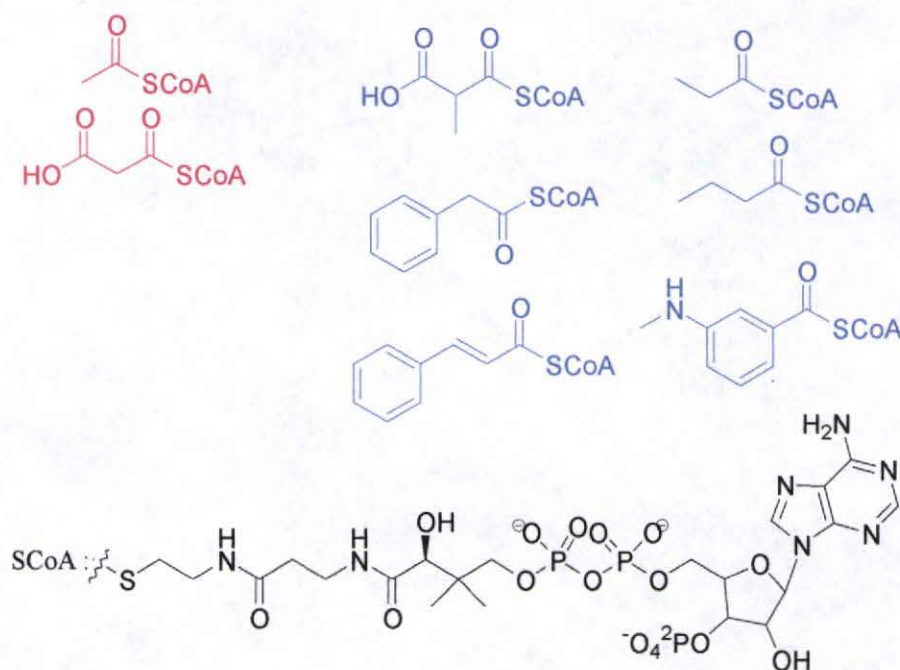
<i>D</i> -Ala <sup>1</sup>	X <sup>2</sup>	<i>D</i> -Masp <sup>3</sup>	Z <sup>4</sup>	Adda <sup>5</sup>	<i>D</i> -Glu <sup>6</sup>	Mdha <sup>7</sup>
<i>D</i> -Ser	Leu	<i>D</i> -Asp	Ala	DMAdda	<i>D</i> -Glu(OMe)	Dha
<i>D</i> -Leu	Ala		Aba	ADMAdda		Ser
	Tyr		Leu	6Z-Adda		Dhb
	Glu		Arg			Me-Ser
	Val		Glu			Me-Lan
	Glu(OMe)		Glu(OMe)			
	Arg		Phe			
	Phe		Tyr			
	Met-Ox		HArg			
	HPhe		Trp			
	HTyr		Met-Ox			
	Trp		HArg			
	HArg					
	Hlle					
	(H <sub>4</sub> )Tyr					

Glu(OMe): Methyl glutamate, Met-Ox: Methionine sulfoxide, HPhe: Homophenylalanine, HTyr: Homotyrosine, HArg: Homoarginine, Hlle: Homoisoleucine, (H<sub>4</sub>)Tyr: 1,2,3,4-tetrahydrotyrosine, Aba: amino-isobutyric acid, DMAdda: Desmethoxy Adda, ADMAdda: O-acetyl-O-Desmethyl Adda, Dha: Dehydroalanine, Dhb: Dehydrobutyrate, Me-Ser: N-methyl serine, Me-Lan: N-methyl lanthionine

### 1.1.4 Polyketide synthetases and non-ribosomal peptide synthetases

Polyketides are synthesized by polyketide synthases, while peptides (linear and cyclic) and depsipeptides are synthesized by non-ribosomal peptide synthetases. There are three types of PKS systems, Type I, II and III, and they are classified by their mechanistic characteristics and resulting products.<sup>13</sup> Type I PKS systems are the aforementioned large multifunctional enzymes found in the microcystin gene cluster. Type I polyketide synthases (from hereon out simply referred to as PKS) are related to fatty acid synthases (FAS) in that they both contain conserved domains for selection, coupling and modification of carbon extender units and both rely on the *thio*-template mechanism for product construction. Unlike FAS, which can only incorporate acetyl-CoA or malonyl-CoA building blocks, PKSs can utilize a wide variety of starting materials (Figure 1). After the Claisen reaction, the FAS system reduces the resulting  $\beta$ -keto thioester

to the  $\beta$ -hydroxy thioester, followed by dehydration to form the  $\alpha,\beta$ -unsaturated thioester and lastly



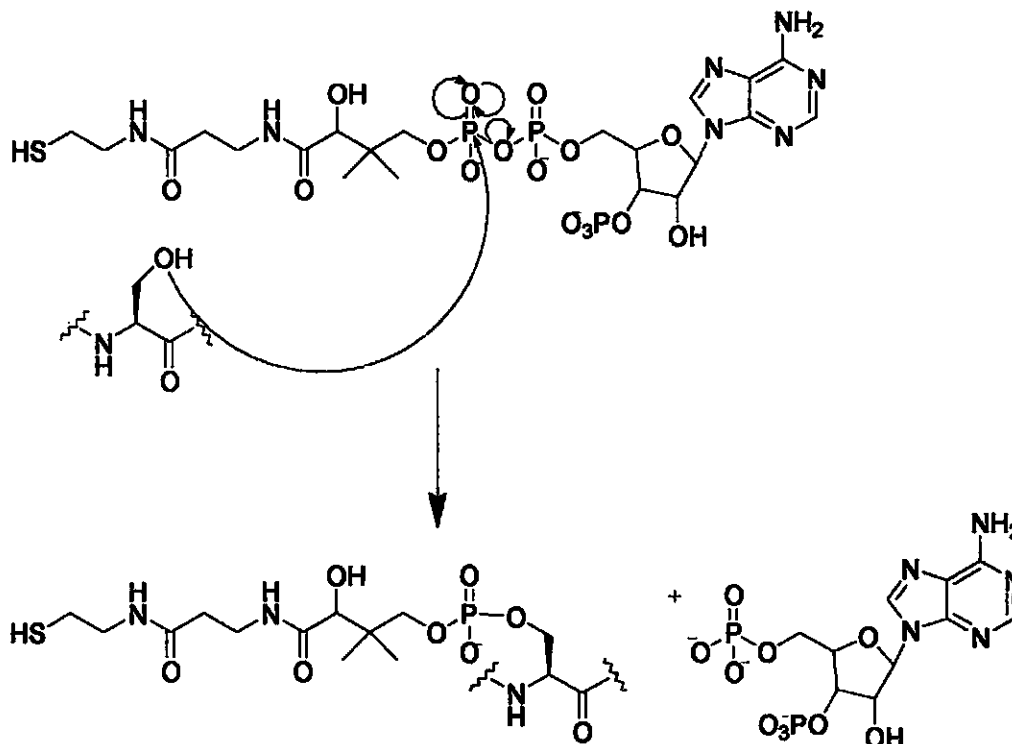
**Figure 1:** Starter units for fatty acid synthase systems (red) and polyketide synthase systems (red and blue). Adapted from <http://203.90.127.50/~pkfdb/polyketide.htm>

reduction of the resulting alkene to an alkane. Done repetitively, this results in a long saturated alkyl chain. In contrast PKS systems can do all or none of these reactions following the condensation. Moreover they may introduce methyl groups at the C1 and C2 position of the incorporated acetate unit. This flexibility, along with the diverse pool of extender units contributes to the large structural diversity of polyketide natural products. NRPSs operate by a similar means only using amino acids instead. NRPSs also modify the incorporated amino acids by N-methylation, epimerization, cyclization (of Thr, Ser and Cys residues) and oxidation or reduction of the cyclized amino acid.

NRPS and PKS contain conserved domains that perform different aspects of the biosynthesis.<sup>14</sup> Adenylation or A domains (NRPS) and ketosynthase or KS domains (PKS) are analogous in function. They select and activate the correct building block for incorporation. Adenylation domains activate the amino acid as the AMP mixed anhydride with the formation of pyrophosphate while ketosynthase domains utilize Co-enzyme A bound substrates. Condensation or C domains (NRPS) and acyl transferase or AT domains (PKS) perform the coupling of neighboring units by amide bond formation or a thio-Claisen reaction respectively. Thiolation or T domains (NRPS) and acyl carrier proteins or ACPs (PKS) are responsible for tethering the growing chain to the enzyme through a thioester linkage.

There are two forms of each enzyme, the *apo* or inactive form and the *holo* or active form. An enzyme known as a phosphopantetheinyl transferase mediates the conversion of the *apo* to the *holo* form. This enzyme transfers a 4'-phosphopantethienyl arm (4'-PP) onto a conserved serine residue of the T or ACP domain. The *holo* form of

the enzyme is established by nucleophilic attack of the  $\beta$ -phosphate of Co-enzyme A by the serine hydroxyl group (Figure 2).



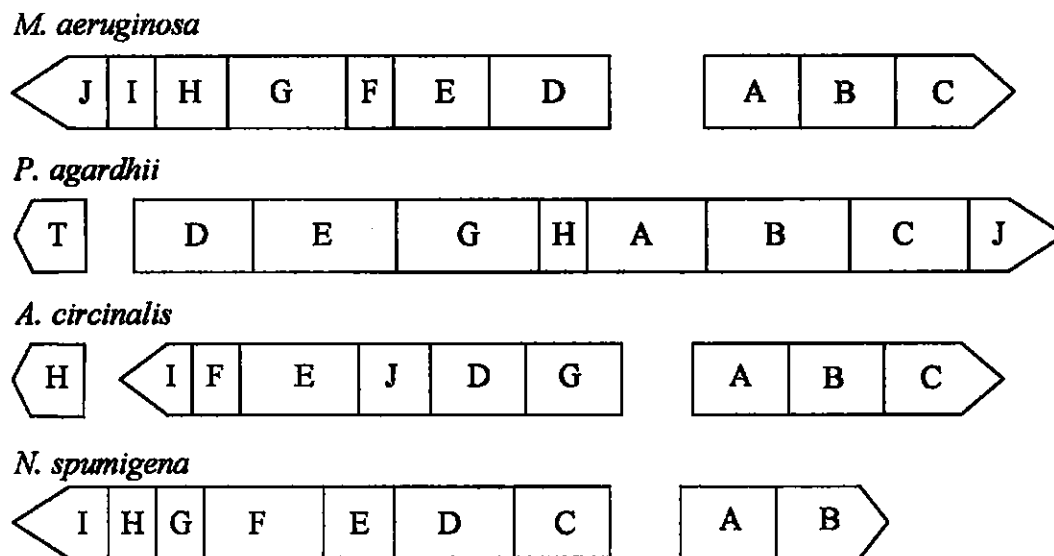
**Figure 2:** Conversion of *apo* protein into *holo* protein by nucleophilic attack of the conserved serine residue on coenzyme A

Type I thioesterase domains (NRPS/PKS) are responsible for cleavage of the final product from the enzyme either by hydrolysis to form the linear product or by cyclization to yield a lactone or lactam type product. Alternatively, a reductase domain can also cause the release of the peptide as the aldehyde, as in the case of nostocyclopeptide biosynthesis.<sup>15</sup> Exclusive to PKS systems are the C-methyl transferase (CMT), the ketoreductase (KR), the dehydratase (DH), and the enoyl reductase (ER) domains. These domains are responsible for methylating the C2 of an acetate extender unit (the methyl is derived from S-adenosyl methionine (SAM)), reducing the ketone to the  $\beta$ -hydroxy-

thioester, dehydration to generate the  $\alpha,\beta$ -unsaturated thioester and finally reduction to the saturated thioester respectively. NRPSs have their own associated domains including the N-methyl transferase (NMT) and the epimerization (Ep) domains. The NMT catalyzes the addition of a methyl group to the  $-\text{NH}_2$  group prior to amide bond formation while the epimerization domain converts the naturally occurring L-amino acid to the unnatural D-amino acid.

The microcystins are synthesized by a PKS-NRPS hybrid system. The gene clusters from three limnic cyanobacterial strains, *Microcystis aeruginosa* PCC 7806,<sup>16</sup> *Planktothrix agardhii* CYA 126/8<sup>17</sup> and *Anabaena circinalis* strain 90<sup>18</sup> have been sequenced in addition to the gene cluster for the related natural product nodularin. The gene clusters contain either nine or ten genes involved in microcystin biosynthesis. All three contain *mcyA*, *-B*, *-C*, *-D*, *-E* and *-G* which are NRPS/PKSs that are responsible for the assembly of the structural backbone of microcystin, *mcyJ*, which encodes for an O-methyl transferase and *mcyH* which codes for a putative ABC transporter. *McyJ* is responsible for the methylation of the hydroxyl group to form the methoxy group found on the Adda side chain. The clusters obtained from *M. aeruginosa* and *A. circinalis* contained two modifying genes named *mcyF*, an aspartate racemase and *mcyI*, a D-3-phosphoglycerate dehydrogenase homolog. *P. agardhii* was discovered to be deficient in these genes but did contain a putative type II thioesterase gene, *mcyT* (Figure 3). Figure 4 shows the domains present in each gene and what aspect of the biosynthesis each gene is responsible for. The Te domain found at the 3' end of *mcyC* is responsible for cyclization by forming the amide bond between the carboxyl group of the variable amino acid Z and the amino group found on the Adda chain.





**Figure 3:** Schematic diagram of the microcystin gene clusters from *M. aeruginosa*, *P. agardhii*, *A. circinalis* and the nodularin gene cluster from *N. spumigena*. Genes not drawn to scale.

### 1.1.5 Microcystin toxicity

The toxicity of the microcystins has been studied intensively because of the danger toxic blooms pose to humans and livestock. The microcystins have recently been blamed for the deaths of approximately 50 dialysis patients in a Brazilian hospital,<sup>19</sup> and currently microcystins are being investigated as the cause of high levels of liver cancer in China.<sup>20,21</sup> The World Health Organization (WHO) has set the limit for microcystin-LR at 1 µg per liter of water as safe for human consumption.<sup>22</sup> Microcystin-LR was found to be the most toxic variant with an LD<sub>50</sub> in mice of 43 µg per kg body weight while microcystin-RR was found to be less toxic with a LD<sub>50</sub> of 235 µg per kg body weight.

The mechanism of action of the MCs is a result of their high affinity for the eukaryotic PP1 and 2a. Most microcystin variants described to date have some level of inhibition of eukaryotic PP1 and PP2A. Protein phosphatases 1 and 2A regulate a wide



variety of cellular functions by working in concert with protein kinases to regulate the activity of enzymes. Processes such as cell signaling, glycogen storage and cell division are regulated by PP1 and 2A. Goldberg and coworkers co-crystallized human PP1 and MC-LR and therefore were able to deduce the key interactions by X-ray crystallography.<sup>23</sup> One observation was that the solution and bound conformations of MC-LR did not differ significantly which results in a low activation barrier for binding. The Adda side chain fits snugly in the hydrophobic groove of PP1 where it interacts with the hydrophobic residues. The carboxylate group and the amide carbonyl of the *D-iso*Glu residue hydrogen bonds to two water molecules found in the active site of the enzyme. The water molecules are in turn coordinated to two metal ions ( $\text{Fe}^{3+}$ ,  $\text{Mn}^{2+}$ ) found in the active site. The carboxylate of the *D-erythro*-Masp hydrogen bonds to two residues (Arg-96 and Tyr-134) while the Mdha unit was covalently bound to Cys-273. However this last interaction was found to be non-essential as mutation of the cysteine did not reduce the binding affinity of MC-LR.

#### 1.1.6 *Hapalosiphon hibernicus*

*Hapalosiphon hibernicus* W & GS West BZ-3-1 was isolated from a soil sample collected on Maui, HI. It was shown to produce microcystin-LA during screens for protein phosphatase inhibition by Moore and co-workers.<sup>24</sup> This became an important discovery because it was the first terrestrial strain as well as first subsection V cyanobacterium to be identified as a microcystin producer. Cyanobacteria in subsection I (*Microcystis*), III (*Planktothrix*), and IV (*Anabaena*) had previously been shown to produce microcystin. More recently another terrestrial, lichen-associated cyanobacterium (*Nostoc* sp. strain IO-102-I) was shown to produce a few microcystin variants.<sup>25</sup>

### **1.1.7 Research objectives**

These studies were undertaken to 1) study the gene cluster of a terrestrial strain in the hopes that new proteins involved in the biosynthesis of microcystin could be located. As well as examine whether the biosynthesis of microcystin was conserved between limnic and terrestrial strains. 2) to develop a PCR based method for sequencing microcystin gene clusters from any cyanobacterial source.

## 1.2 EXPERIMENTAL SECTION

### 1.2.1 General

#### 1.2.1.1 General laboratory procedures

All cell handling and transfer was performed using sterile technique under a laminar flow hood (Labconco, Kansas City, MO). All centrifugations were performed in a Sorvall RC 5B Plus centrifuge (Thermo Electron Corporation, Asheville, NC). All DNA primers were obtained from Integrated DNA Technologies (Coralville, IA) with standard desalting purification and were resuspended in dH<sub>2</sub>O to a stock concentration of 100 pmol/μl and used without further purification. Growth media used for *Escherichia coli* were prepared and sterilized according to standard protocols.<sup>26</sup> All PCR reactions were performed in a GeneAmp PCR system 2700® (Applied Biosciences). All OD<sub>600</sub> measurements were taken in a SP-830 spectrophotometer (Barnstead International, Dubuque, IA) using disposable polypropylene cuvettes (Fisher Scientific, Pittsburgh, PA). *E. coli* cells in liquid media were grown in an Environ shaker (Lab-line) at 37°C with shaking at 220 rpm. Agar plates were grown in an IsoTemp incubator (Fisher Scientific, Pittsburgh, PA) at 37°C for 15-20 h.

#### 1.2.1.2 Preparation of electro-competent *E. coli* cells

An individual colony was inoculated into 5 ml of LB broth and grown to saturation (OD<sub>600</sub> > 2). The following day 2.5 ml of the culture was inoculated into 500 ml YENB media.<sup>27</sup> The cells were grown to an OD<sub>600</sub> between 0.5 and 0.9 at which time the cells were transferred to sterile 250 ml collection bottles and pelleted in a SLA-1500 rotor pre-cooled to 4°C (5000 rpm/15 min/4°C). The supernatant was decanted and the cells were resuspended in 25 ml of sterile reverse osmosis water (pre-cooled to 4°C) and

then diluted up to 200 ml. The cells were then pelleted (5000 rpm/15 min/4°C) and the washing and centrifugation repeated. The cells were then resuspended in 50 ml total volume of 10% glycerol (pre-cooled to 4°C, sterile filtered). The cells were centrifuged again in a SH-3000 rotor pre-cooled to 4°C (4500 rpm/15 min/4°C). The cells were resuspended in 1-1.5ml 10% glycerol and allocated in sterile 1.5 ml micro-centrifuge tubes in 40 µl aliquots, which were then flash frozen in liquid nitrogen and stored at -80°C until use.

### 1.2.1.3 Transformation of electro-competent cells

Previously aliquoted electro-competent cells were defrosted on ice while ligation mixtures were dialyzed against reverse osmosis water using a 0.2 µ nylon membrane (Phenomenex, Torrance, CA) for 5 minutes. The ligation mixture was added to the thawed cells and the mixture was transferred to a pre-cooled (-20°C) 0.2 cm Gene Pulser® cuvette (Bio-Rad, Hercules, CA). The electroporation was performed using a Gene Pulser II® (Bio-Rad) (1.7 kV, 200 Ω, 25 µF) resulting in time constants of 4.65 – 4.91 msec. Immediately after electroporation the cells were resuspended in 250 µL room temperature SOC media and transferred to a sterile 1.5ml micro-centrifuge tube. The cells were then grown in an orbital shaker for 1 hour. The cells were then spread on LB agar plates containing the appropriate antibiotics using a glass spreader.

### 1.2.1.4 Preparation of chemical competent cells

Transformation Buffer A (100 ml):	RbCl	1.2 g
	MnCl <sub>2</sub> •4 H <sub>2</sub> O	0.99 g

KOAc (1M, pH 7.5)	3 ml
CaCl <sub>2</sub> •2 H <sub>2</sub> O	0.15 g
Glycerol	15 g

-Adjust pH to 5.8 with 0.2M acetic acid

-Sterilize with 0.2 µm filter

Transformation Buffer B (25 ml):	MOPS (0.5M, pH 6.8)	0.5 ml
	RbCl	0.03 g
	CaCl <sub>2</sub> •2 H <sub>2</sub> O	0.275 g
	Glycerol	3.75 g

-Adjust pH 6.8 with NaOH

- Sterilize with 0.2 µm filter

An individual colony was inoculated into a 3 ml overnight culture in LB media. The following day 500 µl of the overnight culture was inoculated in 50 ml SOB broth containing the appropriate antibiotics. The cells were grown in an orbital shaker until the OD<sub>600</sub> was between 0.3 and 0.5. The cells were transferred to a 50 ml Falcon tube, placed on ice for 15 min and pelleted using a SH-3000 rotor pre-cooled to 4°C (4500 rpm/15 min/4°C). The supernatant was decanted, the pellet resuspended in 16 ml of Transformation Buffer A, and allowed to sit on ice for 15 min. The cells were pelleted in a SH-3000 rotor (4500 rpm/15 min/4°C), the supernatant drained and the pellet resuspended in 4 ml Transformation Buffer B. The cells were then aliquoted in sterile 1.5 ml micro-centrifuge tubes in 200 µl aliquots and flash frozen in liquid nitrogen and stored at -80°C until use.

### 1.2.1.5 Transformation of chemically competent cells

Chemically competent cells were thawed on ice, DNA was then added and the tube mixed by quickly vortexing. The cells were incubated on ice for 30 min followed by heat shock at 42°C for 40-45 sec and then placed back on ice for 1 min. 500 µl SOC media (pre-warmed to 37°C) was added to the cells and they were incubated in an orbital shaker for 1 hour. Using a glass spreader the cells were then plated out on LB agar plates containing the appropriate antibiotic.

### 1.2.1.6 Colony PCR using *Taq* polymerase kit from Qiagen (Valencia, CA)

#### Master mix

36.25 µl reverse osmosis water  
5 µl 5X PCR buffer 3 µl 25mM MgCl<sub>2</sub> solution  
1.5 µl dNTP mixture (10 µM each)  
1.5 µl forward primer (10pmol/µl)  
1.5 µl reverse primer (10pmol/µl)  
0.25 µl *Taq* DNA polymerase

Individual colonies were suspended in 10 µl of reverse osmosis water. The master mix was mixed and aliquots of 9 µl were placed in 0.2 µl thin walled PCR tubes (Corning, Acton, MA) and 1 µl of the cell suspension in water was added to each PCR tube. The PCR reaction was performed using the following program. 95 °C for 3 min followed by 32 cycles of 95°C for 30 sec, 66°C for 30 sec then 72°C for a variable amount of time (1 min per kb of DNA to be amplified) followed by cooling to 25°C. The PCR products were then observed by gel electrophoresis.

### 1.2.1.7 Plasmid preparation by hand

Plasmid Isolation Buffer (PIB) 1: 50 mM Tris•HCl, 10mM Na<sub>2</sub>EDTA, pH 8.0



Plasmid Isolation Buffer (PIB) 2: 0.2 M NaOH, 1% SDS

Plasmid Isolation Buffer (PIB) 3: 4 M guanidine-HCl, 0.5 M KOAc, pH 4.2

The cell pellet was resuspended in 250  $\mu$ l PIB 1 by vortexing. PIB 2 (250  $\mu$ l) was added and mixed by inverting gently 4-5 times. After 5 min at RT, PIB 3 (350  $\mu$ l) was then added and the tubes mixed by inversion 6 times followed by incubation on ice for 5 min. The white precipitate formed was pelleted by centrifugation in a F-20/Micro rotor (11,000 rpm/10 min/4°C). The supernatant was removed by pipette and transferred to a second 1.5 ml micro-centrifuge tube. The tubes were centrifuged down again as described above and the supernatant was carefully transferred to a new 1.5 ml micro-centrifuge tube. Isopropanol (565  $\mu$ l) was added to the cleared supernatant and mixed by gentle inversion. The tube was allowed to stand at room temperature for 30 min and then the DNA was precipitated by centrifugation in a F-20/Micro rotor (11,000 rpm/30 min/4°C). The supernatant was removed and the pellet washed with 750  $\mu$ l 70% (v/v) aqueous ethanol. The ethanol was aspirated and the pellet air-dried for 10-12 min. The dried pellet was resuspended in 100  $\mu$ l 10mM Tris buffer (pH 7.5). 1  $\mu$ l RNase A/T1 mixture (Fermentas, Hanover, MD) was added and the tube mixed followed by incubation at 37°C for 1 hour. The plasmid was then used without further purification.

#### **1.2.1.8 Cyanobacterial DNA Isolation**

TE Buffer: 10mM Tris•HCl, pH 8.0, 1mM EDTA, pH 8.0

TES Buffer: 10mM Tris•HCl, pH 8.0, 1mM EDTA, pH 8.0, 40% (w/v) sucrose

Isolation of DNA from cyanobacteria cultures was performed essentially described<sup>28</sup> but with a few minor modifications. Cyanobacterial cells were pelleted in a SH-3000 rotor (5000 rpm/15 min/RT). The supernatant was removed and the cells were washed three times with 10 volumes (compared to the volume of cells) of TE buffer. The cells were suspended in 50 - 300  $\mu$ l of TES buffer and kept on ice for 1 hour. Lysozyme (0.1 volumes) (Sigma-Aldrich, 50 mg/ml) was added, the solution mixed thoroughly and incubated at 37°C for 30 – 45 min. 20% SDS (0.1 volumes) and Proteinase K (0.1 volumes) (Fermentas, Hanover, MD, 1mg/ml) were added to the cells and the mixture stored at 65°C overnight. The resulting brown mixture was diluted to 700  $\mu$ l with TES buffer and extracted with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1) with vigorous shaking for 5 min. The resulting emulsion was separated by centrifugation (Eppendorf 5417C centrifuge, 14,000 rpm/10 min). The aqueous layer was carefully removed by pipette and transferred to a new 1.5 ml micro-centrifuge tube. This extraction was repeated until no precipitate was observed at the interface of the two resulting layers. The aqueous layer was washed with an equal volume of chloroform followed by centrifugation to separate the layers (Eppendorf 5417C centrifuge, 14,000 rpm/10 min). The aqueous layer was carefully removed and transferred to a 1.5 ml micro-centrifuge tube where isopropanol (0.7 volumes) was added. The tube was mixed by inversion and stored at room temperature for 30 – 60 min. The DNA was then pelleted by centrifugation in an F20/Micro rotor (11,000 rpm/60 min/4°C). The supernatant was carefully removed and the pellet washed with 1 ml 70% (v/v) aqueous ethanol. The ethanol was removed, the pellet air dried for approximately 10 – 15 min and then suspended in an appropriate volume (100 – 300  $\mu$ l) of TE buffer. To the mixture was

added 1  $\mu$ l RNase A/T1 (Fermentas, Hanover, MD) followed by incubation at 37°C for 1 hour. The DNA concentration was determined by using a NanoSpec spectrophotometer (NanoDrop, Wilmington, DE) and then diluted to 100 pmole/ $\mu$ l for use in PCR applications.

## 1.2.2 Cyanobacterial culture

### 1.2.2.1 Z+G Media<sup>29</sup>

Solution 1: 0.590 g  $\text{Ca}(\text{NO}_3)_2 \cdot 4 \text{H}_2\text{O}$  and 0.476g  $\text{NaNO}_3$  in 100 ml  $\text{dH}_2\text{O}$

Solution 2: 0.410 g  $\text{K}_2\text{HPO}_4 \cdot 3 \text{H}_2\text{O}$  in 100 ml  $\text{dH}_2\text{O}$

Solution 3: 0.250 g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  in 100ml  $\text{dH}_2\text{O}$

Solution 4: 1.680 g  $\text{NaHCO}_3$  in 100 ml  $\text{dH}_2\text{O}$

Solution 5: 1.145 g  $\text{Na}_2\text{EDTA}$  and 0.300 g  $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$  in 100 ml  $\text{dH}_2\text{O}$ .

The solution was boiled for 5 minutes in a microwave and the diluted up to 1 L with  $\text{dH}_2\text{O}$ . Solution was then autoclaved for 20 min at 15 lbs./sq. in.

Solution 6: 0.310 g  $\text{H}_3\text{BO}_3$ , 0.169 g  $\text{MnSO}_4$ , 0.009 g  $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24} \cdot 4 \text{H}_2\text{O}$ , 0.029 g  $\text{ZnSO}_4 \cdot 7 \text{H}_2\text{O}$ , 0.015 g  $\text{Co}(\text{NO}_3)_2 \cdot 6 \text{H}_2\text{O}$  and 0.013 g  $\text{CuSO}_4 \cdot 5 \text{H}_2\text{O}$  in 100 ml  $\text{dH}_2\text{O}$

10 ml of solutions 1 – 5 and 80  $\mu$ l of solution 6 was added to an autoclavable vessel. The mixture was diluted up to 1 L with distilled water and the media autoclaved for 20 min at 15 lbs/sq. in. The media was cooled to room temperature before use.

### 1.2.2.2 BG-11 media<sup>30</sup>

Solution 1: 15.0 g  $\text{NaNO}_3$  in 100 ml  $\text{dH}_2\text{O}$

Solution 2: 2.0 g  $\text{K}_2\text{HPO}_4$  in 50ml  $\text{dH}_2\text{O}$

Solution 3: 3.75 g  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  in 50 ml  $\text{dH}_2\text{O}$

Solution 4: 1.80 g  $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$  in 50 ml  $\text{dH}_2\text{O}$

Solution 5: 0.30 g Citric acid in 50 ml  $\text{dH}_2\text{O}$

Solution 6: 0.30 g Ammonium ferric citrate in 50 ml  $\text{dH}_2\text{O}$

Solution 7: 0.05 g  $\text{Na}_2\text{EDTA}$  in 50 ml  $\text{dH}_2\text{O}$

Solution 8: 1.00 g  $\text{Na}_2\text{CO}_3$  in 50ml  $\text{dH}_2\text{O}$

Solution 9: 2.86 g  $\text{H}_3\text{BO}_3$ , 1.81 g  $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$ , 0.22 g  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ , 0.39 g  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ , 0.08 g  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ , 0.05 g  $\text{Co}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$  in 1L  $\text{dH}_2\text{O}$

In an autoclavable vessel 10ml of solution 1 and 1 ml solutions 2-8 were mixed. The volume was then adjusted to 1 L with dH<sub>2</sub>O. The media was autoclaved for 20 min at 15 lbs./sq. in. The media was cooled to room temperature and before use 1 ml solution 9 was added using sterile technique.

### 1.2.2.3 Cultivation of alga

Cultures of *H. hibernicus* BZ-3-1 were cultured under conditions described previously<sup>31</sup> with slight modifications. The cyanophyte was cultured in autoclaved 20-L glass carboys containing an inorganic medium (Z+G or BG-11 media). Cultures were continuously illuminated at an incident intensity of 80-100  $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$  (photosynthetically active radiation) from banks of cool-white fluorescent tubes and sparged with 1% carbon dioxide in air at a rate of 4.5 L/min. The temperature was maintained at  $23 \pm 1$  °C and the culture was harvested after 30 d. The yield was 1.25 g/l dry cell mass.

### 1.2.3 *E. coli* culturing

#### 1.2.3.1 Antibiotics

Ampicillin, sodium salt (Fisher Scientific), kanamycin (Fisher Scientific) and chloramphenicol (Fisher Scientific) were prepared according to standard protocols<sup>26</sup> and used at the following concentrations for the plasmids/strains listed below.

Plasmid	Antibiotic (conc.)
pDrive	Amp (100 $\mu\text{g/ml}$ )
pUC19	Amp (100 $\mu\text{g/ml}$ )

<i>E. coli</i> strain	Antibiotic (conc.)
DH5 $\alpha$	None
BL21(DE3)	None

### **1.2.3.2 Agar plate preparation**

LB media plus agar was prepared according to standard protocols.<sup>26</sup> After autoclaving and cooling to less than 55°C antibiotics were added to the appropriate concentration. The plates were then poured and allowed to harden and were stored in a sealed bag at 4°C until use. For cases where X-Gal and IPTG were necessary for blue/white screening the plates were dried under a laminar flow hood for 45 min and then 40 µl of both X-Gal (2g/ml in DMF) and IPTG (1M in dH<sub>2</sub>O) were spread evenly around the plate and allowed to adsorb for 15 min. Then using a glass spreader the cells were applied to the agar plate.

## **1.2.4 Recombinant DNA procedures**

### **1.2.4.1 Ligation using T4 DNA Ligase**

T4 DNA ligase was purchased from Fermentas (Hanover, MD) and stored at –20°C until use. The ligation reaction was performed in a 0.2 ml thin walled PCR tube at 4°C for at least 12 hours. The vector DNA and insert DNA were combined in a ratio of approximately 1:3 as judged by comparison to a known concentration of DNA on an agarose gel. For ligations utilizing DNA cut with restriction enzymes that yield overhanging ends, 1 µl of 10X buffer (supplied by the company) which contained 10mM ATP and 1 µl of T4 DNA ligase (1 Weiss unit) were added to the tube to make the total volume 10 µl. In the case of ligation of blunt end DNA 1µl of 10X buffer, 1 µl PEG 4000 (final concentration of PEG: 5%) and 1 µl of T4 DNA ligase (1 Weiss unit) were added to a final volume of 10 µl.

#### 1.2.4.2 Ligation into pDrive vectors

pDrive TA cloning kits were purchased from Qiagen (Valencia, CA) for use in cloning PCR products. The supplied components, 2X ligation master mix (5  $\mu$ l) and pDrive cloning vector (1  $\mu$ l) were aliquoted into 0.2 ml thin walled PCR tube on ice. The aliquots were then stored at  $-20^{\circ}\text{C}$  until use. The aliquots were removed from the freezer and thawed by holding the tube between two fingers. When defrosted the tube was placed on ice and the PCR product (4  $\mu$ l) was added to the tube and mixed gently by pipetting 2-3 times. The ligation mixtures were then stored at  $4^{\circ}\text{C}$  for at least 12 h.

#### 1.2.4.3 PCR using *Taq* DNA polymerase from Qiagen

For PCR reactions the following mixture was used unless otherwise stated.

Reagents were added in this order.

36.25  $\mu$ l dH<sub>2</sub>O  
 5  $\mu$ l 10X PCR buffer (Qiagen, Valencia, CA)  
 3  $\mu$ l MgCl<sub>2</sub> solution (25mM)  
 1.5  $\mu$ l dNTPs (10 mM each)  
 1.5  $\mu$ l Primer 1 (10 pmol/ $\mu$ l)  
 1.5  $\mu$ l Primer 2 (10 pmol/ $\mu$ l)  
 0.25  $\mu$ l *Taq* DNA polymerase (1 unit)  
 1  $\mu$ l DNA (100 pmol/ $\mu$ l)  


---

 $\Sigma$  50  $\mu$ l total volume

#### 1.2.4.4 PCR using Advantage 2 PCR kit (BD Biosciences, Mountainview, CA)

For PCR reactions the following mixture was utilized. Reagents were added in this order.

40  $\mu$ l dH<sub>2</sub>O  
 5  $\mu$ l Advantage 2 PCR 10X buffer  
 1  $\mu$ l 50X dNTP solution

1 $\mu$ l Primer 1 (10 pmol/ $\mu$ l)
1 $\mu$ l Primer 2 (10 pmol/ $\mu$ l)
1 $\mu$ l DNA polymerase mix
1 $\mu$ l DNA (100 pmol/ $\mu$ l)
$\Sigma$ 50 $\mu$ l total volume

#### 1.2.4.5 Gel electrophoresis for DNA analysis

Agarose (Fisher Bioscience) was dissolved in 1X TAE buffer by use of a microwave oven. Agarose gels are expressed as percent weight agarose per volume TAE buffer (i.e. 0.8 g agarose dissolved in 100 ml TAE buffer results in a 0.8% agarose gel). Prior to use, EtBr (Bio-Rad) was added to a final concentration of 0.4  $\mu$ g/ml. The gels were developed in a submersion electrophoresis chamber containing 1X TAE buffer. The DNA was then visualized using a Transilluminator (Fisher Biotech).

#### 1.2.5 DNA sequencing

Plasmids were sequenced at the Greenwood molecular biology facility using standard Sanger sequencing techniques. Sequences were viewed and analyzed using Chromas v1.45. Plasmids were sequenced using the vector specific primers pDriveseq+ (5'-ACGACGTTGTAACGACGG) and pDriveseq- (5'-ACACAGGAAACAGCTATGAC)

#### 1.2.6 Sequencing of the *mcy* cluster

##### 1.2.6.1 Gene amplification of *mcyA*, -B, -C, and -G

The nucleotide sequences for *mcyA*, -B, -C, and -G from *M. aeruginosa* PCC 7806 (accession number AF183408), *P. agardhii* CYA 126/8 (accession number

*AJ441056*) and *A. circinalis* sp. 90 (accession number *AJ536156*) were obtained from Entrez nucleotide (<http://www.ncbi.nlm.nih.gov/>). The genes were aligned using the MultAlin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>) algorithm<sup>32</sup> and conserved regions were located. Degenerate primer mixes were then designed for the 5' and 3' end of the genes (Table 2). These mixes had calculated T<sub>m</sub> values with ranges of 60-69°C and the degeneracy was kept as low as possible and below 100 except for mcyH5fwd (228-fold degeneracy). The genes were amplified using the Advantage 2 PCR kit (BD Biosciences, Mountainview), using the supplied protocol. The program used was 95°C-1 min then 35 cycles of 95 °C-30 sec, 68 °C-6 min followed by 68 °C-6 min. The reaction mixtures were then applied to a 0.8% agarose gel and visualized using EtBr. The resulting bands were excised from the gel and recovered using the QiaQuick Gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol. The DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min to inactivate the ligase and the mixture was transformed into *E. coli* DH5α cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C. Positive clones were grown overnight in 3ml LB<sub>Amp</sub> media in an incubated shaker and the plasmids were isolated using the High pure plasmid isolation kit (Roche). Plasmids were sequenced with the vector specific primers pDriveseq+/ pDriveseq- to ensure that they contained the desired gene fragment. The gene fragments were then sequenced as described below.



### 1.2.6.2 Amplification of *mcyD* and *-E*

The nucleotide sequences for *mcyD* and *-E* from *M. aeruginosa* PCC 7806 (accession number *AF183408*), *P. agardhii* CYA 126/8 (accession number *AJ441056*) and *A. circinalis* sp. 90 (accession number *AJ536156*) were obtained from Entrez nucleotide (<http://www.ncbi.nlm.nih.gov/>). The genes were aligned using the MultAlin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>) algorithm<sup>32</sup> and conserved regions were located. At the 5' and 3' ends of the genes degenerate primer mixes were designed to amplify 1kb fragments of the gene (Table 2). The primer mixes had T<sub>m</sub> ranges of 54-59°C and a calculated degeneracy of less than 40. The 1kb fragments were amplified using *Taq* DNA polymerase (Qiagen, Valencia, CA) with the following program: 95°C-3min, 40 cycles of 95°C-30 sec, 55°C-30 sec, 72°C-1 min followed by 72°C-2 min. The products were purified over a 1% agarose gel and visualized with ethidium bromide. The bands were removed from the gel and isolated with the QiaQuick gel elution kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol. The DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min and the mixture was transformed into *E. coli* DH5α cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C. Positive clones were grown overnight in 3 ml LB<sub>Amp</sub> media in an orbital shaker and the plasmids were isolated using the High pure plasmid isolation kit (Roche). Plasmids were sequenced with the vector specific primers pDriveseq+/ pDriveseq- to ensure that they contained the desired gene fragment. The sequence was confirmed by homology to the known *mcy* genes by using the blastx search function provided by NCBI (<http://www.ncbi.nlm.nih.gov/BLAST/>). Once the sequence of the 1kb gene fragments

was known, gene specific primers were designed (Table 2). These primers had a calculated  $T_m$  of 66-68°C and overlapped with the known sequence by approximately 125bp. The unknown sequence of the gene was amplified using the Advantage 2 PCR kit (BD Biosciences, Mountainview, CA) with the following program: 95°C-1 min, 35 cycles 95°C-30 sec, 68°C-6 min followed by 68°C-6 min. The bands were purified using a 0.8% agarose gel and isolated with the QiaQuick gel elution kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol. The DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min and the mixture was transformed into *E. coli* DH5 $\alpha$  cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C. Positive clones were grown overnight in 3 ml LB<sub>Amp</sub> media in an orbital shaker and the plasmids were isolated using the High pure plasmid isolation kit (Roche). Plasmids were sequenced with the vector specific primers pDriveseq+/ pDriveseq- to ensure that they contained the desired gene fragment. The gene fragments were then sequenced as described below.

### 1.2.6.3 Amplification of *mcyH* and *-J*

The nucleotide sequences for *mcyH* and *-J* from *M. aeruginosa* PCC 7806 (accession number AF183408), *P. agardhii* CYA 126/8 (accession number AJ441056) and *A. circinalis* sp. 90 (accession number AJ536156) were obtained from Entrez nucleotide (<http://www.ncbi.nlm.nih.gov/>). The genes were aligned using the MultAlin (<http://prodes.toulouse.inra.fr/multalin/multalin.html>) algorithm<sup>32</sup> and conserved regions were located. Degenerate primer mixes were then designed for the 5' and 3' end of the

genes (Table 2). These primer mixtures had calculated  $T_m$  values within the range of 65-69°C and the degeneracy was kept as low as possible. The genes were amplified using the *Taq* DNA polymerase (Qiagen, Valencia, CA). The program used was 95°C-3 min then 40 cycles of 95 °C-30 sec, 65 °C-30 sec, 72°C-1 min followed by 72 °C-2 min. The reaction mixtures were then applied to a 1.0% agarose gel and visualized using EtBr. The resulting bands were excised from the gel and recovered using the QiaQuick Gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol. The DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min to inactivate the ligase and the mixture was transformed into *E. coli* DH5 $\alpha$  cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C. Positive clones were grown

**Table 2:** Degenerate primer sequences used in the amplification *mcyA*, *-B*, *-C*, *-D*, *-E*, *-G*, *-H*, and *-J*

Name	Sequence
<i>mcyA5fwd</i>	5'-AGTTTTGATGTTTCRGTVTGGGAATTCTTTTG
<i>mcyA3rev</i>	5'-CTGTTGATTYAAGTKSGCHACKGGAAAATC
<i>mcyB5fwd</i>	5'-CCATGCARSARGGGATGCTCTTTCATAG
<i>mcyB3rev</i>	5'-CCAAANSRARGCTAATAAAGCYTCTTGTGKATTAAA
<i>mcyC5fwd</i>	5'-GCATGGCAAATCAGTGGTTAGAAAGTGA
<i>mcyC3rev</i>	5'-GAATWCCARAGAAAATTAAGTATTCCCANACTTGAT
<i>mcyD5SRfwd</i>	5'-ATTCCTTAYCTTTTRAATCTYAATGG
<i>mcyD5SRrev</i>	5'-GTTCCYGATGCTCCAAAAGA
<i>mcyD3SRfwd</i>	5'-TGGGGATTAGGAGCAGTNATT
<i>mcyD3SRrev</i>	5'-CCTAAACTTTGYCGTTGCTG
<i>mcyDLR5fwd</i>	5'-GCGATCGCTGATGCTTTGTGTGTAA
<i>mcyDLR3rev</i>	5'-GGCGTAATCACTACTGAAAGTATGAAGTCTTGG
<i>mcyE5SRfwd</i>	5'-ATGGAGGYNATGCAGGCTCG
<i>mcyE5SRrev</i>	5'-CCRAARGAACTNACYCCAGC
<i>mcyE3SRfwd</i>	5'-GATAACTTYTTTGAAATYGGCGG
<i>mcyE3SRrev</i>	5'-GWKMTTCTAAATCRGGATGATTTC
<i>mcyG5fwd</i>	5'-AAGAYTTTATTCCYGCACNTGGGCTTGTTTT
<i>mcyG3rev</i>	5'-GARTCAATCCCCATTTCTRWNAATCCTTGTTT
<i>mcyH5fwd</i>	5'-GATARGATTASKCAAGARGAAYTKGAATCYAAAGC
<i>mcyH3rev</i>	5'-GCCAACTAGARTYYTSTRAAAGTTCYARMACCC
<i>mcyJ5fwd</i>	5'-ATCTTCABGATCCYAATDTTCCCAATGGCT
<i>mcyJ3rev</i>	5'-AAAAGATAATCRTCAAHGCCATAAACCCAMCC

R - A or G, M - A or C, S - G or C, W - A or T, K - G or T, Y - C or T, B - C, G or T, D - A, G or T, V - A, C or G, H - A, C or T, N - A, G, C or T

overnight in 3ml LB<sub>Amp</sub> media in an orbital shaker and the plasmids were isolated using the High pure plasmid isolation kit (Roche). The inserts were sequenced with the vector specific primers pDriveseq<sup>+</sup>/ pDriveseq<sup>-</sup>. This section was performed in triplicate to ensure proper sequencing.

#### 1.2.6.4 Sequencing of *mcyA*, *-B*, *-C*, *-D*, *-E*, and *-G*

To expedite sequencing, the genes were fragmented and sub-cloned into the pUC19 vector. Plasmids containing the individual genes (10 µg DNA in 100 µl) were sonicated in a Bachman ultrasonic cleaner in 10 sec intervals until the plasmid was sheared as observed by gel electrophoresis. The sheared DNA was separated over a 1.2% agarose gel and the region at approximately 1kb was excised from the gel and purified using the QiaQuick gel extraction kit. The isolated DNA was then blunted by use of a Klenow fragment *exo*- (Fermentas, Hanover, MD) according to the manufacturer's protocol. The now blunt ended DNA was ligated into the pUC19 plasmid, which had been previously restricted by *Sma*I and dephosphorylated (New England Biolabs), using T4 DNA ligase. The ligation mixture was transformed into *E. coli* DH5α cells and plated on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG. Colony PCR was performed on a variable number of colonies for each gene (number of colonies = kb DNA to be sequenced x 5). Positive clones were grown overnight in 1.5 ml LB<sub>Amp</sub> media in sterile micro-centrifuge tubes and the plasmids were isolated by hand as described above. The inserts were sequenced using the vector specific primers pDriveseq<sup>+</sup> and pDriveseq<sup>-</sup>. The appropriate sequences, as determined by sequence homology with known *mcy* genes according to blastx (<http://www.ncbi.nlm.nih.gov/BLAST/>) were then assembled using the ContigExpress function of VectorNTI 9.0<sup>33</sup> with the following parameters, pair wise

assembly algorithm with a minimum overlap of 20bp, minimum identity of 0.85 and minimum cutoff score of 40. Primer walking of the original pDrive plasmid filled in any gaps utilizing sequencing primers that had a calculated  $T_m$  of 57°C and overlapped the known sequences by approximately 125bp.

#### 1.2.6.5 Orientation of the *mcy* gene cluster

Once the 5' and 3' sequences of the genes had been established, gene specific primers that were aimed toward the extremities were designed (Table 3). These primers had a calculated  $T_m$  of 57-58°C and overlapped known sequences by approximately 100bp. A variety of combinations based on known cluster orientations were attempted. PCR products were observed on 0.8% gels, (*mcyA-G*, E-H) and 1.2% gels (all others). The DNA was purified using the QiaQuick Gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol. The DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min and the mixture was transformed into *E. coli* DH5 $\alpha$  cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C.

**Table 3: Primers used during the orientation PCR reactions**

Name	Sequence*
<i>mcyA5revseq</i>	5'-CTATCATCATTAACGACGACAACATAGT
<i>mcyA5rev</i>	5'-GATAGTAGTAATTTGCTGCTGTTGTATCA
<i>mcyA3fwd</i>	5'-TTACAAGTAACTGGAGCTACAGTCG
<i>mcyB5rev</i>	5'-ACTTTTTCCCAAGCTCGCTG
<i>mcyB3fwd</i>	5'-TGCACTTGGTAGATTTCCGTTATT
<i>mcyC5rev</i>	5'-ATTGCGGTCTAGGAAAATCGG
<i>mcyD5rev</i>	5'-CAAAGCATCAGCGATCGC
<i>mcyD3fwd</i>	5'-CCAAGAACTTCATACTTTCAGTAGTGA
<i>mcyE5rev</i>	5'-GACAGTAATTCCAATAAGTCTCAGGAG
<i>mcyE3fwd</i>	5'-CATCGAGATTACAAGCGCAA
<i>mcyG5rev</i>	5'-TCTAAAATCTGCCAGACATTGTAAG
<i>mcyG3fwd</i>	5'-GGAGAACGTCGAGAAAAATTATTAGA
<i>mcyH5rev</i>	5'-TTTCGAGGAGATTATTAATCGGCGCAG
<i>mcyJ5rev</i>	5'-GGGAACATTAGGATCGTGAAGAT
<i>mcyJ3fwd</i>	5'-TTTATGGCTTTTGACGATTATCTTT

\*: all 5rev primers were created as the reverse compliment to the coding strand with the exception of *mcyArevseq*, which is simply the reverse sequence of the coding strand

Positive clones were grown overnight in 3ml LB<sub>Amp</sub> media in an orbital shaker and the plasmids were isolated using the High pure plasmid isolation kit (Roche). The inserts were sequenced with the vector specific primers pDriveseq+/ pDriveseq- and sequencing by primer walking filled in any gaps.

#### 1.2.6.6 Completing *mcyH* and *-C*

Because of the high nucleotide similarity between our cluster and the cluster sequenced from *A. circinalis* sp. 90, the 3' areas of *mcyH* and *-C* were amplified using gene specific primers (Table 4) derived from the *A. circinalis* gene cluster. *Taq* DNA polymerase (Qiagen, Valencia, CA) was used to amplify the products using the following program, 94°C-3 min, 40 cycles 95°C-30 sec, 57°C-30 sec, 72°C-1 min, then 72°C-2 min. The DNA was purified over a 0.8% agarose gel visualizing with EtBr. The resulting bands were excised from the gel and recovered using the QiaQuick Gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol, the DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min and the mixture was transformed into *E. coli* DH5 $\alpha$  cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C. Positive clones were grown overnight in 3ml LB<sub>Amp</sub> media at 37°C at 250 rpm and the plasmids were isolated using the High pure plasmid isolation kit (Roche). The inserts were sequenced with the vector specific primers pDriveseq+/ pDriveseq-.

To ensure that the stop codon was indeed true a vector assisted PCR was performed. Cyanobacterial DNA was completely digested with *Sau3AI* (Fermentas,

Hanover, MD) using the buffer supplied by the manufacturer and an appropriate number of units of enzyme. The DNA was purified using of the QiaQuick PCR purification kit (Qiagen, Valencia, CA) according to the supplier's protocol. The restricted DNA was ligated into a pUC19 vector that had been digested with *Bam*HI and dephosphorylated (New England Biolabs) using T4 DNA ligase (Fermentas, Hanover, MD) under the following conditions.

6.5  $\mu$ l genomic DNA *Sau*3AI restricted (150 ng)

0.5  $\mu$ l pUC19 (250 ng)

1 $\mu$ l 10X T4 DNA ligase buffer (Fermentas)

2  $\mu$ l T4 DNA ligase (2 Weiss units)

The ligation was stored at 4°C for 18 h and then the DNA was purified by use of a QiaQuick PCR purification kit according to the manufacturer's protocol.

Gene specific primers were designed as close to the 3' end of the gene as possible albeit without including the stop codon (Table 4). A standard PCR reaction utilizing *Taq* DNA polymerase was then performed with one modification, the vector specific primer (pDriveseq+) was used at 100 pmol/ $\mu$ l stock concentration.

The PCR reaction mixture was applied to a 2% agarose gel and visualized using EtBr. The resulting bands were excised from the gel and recovered using the QiaQuick Gel extraction kit (Qiagen, Valencia, CA) according to the manufacturer's supplied protocol, The DNA was then ligated into the pDrive vector (Qiagen, Valencia, CA). The ligation mixture was heated at 65°C for 10 min and the mixture was transformed into *E. coli* DH5 $\alpha$  cells. The cells were plated out on LB<sub>Amp</sub> agar plates containing X-Gal and IPTG and incubated overnight at 37°C. Positive clones were grown overnight in 3ml

LB<sub>Amp</sub> media at 37°C at 250 rpm and the plasmids were isolated using the High pure plasmid isolation kit (Roche). The inserts were sequenced with the vector specific primers pDriveseq+/ pDriveseq-.

**Table 4: Primers used during the completion of *mcyC* and *-H***

<b>Name</b>	<b>Sequence*</b>
<b>mcyH AnabaenaRev</b>	<b>5'-CAAAGATTATCGAAATAAAAAAAAAATAGTTTATAA</b>
<b>mcyC AnabaenaRev</b>	<b>5'-TCAATTCTGACTAAGAATCGCATTATAA</b>
<b>mcyH vectorFwd</b>	<b>5'-AATGGGTTCTGGAACITTCACAAGATTC</b>
<b>mcyC vectorFwd</b>	<b>5'-ATGATTGATCCTCCCTATCTCAATCAAAC</b>



## 1.3 RESULTS AND DISCUSSION

### 1.3.1 Amplification/Identification of the *mcy* gene cluster

Since three microcystin gene clusters had been fully characterized at the beginning of this investigation we chose to take a unique approach utilizing PCR to sequence the microcystin gene cluster from *H. hibernicus*. The known nucleotide sequences from *mcyA*, *-B*, *-C*, *-D*, *-E*, *-G*, *-H*, and *-J* were aligned to reveal conserved nucleotide regions which were then used to derive degenerate primer mixes as described in the experimental section. Six of the primer pairs (*mcyA*, *-B*, *-C*, *-H*, *-G*, *-J*) gave the expected amplicons, however the primer pairs for the two largest PKS genes (*mcyD*, *-E*) did not yield a PCR product. One could attribute the lack of PCR products simply to the fact that the desired amplicon was too large (10kb and 11kb respectively). In comparison, the largest fragment that was amplified successfully was *mcyA* (9kb). Another explanation is that the degenerate primer mixes that we used were located in conserved regions of the PKS. This would prevent the desired amplicon from being replicated by scattering the primers throughout the genomic DNA. To circumvent this problem 1kb stretches from the 5' and 3' regions of *mcyD* and *-E* were amplified and sequenced. Sequence specific primers were then created to amplify the entire gene while overlapping approximately 150 bp with the known sequence. This approach yielded PCR products of the appropriate size that upon sequencing were verified to be the *mcyD* and *mcyE* gene fragments.

The previously published gene clusters contained three other genes of which two, *mcyF*, and *-I*, have been found in two of the three clusters,<sup>16,18</sup> while *mcyT* was discovered only in the *Planktothrix* cluster.<sup>17</sup> It was therefore decided that these genes

would be sequenced during the PCR reactions used for joining the gene fragments or discovered during the sequencing of the cluster's flanking regions.

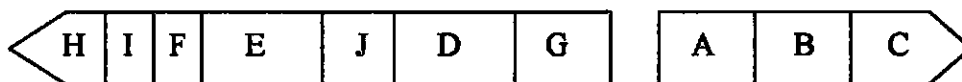
### 1.3.2 Organization of the microcystin gene cluster

After the genes were amplified and sequenced, gene specific primers were designed to amplify regions between the known sequences. PCR reactions were tried in the order and orientation known from the other clusters. The previously sequenced gene clusters all contained *mcyB* flanked by *mcyA* at the 5' end and *mcyC* at the 3' end. Since this was observed to be the only commonality in gene orientation between the known clusters, it was assumed to be present in the *H. hibernicus* cluster and therefore was the first set of PCR reactions attempted. The next step was to establish the orientation of the other genes. The next three PCR reactions that were attempted were between the primer pairs *mcyA5revseq/mcyD5rev*, *mcyA5revseq/mcyG5rev* and *mcyA5rev/mcyH3fwd* which mimic the arrangement of the genes in *Microcystis*, *Anabaena* and *Planktothrix*, respectively. Only the primer pair *mcyA5revseq/mcyG5rev* gave an amplicon of the appropriate size that upon sequencing was confirmed to be the desired product. To limit the amount of PCR reactions that were performed the next set was modeled after the arrangement in the *Anabaena* cluster. All the reactions afforded predominant bands of the expected size. Given the similarities between the *H. hibernicus* and *A. circinalis* clusters it was anticipated that *mcyF* and *-I* would be located between *mcyE* and *-H* and this was discovered to be the case during sequencing. The PCR reactions established the order and orientation of the genes as seen in Figure 5. The resulting sequences were assembled using the ContigExpress platform in the VectorNTI software. Each of the original gene

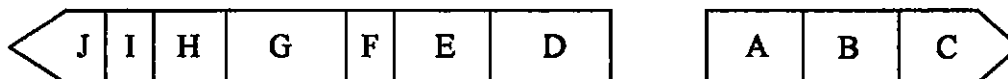
fragments amplified from the degenerate primer mixtures were assembled into contigs first. These contigs along with the sequences from the second round of PCR were then assembled into the gene cluster.

The gene cluster is 55.5kb in length and was discovered to contain two putative operons. The first operon (*mcvABC*) is transcribed in the opposite direction, from a putative bi-directional promoter, as the second operon (*mcvGDJEFIH*). The promoter region is 891 bp in length and is proposed to be bi-directional in nature. Use of BPRM, a web based program for sigma70 promoter recognition, yielded possible -10 and -35 boxes in the analyzed region for each direction.<sup>34</sup>

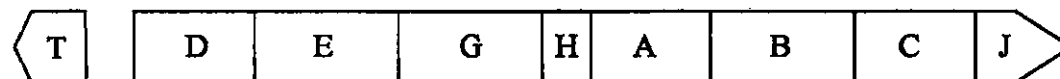
*H. hibernicus*



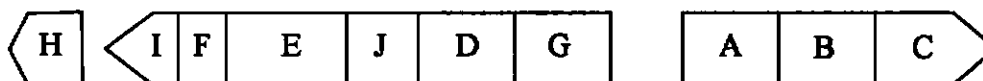
*M. aeruginosa*



*P. agardhii*



*A. circinalis*



*N. spumigena*

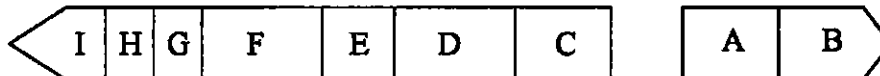


Figure 5: Schematic diagram of the *mcv* cluster sequenced from *H. hibernicus* as well as the known clusters from *M. aeruginosa*, *P. agardhii*, *A. circinalis* and the nodularin gene cluster from *N. spumigena*. Genes not drawn to scale.

### 1.3.3 Characterization of the microcystin genes

Upon assembly of the gene cluster the nucleotide sequences of the individual genes were translated and compared to the corresponding proteins of the known microcystin gene clusters and the nodularin gene cluster using the Basic Local Alignment Search Tool (BLAST) available from the NCBI (Table 5). All genes held significant similarity on both the amino acid and nucleotide level. The percent similarity on the amino acid level ranges from 53% (NdaA-*N. spumigena*) to 92% (McyJ-*A. circinalis*). The lowest similarity to a *mcy* gene was 67% (McyB-*P. agardhii*/McyH-*M. aeruginosa*). The similarity on the nucleotide level ranges from 79% (*mcyD*-*M. aeruginosa*) to 89% (*ndaH*-*N. spumigena*/*mcyJ*-*A. circinalis*).

**Table 5:** Percent similarity of genes from *H. hibernicus* to known *mcy* genes. % homology to amino acid (% homology to nucleotide)

	<i>mcyA</i>	<i>mcyB</i>	<i>mcyC</i>	<i>mcyD</i>	<i>mcyE</i>	<i>mcyF</i>	<i>mcyG</i>	<i>mcyH</i>	<i>mcyI</i>	<i>mcyJ</i>
<i>A. circinalis</i>	77(85)	82(86)	80(85)	78(86)	78(84)	72(80)	79(86)	75(85)	78(85)	92(89)
<i>N. spumigena</i>	53(82)	N/A	77(84)	74(85)	79(85)	79(81)	77(85)	75(84)	83(89)	87(87)
<i>P. agardhii</i>	62(84)	67(80)	75(82)	71(82)	75(84)	N/A	73(80)	70(80)	N/A	84(82)
<i>M. aeruginosa</i>	64(83)	68(81)	70(85)	67(79)	73(81)	69(79)	69(81)	67(80)	69(79)	81(84)

The low homology to NdaA (53%) can be explained by the proposed evolution of the *nda* cluster. Nodularin is missing two amino acid residues when compared to microcystin, D-Ala and the L-X residue. These amino acids are activated by *mcyA*-A2 and *mcyB*-A1 respectively. Moffitt and Neilan proposed that the nodularin cluster arose from a deletion event involving excision of the second CAT from *mcyA* and the first CAT from *mcyB*, which would result in *ndaA*.<sup>35</sup> When *McyA* and *McyB* are compared to *NdaA*, similarity values of 37% for both are obtained. When *NdaA*-A<sub>1</sub>T<sub>1</sub> and *NdaA*-

C<sub>2</sub>A<sub>2</sub>T<sub>2</sub> are compared to the corresponding region from McyA and McyB the similarity is 54% and 73%, respectively. The lower similarity of NdaA-A<sub>1</sub>T<sub>1</sub> can be attributed the fact that nodularin contains Mdhb (derived from Thr) and microcystin contains Mdha (derived from Ser).

The first operon (*mcyABC*) contains 3 ORFs. The proposed ORFs for *mcyA* and *mcyC* begin with an ATG start codon while the ORF for *mcyB* is proposed to begin with a GTG codon. A section of 20bp separates the termination codon of *mcyA* for the start codon of *mcyB* while 2bp separates the TAA of *mcyB* from the ATG of *mcyC*. The lengths of *mcyA*, -*B*, -*C* are 8,388bp, 6,390bp and 3,858bp and encode *apo* proteins of a predicted mass of 314,605 Da, 241,944 Da and 147,174 Da respectively.

The second operon contains the PKS genes (*mcyG*, -*D*, -*E*), the tailoring enzymes (*mcyJ*, -*F*, -*I*) and the putative ABC transporter (*mcyH*). There are 71bp between *mcyG* and *mcyD*, 78 bp separating *mcyD* and *mcyJ*, 35 bp between *mcyJ* and *mcyE*, 47 bp separating *mcyE* and *mcyF*, 35 bp separating *mcyF* and *mcyI* and a 21 bp spacer between *mcyI* and *mcyH*. The lengths of *mcyG*, -*D*, -*J*, -*E*, -*F*, -*I*, -*H* are 7,923 bp, 11,626 bp, 924 bp, 10,392 bp, 766 bp, 1,026 bp and 1,780 bp, respectively, and, correspondingly, encode for *apo* proteins of 294,142 Da, 43,1274 Da, 36,193 Da, 387,621 Da, 28,723 Da, 37,189 Da, 67,498 Da in size. McyG, -D, and -J were most similar to *A. circinalis* with similarities ranging from 69 to 79%, 67 to 78%, 81 to 92%, respectively, while McyE, -F, -I held the highest similarity to *N. spumigena* and had ranges of 73 to 79%, 75 to 79%, and 69 to 83%. McyH was equally similar (75%) to both *A. circinalis* and *N. spumigena*.

The second operon (*mcyGDJEFIH*) from *H. hibernicus* is most similar in arrangement to that of the nodularin gene cluster. In *Anabaena* there is a putative

transcriptional start point for *mcyH* in the 295 bp spacing between *mcyI* and *mcyH*. In *Hapalosiphon*, as in *Nodularia*, that spacing is less than 30bp (21 bp and 23 bp respectively) and no promoter region can be identified. Not surprisingly the homology for McyF, -I and -H is highest when compared to the corresponding *nda* genes (NdaG, -H and -I)

### 1.3.4 NRPS adenylation domain specificities

The substrate specificity conferring amino acid residues of the adenylation domains from *M. aeruginosa*, *P. agardhii*, *A. circinalis* and *H. hibernicus* were analyzed by the method of Stachelhaus *et al.*<sup>36</sup> (Table 6). The signature sequences of *mcyA*-A1, -A-A2, -BA1, -B-A2 and -C-A from *H. hibernicus* are very tightly conserved with respect to the known microcystin signature sequences. Oddly, the signature sequences from *H. hibernicus* and *A. circinalis* for *mcyB*-A1 and *mcyC*-A are identical although they supposedly activate different amino acids. *A. circinalis* has been shown to produce MC-LR and MC-RR while *H. hibernicus* has previously been shown to produce MC-LA. However, further analysis indicates the presence of MC-AR (R. Kurmayer, G. Christiansen, personal communication). The large differences in amino acid side chain size and polarity raise questions as to the discrimination ability/process. Further biochemical studies are needed to address this issue and are currently in progress.

**Table 6:** Amino acid specificity conferring sequences of adenylation domains as Stachelhaus *et al.*<sup>36</sup>

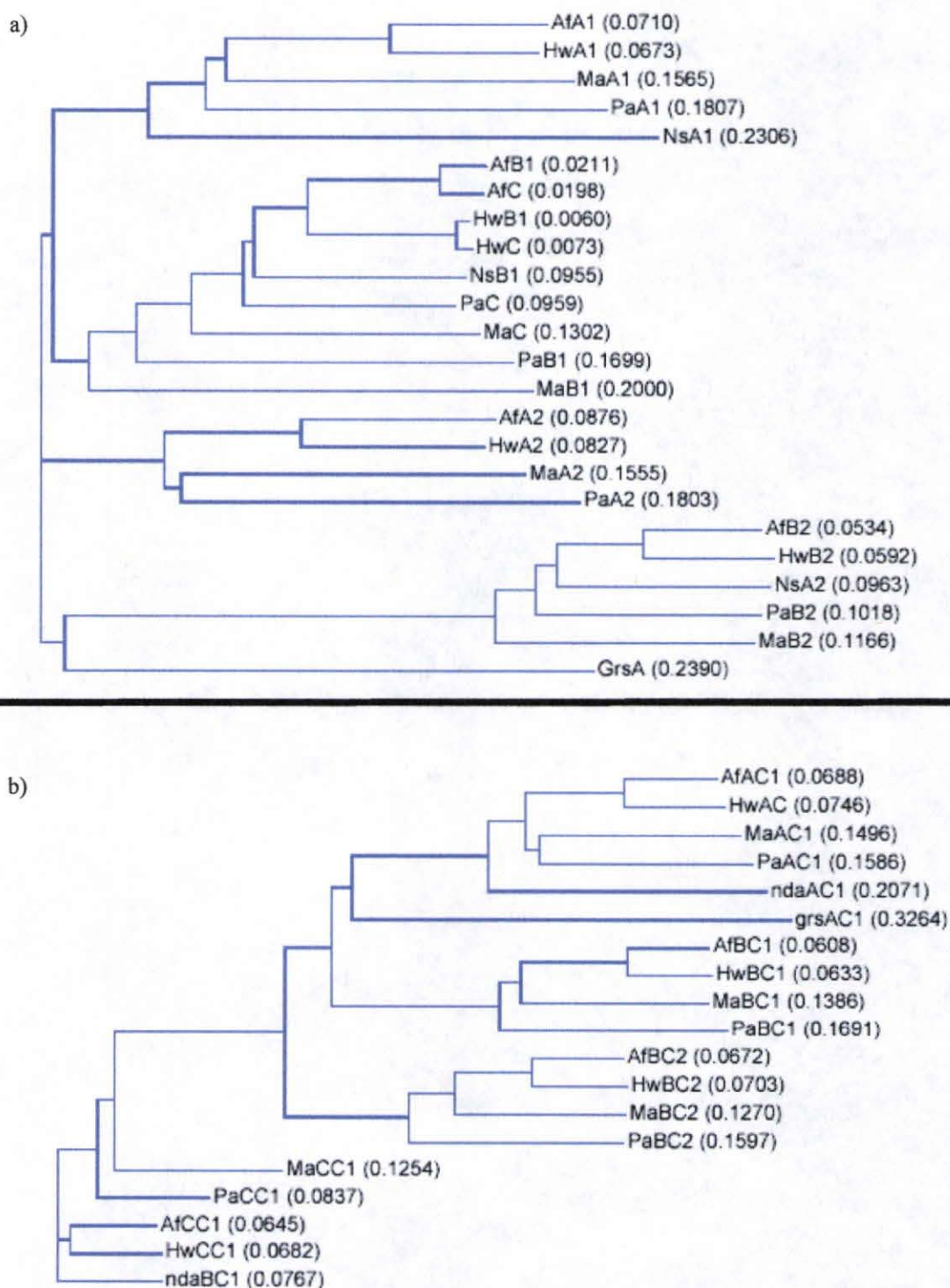
	McyA-A1	McyA-A2	McyB-A1	McyB-A2	McyC-A
<i>H. hibernicus</i>	DVWHISLIV	DLFNNALTS	DVWFFGLVA	DARHIGIFK	DVWIFGLVV
<i>P. agardhii</i>	DVWHISLIV	DLFNNALSS	DALFFGLVA	DPRHVGIFK	DPWIFGLVA
<i>M. aeruginosa</i>	DVWHFSLIV	DLFNNALTS	DAWFLGNVT	DARHVGIFK	DVWIFGAVT
<i>A. circinalis</i>	DVWHISLIV	DLFNNALTS	DVWFFGLVA	DARHVGIFK	DVWIFGLVA

### 1.3.5 Phylogenetic analysis of adenylation domains

Using the nucleotide sequence of the adenylation domains from the conserved motif A1 (L(TS)YxEL) up to and including the amino acid just prior to the conserved motif of the thiolation domain (DxFFxxLGG(HD)S(LI)),<sup>37</sup> a phylogenetic analysis was performed, the results of which can be seen in Figure 6. Most of the adenylation domains cluster together as one would expect due to the proposed linear evolution of the microcystin synthetase gene cluster (i.e. all *mcyA*-A1 domains show higher similarity to each other than to other A domains). However an abnormality can be observed when analyzing *mcyB*-A1 and *mcyC*-A of both *H. hibernicus* and *A. circinalis*. This clustering seems to be the result of a duplication of the *mcyC*-A domain followed by recombination as suggested by the closer relation to *P. agardhii* *mcyC*-A (Figure 6a). However, we find no evidence that the condensation domains have undergone a similar evolution (Figure 6b).

The microcystin gene cluster is purported to be a site of frequent recombination and mutation, which may help explain the wide variety of microcystin variants seen in nature. Also evidence of recombination of small stretches<sup>38</sup> and entire domains<sup>39</sup> has been reported. When aligning the adenylation domain amino acid sequences obtained from *mcyB*-A1 and *mcyC*-A, one can observe only eleven amino acid substitutions in the entire adenylation domain. And more importantly only one amino acid substitution between the conserved motifs A4 and A5, the purported site for amino acid selectivity.<sup>40</sup>





**Figure 6:** (a) Evolutionary tree diagram of the adenylation domains found in the microcystin clusters and the nodularin cluster using GrsA-A1 as an out-group. (b) Evolutionary tree diagram of the condensation domains found in the microcystin clusters as well as the nodularin cluster. Generated using Vector NTI version 9.0



## 1.4 CONCLUSION

A 55.5 kb gene cluster responsible for the biosynthesis of microcystin in *H. hibernicus* was sequenced using a unique degenerate PCR method. It is envisioned that this procedure could be used on any strain that produces microcystins and could be a valuable tool in genetic analysis of the *mcy* clusters. The time necessary for sequencing the *mcy* gene cluster was significantly reduced (approximately 5 weeks compared to a minimum of 1 year for the standard approach using hybridization of genomic libraries). As an additional benefit, the problems associated with the creation of a genomic library were avoided. This method is quick, inexpensive and relies on materials/kits already contained in all molecular biology labs (PCR reagents, cloning vectors, and DNA sequencing equipment). It is envisioned that this technique will be helpful in studying the *mcy* gene cluster in any isolated strain.

## 2. Genetic knockout of *mcyT* in *P. agardhii* 126/8

### 2.1 INTRODUCTION

#### 2.1.1 McyI

The biosynthesis of microcystin requires the formation of Mdha, which is postulated to arise from dehydration of a serine residue. The gene clusters from *Anabaena*, *Microcystis*, *Hapalosiphon* and the nodularin gene cluster from *Nodularia* contain a *D*-3-phosphoglycerate dehydrogenase homolog, *mcyI* (*ndaH*), while it was absent from the sequenced gene cluster found in *Planktothrix*. It has been hypothesized that McyI is responsible for the formation of the Mdha residue. However, as of this date no *in vitro* characterizations of this protein have been reported.

*D*-3-Phosphoglycerate dehydrogenases are known to contain three domains, the catalytic domain, the NAD binding domain and the ACT domain. The ACT domain is where the allosteric inhibitor serine binds to induce a conformational change thereby inhibiting the formation of 3-phosphopyruvate.<sup>41</sup> During the comparison of the amino acid sequence of McyI to known *D*-3-phosphoglycerate dehydrogenases it was observed that the ACT domain was missing in McyI (Figure 7). This suggested that McyI might not be involved in serine biosynthesis but perhaps was the protein that catalyzed Mdha formation instead. With McyI having such high homology to known *D*-3-phosphoglycerate dehydrogenases (42% identity, *Rubrobacter xylanophilus* DSM 9941) it was postulated that the reactive intermediate in dehydration is phosphoserine, which would reduce the energy of activation necessary for dehydration and produce inorganic phosphate as a by-product as seen in the lantibiotics.

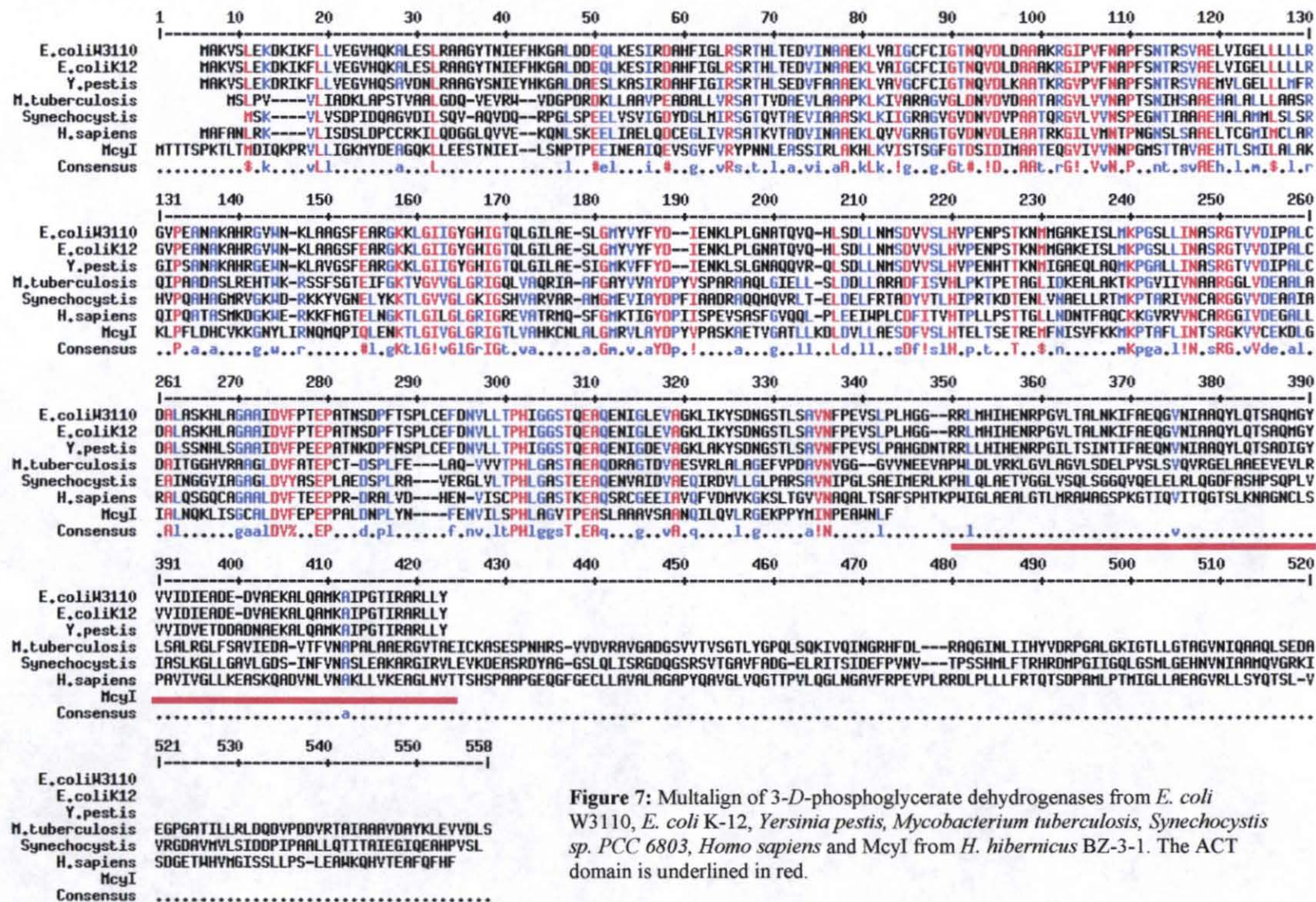


Figure 7: Multialign of 3-D-phosphoglycerate dehydrogenases from *E. coli* W3110, *E. coli* K-12, *Yersinia pestis*, *Mycobacterium tuberculosis*, *Synechocystis* sp. PCC 6803, *Homo sapiens* and *McyI* from *H. hibernicus* BZ-3-1. The ACT domain is underlined in red.

The lantibiotics are a group of ribosomally produced peptide antibiotics that are characterized by the presence of lanthionine (Lan) and/or methyllanthionine (MeLan). These residues are formed by post-translational dehydration of either serine or threonine, followed by Michael addition of the thiol group of a cysteine. The initial step in the dehydration involves the transfer of the gamma phosphate group of ATP onto the serine/threonine hydroxyl group resulting in phosphoserine and ADP. With the hydroxyl group activated, the dehydration can occur with a lower activation barrier. The dehydration and cyclization can be catalyzed by a pair of proteins as in the case of nisin, where NisB and NisC catalyze the dehydration and cyclization respectively,<sup>42</sup> or by a single enzyme as in the case of lactacin 481, where LctM catalyzes both reactions. Interestingly, the C terminus of LctM showed some homology to LanC but the N terminus showed no homology to LanB making *in silico* prediction of the dehydratase activity impossible.<sup>43</sup>

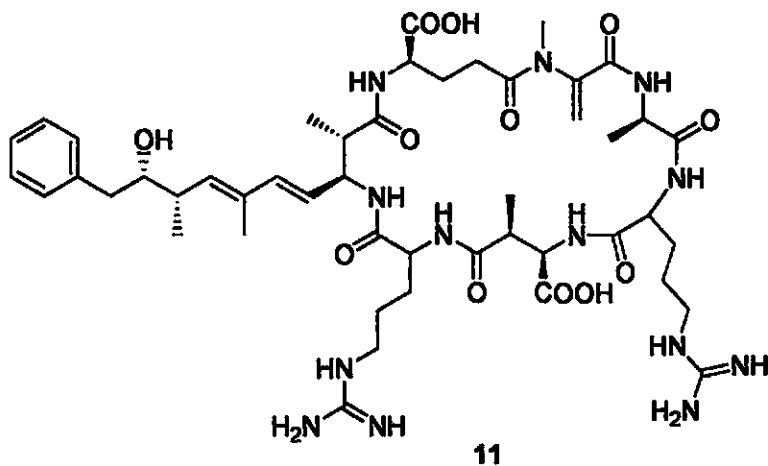
### 2.1.2 McyT

Since *mcyl* is absent from the *Planktothrix* gene cluster efforts were undertaken to locate it at a different locus in the genome. These attempts, which utilized PCR and degenerate primers, were unsuccessful (results not shown). This observation left the question of how the *Planktothrix* genus synthesizes the Mdha residue unanswered. While lacking *mcyl*, *Planktothrix* does contain *mcyT*, a type II thioesterase homolog (68% identity, *Nostoc punctiforme* PCC 73102).

Type II thioesterases are known to have two functions *in vitro*. They are known to cleave either misprimed amino acids or acetyl groups tethered to the 4'-phosphopantetheinyl (4'-PP) arm of thiolation domains.<sup>44</sup> In both cases the thioesterases

are responsible for binding and interacting with thioesters, which they cleave hydrolytically to generate the free thiol and acid. Therefore it is conceivable that the *Planktothrix* genus evolved to form Mdha from O-acetyl-serine convergently.

To probe the function of McyT in the biosynthesis of MC-RR insertional mutagenesis was decided to be the best course of action. Christiansen *et al.*<sup>17</sup> had previously used this technique to inactivate *mcyJ* in CYA 126/8. Analyzing the mutant strain with MALDI-TOF revealed that (*D*-Asp<sup>3</sup>)-MC-RR was no longer present, but a new variant reduced in mass by 14 units was observed. This variant was identified as (DMAdda, *D*-Asp<sup>3</sup>)-MC-RR that contains a desmethoxy-Adda side chain (11). It was envisioned that insertional mutagenesis of *mcyT* might provide insight into the function of McyT by yielding a new structural microcystin variant.



## 2.2 EXPERIMENTAL SECTION

### 2.2.1 Insertional mutagenesis of *mcyT*

#### 2.2.1.1 Construction of the homologous recombination plasmid

The *mcyT* mutant amplicon, consisting of *mcyT*, the bi-directional promoter and the 5'-end of *mcyD*, was amplified using the Advantage 2 PCR kit ® (BD Biosciences) with the primers *mcyT*<sub>fwd</sub> (5'-CGACACTAGTGGATACGGTAGGCGATG) and *mcyT*<sub>rev</sub> (5'-ACTCAGAAAAATTTCCAGCCCTTTCTGC). As expected an amplicon that was 3kb in size was obtained. It was purified over a 0.8% agarose gel and isolated using the QiaQuick Gel extraction Kit according to the manufacturer's protocol. The amplicon was then cloned into the pDrive vector (Qiagen) and transformed into *E. coli* DH5 $\alpha$  cells. The cells were grown overnight at 37°C on an LBA<sub>AMP</sub> plate containing X-Gal and IPTG. Positive colonies were then grown overnight in LB<sub>AMP</sub> media in an orbital shaker and the plasmid was isolated using the High Pure Plasmid Isolation Kit (Roche Applied Sciences). The plasmid was designated pDT-1.

#### 2.2.1.2 Creating the chloramphenicol disrupted *mcyT* gene

After the amplicon was verified via sequencing, the restriction with *PfoI* was performed. This procedure was modified from Parker *et al.*<sup>45</sup>. In a 0.2 ml thin walled PCR tube (Corning) 5  $\mu$ l of pDT-1 in TE buffer, pH 8.0 (500 ng) was combined with 3.8  $\mu$ l dH<sub>2</sub>O, 1  $\mu$ l Buffer Tango (Fermentas), 0.2  $\mu$ l EtBr (1 mg/ml) and 2 units of *PfoI* (Fermentas). The reaction was performed as 16 replicate samples to provide suitable amounts of DNA for the next step. Incubation with the enzyme was performed at 37°C for 20 hours at which time the reactions were run on a 0.8% agarose gel (20 cm, 100V, 45 min) to ensure separation of the singly digested form from the fully restricted form.

The correct band was identified by comparison against pDT-1 linearized with *SacI* and fully restricted with *PfoI* according to the supplier's protocols. The linearized form from all 16 reactions was combined and isolated from the gel using the QiaQuick Gel extraction kit (Qiagen) according to the supplier's protocol. The 5'-overhangs generated by *PfoI* were blunted by exposure to the Klenow *exo-* fragment (Fermentas) according to the manufacturer's protocol. The linearized *mcyT*-pDrive construct was ligated to the chloramphenicol resistance cassette of pACYC184 (NEB). The ligation was accomplished using T4 DNA ligase (Fermentas) as described above. The chloramphenicol resistance cassette was isolated by restriction of pACYC184 using sufficient units of *BsaAI* (New England Biolabs). The 1.2kb fragment was isolated from a 0.8% agarose gel using the QiaQuick Gel extraction Kit (Qiagen). The ligation mixture was transformed into electro-competent *E. coli* DH5 $\alpha$  cells. The cells were grown overnight on LBA plates containing 100  $\mu$ g/ml ampicillin and 50  $\mu$ g/ml chloramphenicol. Colonies were then checked for the appropriate insert location by colony PCR using the primers *mcyT5seq* (5'-GTTAGGTGGGGTAATATGGC) and *mcyT3seq* (5'-CGACTATTCTGCTTTCATTATGC). The positive colony obtained was grown overnight in LB media plus 100  $\mu$ g/ml ampicillin and 50  $\mu$ g/ml chloramphenicol and the plasmid isolated with the High Pure Plasmid Isolation Kit (Roche Applied Sciences) to give plasmid pDTCm-1.

### 2.2.1.3 Preparation of the transformation construct

Plasmid pDTCm-1 was retransformed into *E. coli* DH5 $\alpha$  cells and grown overnight at 37°C on an LB agar plate containing 100  $\mu$ g/ml ampicillin and 50  $\mu$ g/ml

chloramphenicol. Four individual colonies were grown overnight in 5ml LB media plus 100 µg/ml ampicillin and 50 µg/ml chloramphenicol in an orbital shaker. The plasmids were isolated by hand preparation followed by digestion with sufficient units of *SacI*. Two restriction digests were combined and the DNA purified using the QiaQuick PCR purification kit according to the manufacturer's protocol with one exception. The DNA was eluted using two 50 µl aliquots of Buffer EB and stored on ice until use.

#### **2.2.1.4 Transformation of *P. agardhii* CYA 126/8**

*P. agardhii* CYA 126/8 was grown until mid log phase growth (approximately 3 weeks) at 22°C and a light intensity of 80-100 µmol photons m<sup>-2</sup> s<sup>-1</sup> in Z+G media. 50 ml of cell culture was centrifuged down in a 50 ml Falcon tube using the SLA-1500 rotor (11,000 rpm, 15 min, RT). 25 ml of the media was added to a 125 ml Erlenmeyer flask containing 25 ml freshly prepared Z+G media, this was stored for later use. The cells were washed twice with 20 ml 1 mM HEPES (filter sterile) followed by centrifugation (SLA-1500, 11,000 rpm, 15 min, RT). The cells were then resuspended in 1 ml 1 mM HEPES and transferred to a 1.5 ml eppendorf tube and pelleted again, this time using a F-20/Micro rotor (11,000 rpm, 15 min, RT). The supernatant was drained via aspiration and the cells were placed on ice until use.

The linearized DNA was heated at 100°C for 15 min in a heat block followed by cooling on ice. After 5 min on ice the tube was centrifuged for 10 sec to recover the condensed water and then the DNA was stored on ice until use. The cells were then resuspended in the denatured DNA and the slurry was transferred to a 0.2 cm Gene Pulser cuvette (BioRad) (pre-cooled to -20°C). The electroporation was performed on a Gene Pulser II



(BioRad) (1.0 kV, 200  $\Omega$ , 25  $\mu$ F). As quickly as possible after electroporation the cells were removed from the electroporation cuvette and transferred to the 125 ml Erlenmeyer flask containing Z+G media prepared above by washing the cuvette with 1 mL media. The flask was covered with a sheet of white paper and kept under the normal culture conditions. After 24 h the paper was removed and the culture was allowed to grow for an additional 24 h. At this time chloramphenicol was added to a final concentration of 1  $\mu$ l/ml and the culture left alone. Within one week the culture became completely bleached of all color and after five weeks a faint green tint appeared in the flask. The mutant was then sub-cultured two times in Z+G media containing 1  $\mu$ g/ml chloramphenicol to ensure that all the wild type strain had been killed. The mutant strain was named *P. agardhii* CYA 126/8  $\Delta$ *mcyT*.

#### **2.2.1.5 Confirmation of mutant identity**

DNA was isolated from CYA 126/8  $\Delta$ *mcyT* as described above in section 1.2.1.8 and a PCR reaction was performed using the primers *mcyT5seq* and *mcyT3seq* using *Taq* DNA polymerase. The resulting amplicon was compared to wild type DNA and construct DNA amplicons on a 0.8% agarose gel.

#### **2.2.2 Peptide analysis**

##### **2.2.2.1 Peptide extraction of *P. agardhii* CYA 126/8 $\Delta$ *mcyT***

CYA 126/8  $\Delta$ *mcyT* was cultured in 20 L of Z+G media containing 1  $\mu$ g/ml chloramphenicol with air bubbled through at a rate of 3 L/min and a light intensity of 80-100  $\mu$ mol photons  $\text{m}^{-2} \text{s}^{-1}$ . When the cells reached saturation they were concentrated to a

pellet (8.9 g wet weight). The cells were then extracted twice with 50% aqueous methanol (20 ml) by sonicating for 15 min in a Branson 1510 ultrasonic cleaner followed by shaking at 250 rpm at room temperature for 15 min. The debris was removed by centrifugation in an International clinical centrifuge model CL on setting 6 for 10 min. The methanol was removed from the green supernatant at reduced pressure. The residue was then applied to a 4 g ODS-A column (YMC-GEL, 70 mesh) that had been pre-equilibrated with water. The residue was loaded onto the column and then batch eluted using 30 ml each of 25% aqueous methanol, 50% aqueous methanol, 75% aqueous methanol and 100% methanol. The fractions were dried using a SPD Speed Vac (Thermo Savant) without heating. The 50% methanolic fraction yielded 14.1 mg, the 75% methanolic fraction yielded 2.1 mg and the 100% methanol fraction yielded 2.2 mg. Due to the low mass in the 75% methanolic fraction it was combined with the 50% aqueous methanol fraction. The samples were then ready for analysis by HPLC and LC-MS. Extracts of the wild type were prepared in an analogous fashion.

#### **2.2.2.2 HPLC analysis of CYA 126/8 $\Delta$ *mcyT* peptide content**

The combined 50% and 75% aqueous methanol fractions were dissolved in methanol at a concentration of 10 mg/ml and filtered through a 0.22  $\mu$ m filter. The extracts were then analyzed using a Shimadzu LC-10AS HPLC equipped with a HiBar® RT 250-4 column (RP-18, 5  $\mu$ m 5 mm X 22 cm). The sample was run at 1 ml/min flow rate from 30% aqueous acetonitrile plus 0.5% TFA to 100% acetonitrile plus 0.5% TFA over 45 min and observed at 210 and 240 nm. The (*D*-Asp<sup>3</sup>)-MC-RR was a large peak with a retention time of 14.27 min in the wildtype extract.

### 2.2.2.3 LC-MS analysis of CYA 126/8 $\Delta$ *mcyT* peptide content

The combined 50% and 75% aqueous methanol fractions (conc.: 10 mg/ml) for CYA 126/8  $\Delta$ *mcyT* and wild type were analyzed using a Agilent 1100 LC-MS equipped with a HiBar® RT 250-4 column (RP-18, 5  $\mu$ m 5mm X 22cm). The sample was run at 1 ml/min flow rate from 30% aqueous acetonitrile plus 0.5% trifluoroacetic acid to 90% aqueous acetonitrile plus 0.5% trifluoroacetic acid over 30 min. The ions were observed using an Agilent TOF-MSD in positive ESI mode and the data was analyzed using Analyst QS software (ver 1.1, Applied Biosystems)

### 2.2.3 Anabaenopeptin I

#### 2.2.3.1 Isolation of Anabaenopeptin I (14)

The combined 50% and 75% aqueous methanol fractions were repeatedly injected on a Shimadzu LC-10AS equipped with a LUNA 10  $\mu$  C-18 column (250 X 10 mm, Phenomenex). The sample was run at 3 ml/min using the gradient outlined in Table 7. Pure anabaenopeptin I was isolated at a retention time of 21.4 min. The structure was elucidated through the use of NMR and MS techniques.

**Table 7:** HPLC gradient used for the isolation of Anabaenopeptin I. Mobile phase contained 0.05% trifluoroacetic acid.

Time (min)	AcCN (%): H <sub>2</sub> O (%)
0	30:70
10	30:70
25	50:50
30	50:50
32	100:0

**2.2.3.2 Anabaenopeptin I (14)** was isolated as a white amorphous powder:  $[\alpha]_D^{25} -21.43$  ° (c 0.35 mg ml<sup>-1</sup>, MeOH); UV (MeOH)  $\lambda_{max}$  (log  $\epsilon$ ) 280 (2937); IR (film)  $\nu_{max}$  3426, 3960, 2922, 1639 cm<sup>-1</sup>; HR-ESI TOF  $m/z$  916.4825 (calcd for C<sub>48</sub>H<sub>65</sub>N<sub>7</sub>O<sub>11</sub> 916.4821, 0.4 ppm error)

## 2.2.4 *N*-Me homotyrosine synthesis

### 2.2.4.1 *DL-N*-Acetyl-homotyrosine

*DL*-Homotyrosine (250 mg/1.28 mmol) was added to glacial acetic acid (13 mL) to create a suspension. Acetic anhydride (133  $\mu$ L/1.40 mmol) was added and then the reaction was kept at 110°C overnight. The acetic acid was removed and the solid product was suspended in a minimal amount of 1N HCl. The aqueous layer was extracted 3 times with EtOAc. The organic layers were combined and dried over MgSO<sub>4</sub>, filtered and then the solvent was removed yielding *N*-Acetyl-homotyrosine. Yield: 290 mg (95%) <sup>1</sup>H NMR (300 MHz, MeOH-*d*<sub>4</sub>)  $\delta$  1.99 (s, 3H), 2.05 (m, 2H), 2.59 (m, 2H), 4.32 (d,  $J$  = 3.6 Hz, 1H), 6.69 (d,  $J$  = 8.4 Hz, 2H), 7.00 (d,  $J$  = 8.4 Hz, 2H); <sup>13</sup>C NMR (75 MHz, MeOH-*d*<sub>4</sub>)  $\delta$  175.6, 173.4, 156.7, 132.9, 130.4, 116.2, 34.7, 32.2, 22.3

### 2.2.4.2 *L* and *D*-Homotyrosine

*DL-N*-Acetyl-homotyrosine (290mg/ 1.28mmol) was dissolved in 0.1M K<sub>2</sub>HPO<sub>4</sub> buffer (pH 7.0) (25 mL). The pH was then adjusted to 7 by addition of NaOH. Acylase I from *Aspergillus sp.* [Sigma-Aldrich] (50 mg/2.35 units) was dissolved in 0.1M K<sub>2</sub>HPO<sub>4</sub> buffer (pH 7.0) (0.5 mL). The acylase was added to the *DL-N*-Acetyl-homotyrosine solution and allowed to sit at RT for 36h. 75 mL AcCN was added and the resulting solid

filtered through Celite. The AcCN was removed and the remaining aqueous layer was titrated to pH 1 using HCl. The aqueous layer was extracted 4 times with EtOAc.

The organic layers were combined and dried over  $\text{MgSO}_4$ , filtered and then the solvent removed to yield a yellow oil. This oil was suspended in 3N HCl (10ml) and heated at  $65^\circ\text{C}$  overnight. The aqueous HCL was removed yielding *D*-Homotyrosine.

The extracted aqueous layer from above was titrated to pH 5.5 with solid  $\text{NaHCO}_3$  the concentrated to dryness to yield *L*-Homotyrosine.  $^1\text{H}$  NMR (300 MHz,  $\text{MeOH-}d_4$ )  $\delta$  2.01 (m, 2H), 2.52 (dd,  $J = 7.5, 7.8$  Hz, 2H), 3.88 (t,  $J = 5.4$  Hz, 1H), 6.67 (d  $J = 7.8$  Hz, 2H), 7.00 (d,  $J = 7.8$  Hz, 2H)

#### 2.2.4.3 Homotyrosine methyl ester

Resolved homotyrosine (125mg/0.64 mmol) was dissolved in dry MeOH (10mL) and TMS-Cl (813 $\mu\text{L}$ / 6.4 mmol) was added. The reaction was allowed to stir overnight at RT. The reaction mixture was concentrated to dryness and then dissolved in a minimal amount of dry MeOH (2-3mL). The dissolving and resuspending was repeated 3 times.

The crude product was used without further purification.  $^1\text{H}$  NMR (300 MHz,  $\text{MeOH-}d_4$ )  $\delta$  2.121 (m, 2H), 2.67 (m, 2H), 3.82 (s, 3H), 4.03 (t,  $J = 5.4$  Hz, 1H), 6.72 (d  $J = 7.8$  Hz, 2H), 7.06 (d,  $J = 7.8$  Hz, 2H);  $^{13}\text{C}$  NMR (75 MHz,  $\text{MeOH-}d_4$ )  $\delta$  169.9, 156.1, 130.8, 129.9, 115.7, 54.3, 53.2, 33.2, 30.6; IR (film) 3619, 1741, 1646, 1244; HR-ESI

$\text{C}_{11}\text{H}_{16}\text{NO}_3$  calcd. 210.1131, obs.210.1136, 2.4 ppm error

#### 2.2.4.4 *N,N*-Bn,Me homotyrosine methyl ester

Homotyrosine methyl ester was dissolved in MeOH (~0.1M solution) NaOAc (1.1eq) was added, followed by addition of benzaldehyde (1.05eq). The mixture was stirred for 1h at RT, at which time NaBH<sub>3</sub>(CN) (3eq) was added and the mixture stirred for 24h. Powdered paraformaldehyde (1.05eq) was added, the solution was allowed to stir for 4-5h at RT, then NaBH<sub>3</sub>(CN) (3eq) was added and the reaction was stirred for an additional 24h.

The solvent was removed and the oily product was resuspended in 1N HCl. The aqueous layer was extracted 2 times with Et<sub>2</sub>O. The aqueous layer was then adjusted to pH 9 with solid K<sub>2</sub>CO<sub>3</sub> at 0°C followed by extraction with EtOAc (4X). The organic layers were combined and dried over MgSO<sub>4</sub>, filtered and the solvent removed to yield a *N,N*-Bn<sub>2</sub>Me homotyrosine methyl ester. <sup>1</sup>H NMR (300 MHz, MeOH-*d*<sub>4</sub>) δ 1.96 (m, 2H), 2.24 (s, 3H), 2.56 (dd *J* = 7.6, 7.7 Hz, 2H), 3.54 (dd, *J* = 9.81 Hz, 1H), 3.68 (s, 3H), 6.66 (d, *J* = 8.4 Hz, 2H), 6.94 (d, *J* = 8.4 Hz, 2H), 7.27 (m, 5H). *L*: [α]<sup>25</sup><sub>D</sub> -21.3 (*c* 0.79, MeOH), *D*: [α]<sup>25</sup><sub>D</sub> 14.4 (*c* 0.99, MeOH); <sup>13</sup>C NMR (75 MHz, MeOH-*d*<sub>4</sub>) δ 174.4, 156.4, 140.3, 133.5, 130.4, 129.9, 129.2, 128.0, 116.1, 65.6, 59.6, 51.5, 38.0, 32.6, 32.3; IR (film) 3425, 2951, 1728, 1643, 1514, 1452, 1247; HR-ESI C<sub>19</sub>H<sub>24</sub>NO<sub>3</sub> calcd. 314.1757, obs. 314.1759, 0.6 ppm error

#### 2.2.4.5 *N*-Me homotyrosine methyl ester

*N,N*-Bn<sub>2</sub>Me homotyrosine methyl ester was suspended in EtOH acidified with HCl (0.06M solution). Pd(OH)<sub>2</sub> was added and the reaction was stored under hydrogen at atmospheric pressure for 18h. The reaction was filtered through Celite using EtOH acidified with HCl. The ethanol was removed and the solid was resuspended in dH<sub>2</sub>O.

Solid  $K_2CO_3$  was added to adjust the pH to 9 and then the aqueous layer was extracted 4 times with EtOAc. The organic layers were combined and dried over  $MgSO_4$ , filtered and the solvent removed to yield a *N*-Me homotyrosine methyl ester.  $^1H$  NMR (300 MHz, MeOH- $d_4$ )  $\delta$  1.90 (m, 2H), 2.54 (m, 2H), 3.17 (t,  $J = 6.5$  Hz, 1H), 3.70 (s, 3H), 6.68 (d,  $J = 8.6$  Hz, 2H), 6.99 (d,  $J = 8.6$  Hz, 2H). *L*:  $[\alpha]^{25}_D$  21.8 (*c* 0.38, MeOH), *D*:  $[\alpha]^{25}_D$  -18.4 (*c* 0.46, MeOH);  $^{13}C$  NMR (75 MHz, MeOH- $d_4$ )  $\delta$  175.9, 156.7, 133.1, 130.4, 116.2, 63.4, 52.3, 35.8, 34.3, 31.9; IR (film) 3403, 1643, 1454, 1248; HR-ESI  $C_{12}H_{18}NO_3$  calcd. 224.1287, obs. 224.1264, 10.2 ppm error

#### 2.2.4.6 *N*-Me homotyrosine

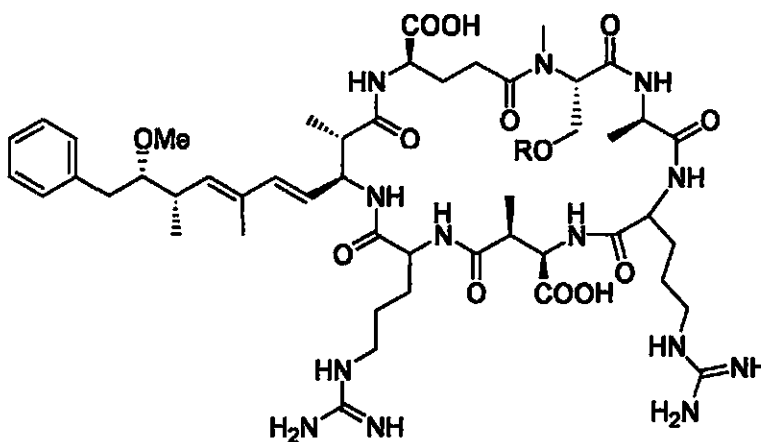
*N*-Me homotyrosine methyl ester was dissolved in dry EtOAc and LiI (5eq) was added. The reaction was refluxed for 20h in darkness. The reaction mixture was concentrated to dryness, resuspended in EtOAc and the solid was collected and washed with EtOAc. The solid was dissolved in  $H_2O$ , filtered and then the water was evaporated to yield *N*-Me homotyrosine.  $^1H$  NMR (300 MHz, MeOH- $d_4$ )  $\delta$  1.90 (m, 2H), 2.43 (m, 2H), 2.45 (s, 3H), 3.31 (t,  $J = 6.1$  Hz), 6.68 (d,  $J = 8.5$  Hz, 2H), 7.01 (d,  $J = 8.5$ , 2H). *L*:  $[\alpha]^{25}_D$  13.4 (*c* 0.36, 1N HCl), *D*:  $[\alpha]^{25}_D$  -12.43 (*c* 0.17, 1N HCl);  $^{13}C$  NMR (75 MHz, MeOH- $d_4$ )  $\delta$  170.9, 157.0, 131.8, 130.4, 116.4, 61.5, 32.6, 32.5, 30.9; IR (film) 3566, 1646; HR-ESI  $C_{11}H_{16}NO_3$  calcd. 210.1130, obs. 210.1127, 1.4 ppm error

## 2.3 RESULTS AND DISCUSSION

### 2.3.1 Creation of CYA 126/8 $\Delta mcyT$

McyT is a putative type II thioesterase. Enzymes of this class have been shown to be responsible for editing and proofreading during natural product biosynthesis. Walsh and co-workers have demonstrated that these Te domains are responsible for cleaning off the 4'-PP arm of thiolation domains by either removing wrongly activated amino acids or acetyl groups (incorporated by attack on acetyl-coenzyme A instead of co-enzyme A during phosphopantetheinylation).<sup>44</sup> The lack of *mcyI* in the genome of CYA 126/8 led us to hypothesize that *mcyT* might be involved in the formation of Mdma with O-Ac-Ser as the activated intermediate. Given that previous experiments with *mcyJ* knockouts produced (DMAdda, *D*-Asp<sup>3</sup>)-MC-RR, we hoped that insertional inactivation of *mcyT* would produce (O-AcSer<sup>1</sup>-*D*-Asp<sup>3</sup>)-MC-RR (12) or (*L*-Ser<sup>1</sup>-*D*-Asp<sup>3</sup>)-MC-RR (13).

The construction of the recombination plasmid required special restriction conditions to only cleave at one of the *PfoI* sites located in the *mcyT* sequence. Previously attempted homologous recombination experiments with *P. agardhii* revealed the need to have approximately 1.5kb regions flanking the chloramphenicol resistance cassette (G.



12 R = Ac  
13 R = H



Christiansen, personal communication). To ensure the appropriate size of the flanking regions we needed to insert the chloramphenicol resistance cassette at the 5'-end of the *mcyT*. Upon *in silico* analysis of the putative amplicon, we observed that the only site suitable for disruption of *mcyT* corresponded to the recognition sequence of *PfoI* (T↓CCNGGA). However another *PfoI* site was located 750bp downstream of the desired insertion site. This created a problem because we could not cleave at both sites as this would reduce the flanking region and result in very low recombination efficiencies if any recombinants were created. An extensive literature search provided a solution to the problem, ethidium bromide. Ethidium bromide is a known DNA intercalation agent and studies by Parker *et al* demonstrated that ethidium bromide could be used to stop restriction after linearization at one site. Ethidium bromide intercalates at a much higher rate into linearized (a double strand cut) and nicked (a single strand cut) plasmids than into the closed circular form that plasmids are usually found in.<sup>45</sup> By optimizing the amount of ethidium bromide present in the reaction mixture (0.02 mg/ml), the amount of singly restricted plasmid could be maximized. After isolation of the linearized plasmid, the 5'-overhanging ends were blunted by use of a Klenow fragment and then ligated to the chloramphenicol (Cm) resistance cassette that had been restricted from the pACYC184 plasmid using *BsaAI* (YAC↓GTR). In order to ensure that the plasmids that were isolated had the Cm cassette in the correct *PfoI* site, a colony PCR was performed. Out of the 20 individual colonies that were tested, only 1 clone contained the chloramphenicol cassette at the desired location.

The DNA was then introduced into the cyanobacterium by electroporation as linear single stranded DNA (See experimental section). After adding the chloramphenicol

to a final concentration of 1  $\mu\text{g/ml}$  the culture became mostly bleached within 2 days and completely bleached within 1 week. After 5 weeks of growing under low light conditions, a faint green tint could be seen. The mutant was then grown to sufficient mass for DNA extraction and peptide analysis. In order to ensure that the desired mutant had been grown and not a random chloramphenicol resistant mutant, we checked the isolated mutant DNA by PCR. The mutant gave an amplicon the same size as the construct indicating that the desired homologous recombination had occurred (Figure 8).

### 2.3.2 Peptide analysis

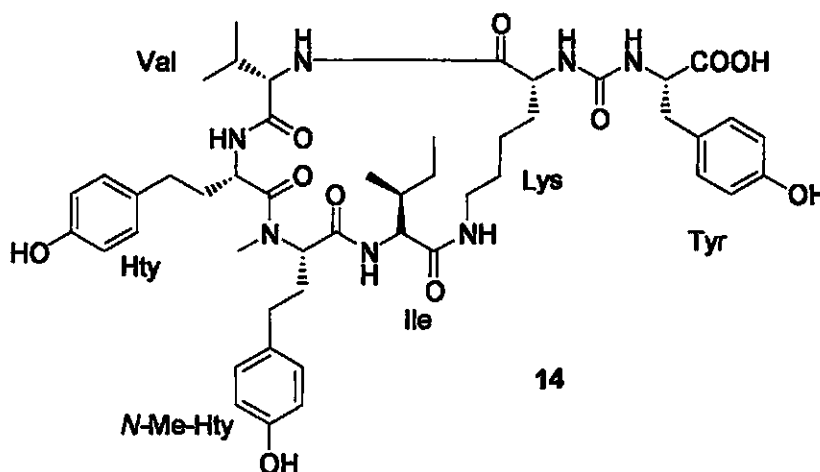
Upon comparison of the peptide content of *CYA 126/8  $\Delta mcyT$*  to *CYA 126/8* wild type, several differences could be observed. The wild type (2.1g wet weight) yielded 40 mg of material from a C18 flash column eluted with 50% aqueous methanol and 40 mg in the fraction eluted with 75% methanol. The latter fraction was shown to contain the microcystins by NMR analysis. *CYA 126/8  $\Delta mcyT$*  (1.2g wet weight), in contrast yielded 10 mg in the 50% aqueous methanol fraction but only 2 mg in the 75% aqueous methanol fraction. The characteristic peaks of the Adda side chain were absent from the crude NMR spectra of the 75% aqueous MeOH fraction (data not shown). This was corroborated by HPLC and LC-MS analysis as the extremely large peak due to (*D*-Asp<sup>3</sup>)-



**Figure 8:** 2.0% agarose gel verifying the correct mutant had been created. PCR performed with primers *mcyT5'*seq and *mcyT3'*seq.

MC-RR had been severely reduced (See appendix), indicating that *mcyT* is crucial for microcystin biosynthesis.

### 2.3.3 Anabaenopeptin I



Some cyanobacterial strains have been shown to upregulate or produce previous unknown peptides when a gene cluster is knocked out. Therefore the combined 50% and 75% aqueous methanol fractions were then analyzed to see if the same held true for CYA 126/8  $\Delta mcyT$ . In the absence of the huge (*D*-Asp<sup>3</sup>)-MC-RR peak characteristic of the wildtype, one uncharacterized peak stood out and was isolated pure after one HPLC column with an increased yield when compared to the wild type (0.59 mg per g dry mass vs. 0.33 mg per g dry mass). It should be noted that the increase in mass of anabaenopeptin I cannot be solely attributed the creation of the mutant and reduction of (*D*-Asp<sup>3</sup>)-MC-RR. It is uncertain whether this increase in mass is due to collection bias or represents a real increase in production by *P. agardhii*. The first thing noticed in the <sup>1</sup>H-NMR data (methanol-*d*<sub>4</sub>) was the presence of 3 *para* substituted benzene rings indicating

homotyrosine or tyrosine residues. Using  $^1\text{H-NMR}$  (methanol- $d_3$ ), HMQC and 1D TOCSY experiments, the compound was found to contain two homotyrosine, tyrosine, valine and an isoleucine residue with one of the residues containing an N-methyl group. It also contained either a lysine or an ornithine residue but the identity could not be determined using the  $^1\text{H-NMR}$  data alone. The presence of a lysine was confirmed by FABMS ( $m/z$   $\text{M}+\text{H}^+$  916.4) and upon review of the  $^1\text{H NMR}$  data it was seen that the resonances of two of the methylenes overlapped. HRESI-TOF gave a mass of 916.4727 ( $\text{M}+\text{H}^+$ , 0.17 ppm error), which corresponds to a molecular formula of  $\text{C}_{48}\text{H}_{65}\text{N}_7\text{O}_{11}$ . When totaling the molecular formula for the known residues all but  $\text{CHO}_2$  was represented. The  $^{13}\text{C NMR}$  contained a signal at 159.83, which hinted at the presence of an ureido bond leaving an OH unaccounted for. These pieces of information led us to believe that we had isolated a member of the anabaenopeptin family. Using 1D TOCSY and HMBC experiments, the 2D structure was established (Table 8). As in the case of all the known anabaenopeptins the lysine 2-NH and the alpha carbon (C2) show a correlation with the ureido carbonyl (C39). Similarly the NH of the tyrosine residue also showed a large correlation with C39. With the Lys and Tyr residues bound to the ureido carbonyl and due to the lack of HMBC correlation to the carbonyl of the Tyr it was determined that C48 was not involved in an amide bond and therefore must be a free carboxylic acid. This last piece of evidence accounted for the OH that had still to be accounted for in the molecular formula. Interestingly the  $\delta_{\text{C}}$  of the carbonyl carbon of the tyrosine subunit could only be determined through an HMBC correlation from

Table 8: NMR data for Anabaenopeptin I (14)

Unit	C/H number	$\delta_H$ (J in Hz) <sup>a</sup>	$\delta_C$ <sup>b</sup>	HMBC <sup>c,d</sup>
Lys	1		175.93	
	2	4.06, m	56.54	1,3,4,39
	2-NH	6.64, d (4.73)		2,3, 39
	3	1.91, m 1.76, m	32.05	
	4	1.53, m 1.31, m	21.93	
	5	1.56, m 1.31, m	29.20	
Ile	6	3.54, m 2.98, m	40.38	4
	6-NH	7.31, dd (7.03, 3.02)		6
	7		174.24	
	8	4.3, dd (9.61, 6.16)	60.11	7,9,10,12,13
	8-NH	8.42, d (9.61)		8,13,
	9	2.03, m	37.68	
N-Me-Hty	10	1.38, m 1.10, m	26.09	
	11	0.79, t (7.38)	11.67	10
	12	0.85, d (6.89)	16.57	9,10
	13		171.69	
	14	4.52, t (6.82)	61.42	13,23,24
	15	2.12, m 1.80, m	34.85	
	16	2.26, dt (-12.44, 5.02) 2.32, dt (-12.48, 4.63)	33.03	17
	17		133.181	
	18/22	6.97, d (8.47)	130.21	17,16,19/21,20
	19/21	6.68, m	116.33	
Hty	20			
	23	2.70, s	29.77	24
	24		175.21	
	25	4.653, m	51.25 <sup>*</sup>	24
	25-NH	8.96, d (4.01)		34
	26	2.12, m 1.80, m	34.85	
	27	2.78, m 2.64, m	31.94	17
	28		132.79	
	29/33	7.064, d (8.31)	131.06	28,30/32,31
	30/32	6.68, m	116.48	
Val	31			
	34		175.01	
	35	3.915, dd (8.03, 5.59)	61.17	34
	35-NH	7.28, d (4.73)		1
	36	1.97, m	31.45	37,38
	37	0.98, d (6.60)	20.07	35,36
Ureido	38	0.98, d (6.60)	19.76	35,36
	39		159.83	
Tyr	40	4.45, m	56.6 <sup>*</sup>	48
	40-NH	6.06, d (8.03)		39
	41	3.03, m 2.91, m	38.37	42
	42		129.09	
	43/47	7.024, d (8.32)	131.48	42,44/46,45
	44/46	6.68, m	116.50	
	45			
	48		177.0 <sup>**</sup>	

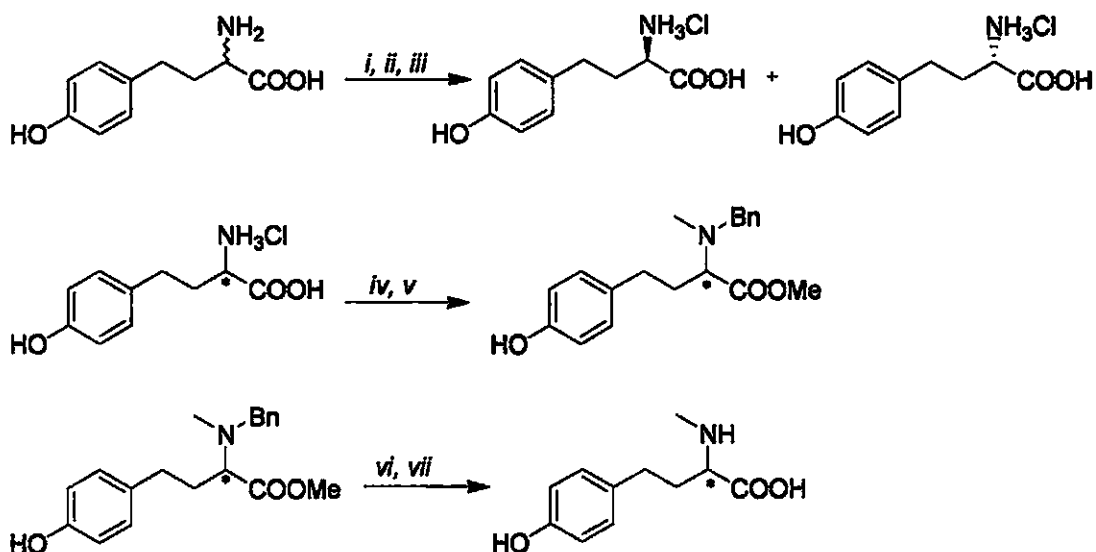
<sup>a</sup> Recorded at 500 MHz <sup>b</sup> Recorded at 125 MHz <sup>c</sup> Protons showing long range correlation with indicated carbon <sup>d</sup> Correlations were observed for <sup>2</sup>J<sub>CH</sub> = 7 Hz \*  $\delta_C$  of this carbon could only be determined through HMQC \*\*  $\delta_C$  of this carbon could only be determined through HMBC

H40. The rest of the residues were placed in their respective positions using the HMBC correlations of the NH and alpha carbons to the carbonyl signals.

### 2.3.4 *N*-Me homotyrosine synthesis

After determining that *N*-Me-Hty was present in the molecule, a synthesis had to be undertaken to create the enantiomerically pure standards for configurational analysis (Figure 9). The first step was to resolve the *DL*-homotyrosine mixture. This was accomplished by exposing *N*-Ac-Hty to Acylase I isolated from *A. melleus* to remove the acetamide unit from the *L* enantiomer. The resulting *D* enhanced mixture was deacetylated by exposure to 1N HCl. The synthesis of each enantiomer was then done in parallel using standard methodology. The benzylation and methylation were performed in one pot using an adaptation of a recently developed protocol.<sup>46</sup> Instead of exposing the Hty methyl ester•HCl salt to Amberlite IRA-410 resin to make the free base form, 1.1 eq of NaOAc was added to the Hty methyl ester•HCl in methanol prior to benzaldehyde addition to act as a proton scavenger. Creation of the free base was essential for the benzylation reaction to proceed as planned. Failure to do this resulted in incomplete benzylation, which in turn resulted in multiple products being isolated from the reaction. Powdered formaldehyde was used as suggested, however owing to the suspended NaOAc from the neutralization step, the dissolution of the formaldehyde could not be observed visually as described.<sup>46</sup> Simultaneously in the lab the synthesis of *D*-*N*-Me-Phe was being attempted. It was observed that exposure of *D*-*N*-Me-Phe-OMe to LiOH in MeOH resulted in racemization. It was assumed that the same problem would be encountered with *N*-Me-Hty-OMe. A literature search revealed that a few research groups had

encountered similar problems with protected amino acids.<sup>47,48</sup> Therefore the removal of the methyl ester was performed using LiI in refluxing EtOAc.<sup>48</sup> The standards were then used for stereochemical analysis without further purification.



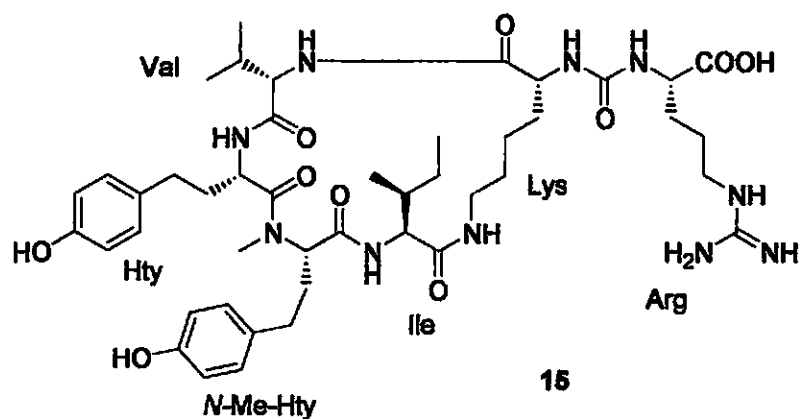
**Figure 9:** Synthesis of enantiomerically enriched *N*-Me homotyrosine

(*i*)  $\text{Ac}_2\text{O}$ ,  $\text{AcOH}$ ,  $110^\circ\text{C}$ , 16h; (*ii*) Acylase I, 0.1M  $\text{K}_2\text{HPO}_4$  buffer (pH 7.0); (*iii*) 1N  $\text{HCl}$ ,  $100^\circ\text{C}$ ; (*iv*)  $\text{TMS-Cl}$ ,  $\text{MeOH}$ ; (*v*) (a)  $\text{NaOAc}$  (1.1 eq),  $\text{C}_6\text{H}_5\text{CHO}$  (1.05 eq),  $\text{MeOH}$ , 1h then  $\text{NaBH}_3(\text{CN})$  (3 eq), 24h (b)  $\text{CH}_2\text{O}$  (1.05 eq) 4.5h then  $\text{NaBH}_3(\text{CN})$  (3 eq), 24h; (*vi*)  $\text{H}_2$ ,  $\text{Pd}(\text{OH})_2$ ,  $\text{EtOH} + \text{HCl}$ , O/N; (*vii*)  $\text{LiI}$ ,  $\text{EtOAc}$  reflux, O/N

### 2.3.5 Stereochemistry of Anabaenopeptin I (14)

The stereochemistry of the amino acid residues was determined by advanced Marfey analysis<sup>49</sup> using FDAA as the derivatization reagent and observed with an LC-MS equipped with an ESI source either in positive or negative mode. This setup greatly reduced the work required in identification because it removed the need for co-injection of standards. After hydrolysis of **14** with 6N  $\text{HCl}$ , D-Lys, L-Val were determined to be present as well as either L-Ile or *L-allo*-Ile. The Ile species could not be resolved using a non-chiral column. Use of a chiral column later revealed that L-Ile was present in the molecule (Cyril Portman, personal communication). The analysis of the stereochemistry

of the Tyr, Hty and N-Me-Hty residues was extremely problematic, as analysis of the hydrolysate did not show any traces of the phenolic residues. Itou *et al* resorted to use of anhydrous hydrazine in order to determine the stereochemistry of the Tyr side chain in anabaenopeptin H. However making the required reagent is time consuming and the by-products toxic, which makes the approach unappealing. Therefore it was decided to scan less harsh hydrolysis conditions to see if the Tyr and Hty residues could be observed. Degradation with 2N HCl at 55°C for 2h surprisingly was harsh enough to liberate the Tyr residue from the ureido bond. Detection of the Marfey derivatized residue using negative ESI-TOF allowed the Tyr to be assigned the *L* configuration. However, the Hty and N-Me-Hty residues were still absent from the spectrum. The configurations of the Hty residues is proposed on the basis of the similarity of the NMR spectra of anabaenopeptin I with those of another anabaenopeptin (15) isolated from wild type CYA 126/8, which differed from 14 simply by replacement of the Tyr residue with an Arg residue. Anabaenopeptin J (0.1mg) was degraded in 6N HCl for 12 at 110°C, followed by derivatization with 3eq of Marfey's reagent. The products were then observed using LC-MS in negative ESI mode. Both the Hty and the N-Me-Hty residues coeluted with the *L* enantiomer standards (Cyril Portman, personal communication)



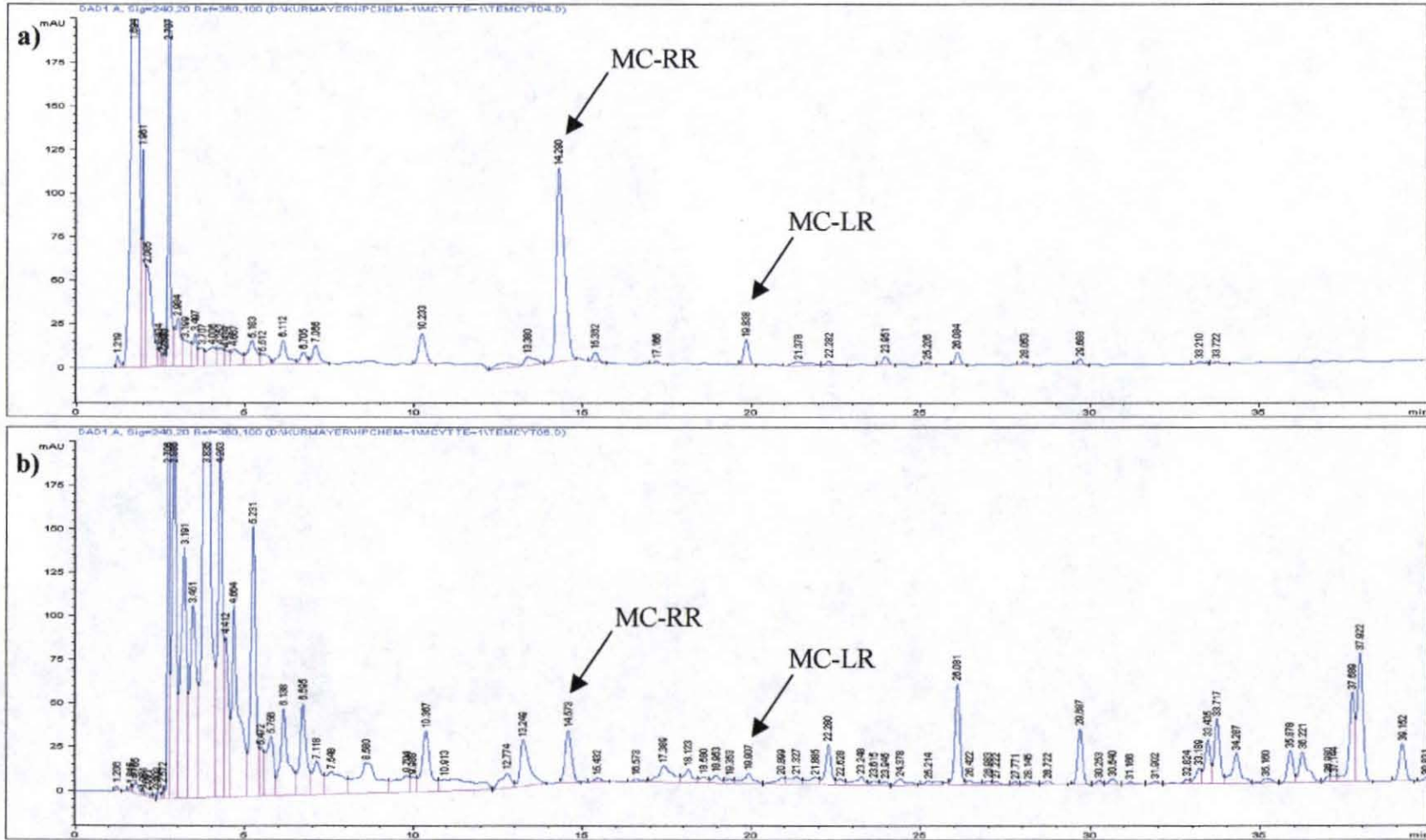


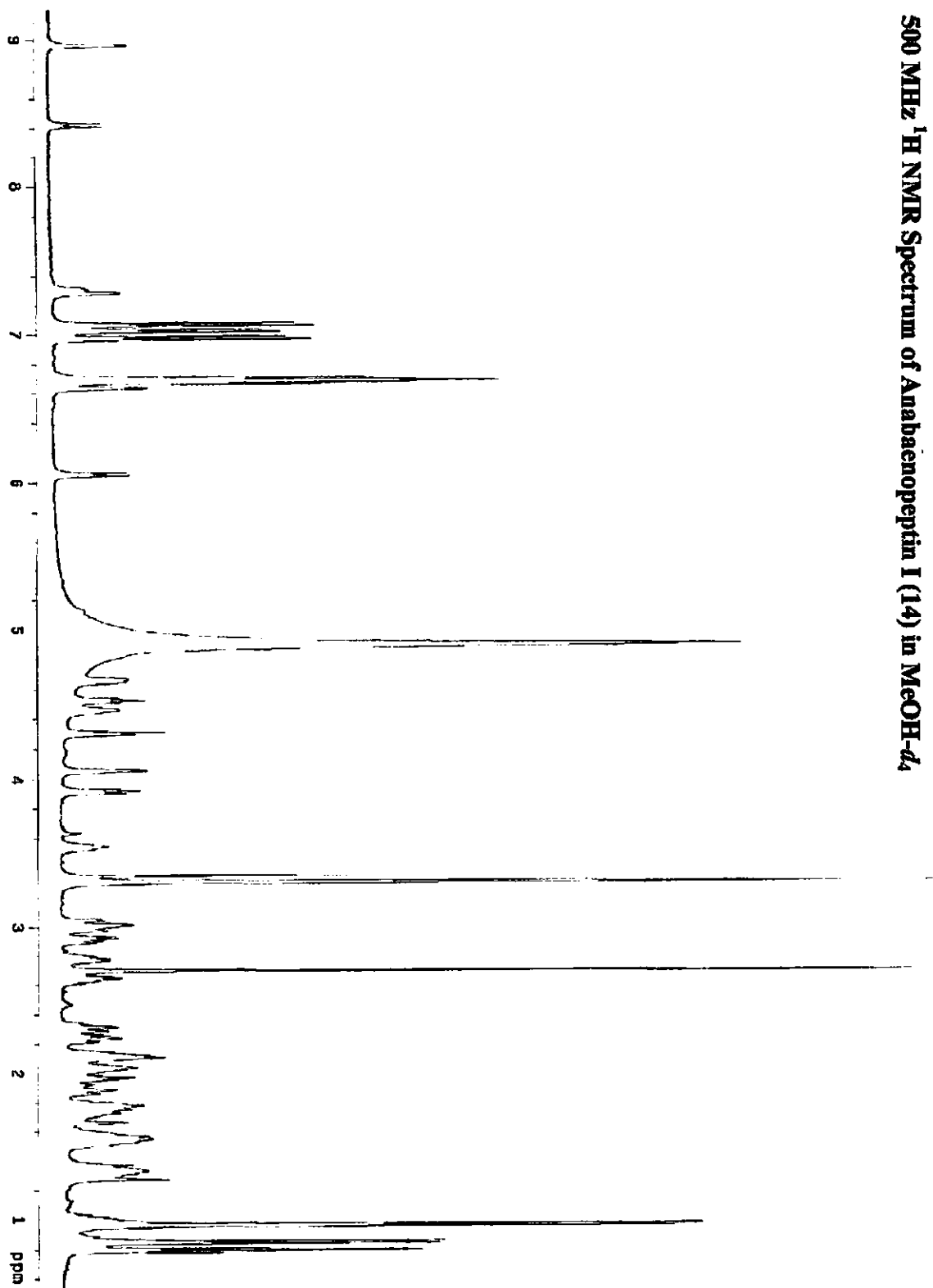
## 2.4 CONCLUSION

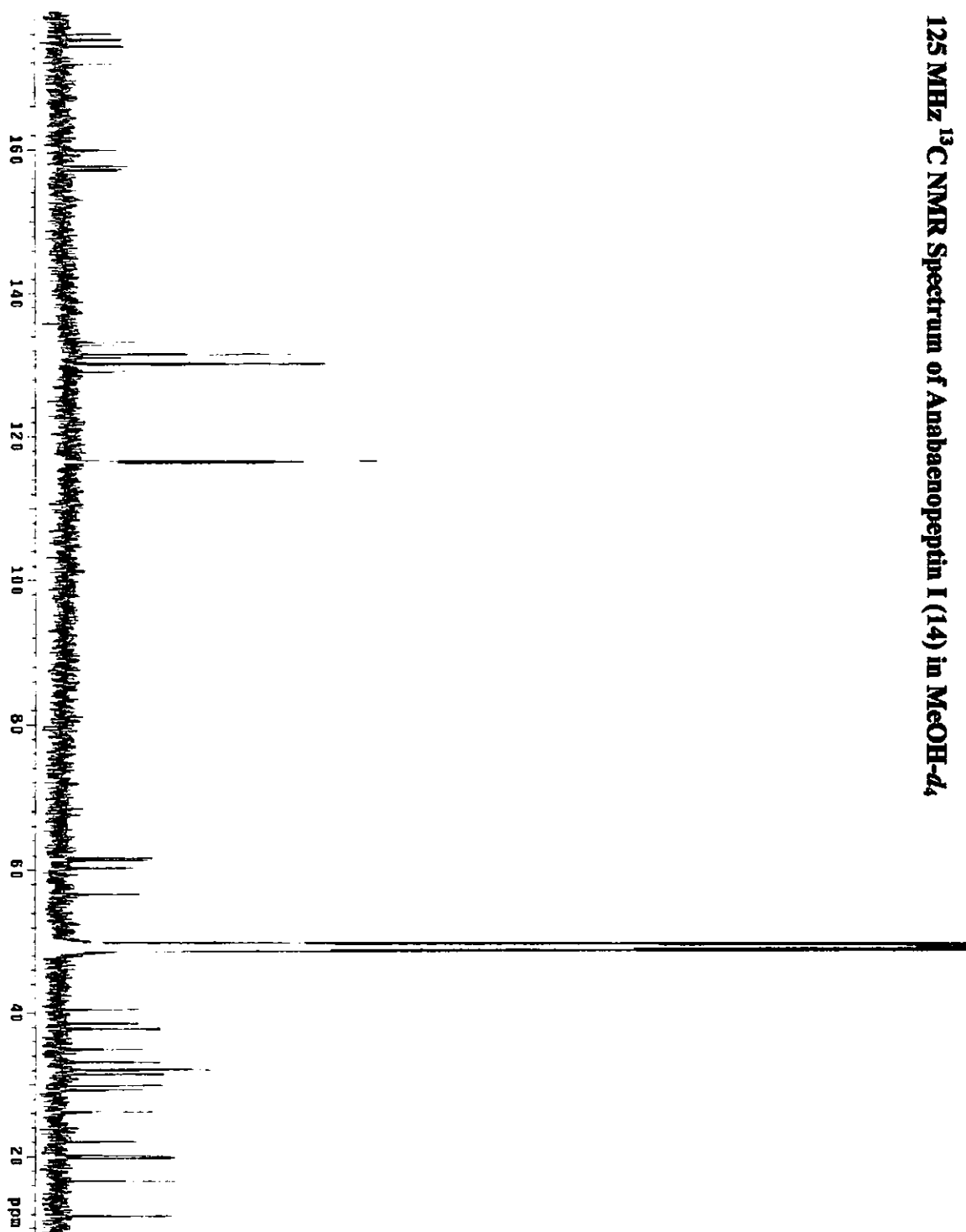
Insertional mutagenesis was utilized to create a knockout mutant of *mcyT* from *P. agardhii* CYA 126/8. This was done to investigate the role of *mcyT* in microcystin biosynthesis. The generated mutant did not produce a structural variant of MC-RR as expected and microcystin biosynthesis reduced to a level far below that of the wild type (<5%) was observed instead. This indicates that *mcyT* is essential for microcystin biosynthesis, even if its precise function remains a mystery. If McyT did function as a *general* type II thioesterase, one would expect that upon disruption all non-ribosomal peptide production would be reduced. However this is not observed. One explanation for why only microcystin biosynthesis is affected is that each NRPS cluster may have its own dedicated type II thioesterase. Therefore, a knockout of *mcyT* would not necessarily lead to disruption of peptide biosynthesis by other clusters. In this connection it is interesting to note that recent studies have elucidated the fact that a population of *P. agardhii* contains individuals carrying defective *mcy* genes, which result from insertions or deletions of gene fragments. All of the resulting mutants are deficient in microcystin production. It is noteworthy, however, that to this date no strain has yet been found containing a defective variant of *mcyT* as ascertained by PCR (G. Christiansen, personal communication). This may suggest that several closely related type II thioesterases are present which could result in false positives in the PCR screen. Unfortunately, at this time too little genomic information exists about *P. agardhii*, or other cyanobacteria with a well-studied spectrum of secondary metabolites, to accurately assess the role and specificity McyT and type II thioesterases in general.

Upon peptide analysis of *P. agardhii* CYA 126/8  $\Delta mcyT$  an uncharacterized peak was isolated and the structure determined. The 2D structure of anabaenopeptin I (14) was elucidated using NMR and HR-FABMS techniques. The stereochemistry of the residues was determined by advanced Marfey analysis. The derivatized residues were observed using LC-MS, which greatly reduced the workload associated with identifying the enantiomer. The phenolic residues (Hty, Tyr) proved problematic to identify and were therefore identified through  $^1\text{H}$  NMR comparison to a closely related compound, anabaenopeptin J (15). The similarities between the proton resonances of the residues in the macrolactam ring residues and the fact that the Lys, Ile and Val residues were of identical configuration in each compound led us to believe that the stereochemistry of the Hty and Tyr residues would also be identical.

HPLC trace of (a) *P. agardhii* CYA 126/8 wild type and (b) *P. agardhii* CYA 126/8  $\Delta$ mcyT. HPLC was performed as described in section 2.2.2.2 and observed at 240 nm.



**500 MHz  $^1\text{H}$  NMR Spectrum of Anabaenopeptin I (14) in  $\text{MeOH-}d_4$** 

**125 MHz  $^{13}\text{C}$  NMR Spectrum of Anabaenopeptin I (14) in MeOH- $d_4$** 

#### 4. References

- (1) Van Lanen, S. G.; Shen, B. *Curr. Opin. Microbiol.* **2006**, *in press*
- (2) Lesney, M. S. *Today's Chemist at Work* **2004**, *July*, 26-32
- (3) <sup>1</sup> Borel, J. F.; Feurer, C.; Gubler, H.U.; Stähelin, H. *Agents and Actions* **1976**, *6*, 468-475
- (4) Jensen, P. R.; Mincer, T. J.; Williams, P. G.; Fenical, W. *Antoine van Leeuwenhoek* **2005**, *87*, 43-48 and references therein
- (5) Kobayashi M.; Kawazoe, K.; Okamoto, T.; Sasaki, T.; Kitagawa, I. *Chem. Pharm. Bull. (Tokyo)* **1994**, *42(1)*, 19-26
- (6) Bewley, C. A.; Holland, N. D.; Faulkner, D. J. *Experientia* **1996**, *52*, 71 –722
- (7) Andrianasolo, E. H.; Gross, H.; Goeger, D.; Musafija-Girt, M.; McPhail, K.; Leal, R. M.; Mooberry, S. L.; Gerwick, W. H. *Org. Lett.* **2005**, *7(7)*, 1375-1378
- (8) Burgess, J. G.; Jordan, E. M.; Bregu, M.; Mearns-Spragg, A.; Boyd, K. G. *J. Biotechnol.* **1999**, *70*, 27-32
- (9) Chang, Z.; Flatt, P.; Gerwick, W. H.; Nguyen, V. A.; Willis, C. L.; Sherman, D. H. *Gene* **2002**, *296(1-2)*, 235-247
- (10) Harrigan, G. G.; Yoshida, W. Y.; Moore, R. E.; Nagle, D. G.; Park, P. U.; Biggs, J.; Paul, V. J.; Mooberry, S. L.; Corbett, T. H.; Valeriote, F. A. *J. Nat. Prod.* **1998**, *61(10)*, 1221-1225
- (11) Ikeda, H.; Ishikawa, J.; Hanamoto, A.; Shinose, M.; Kikuchi, H.; Shiba, T.; Sakaki, Y.; Hattori, M.; Omura, S. *Nat. Biotechnol.* **2003**, *21*, 526-531
- (12) <sup>1</sup> (a) Golakoti, T.; Ogino, J.; Heltzel, C. E.; Husebo, T. L.; Jensen, C. M.; Larsen, L. K.; Patterson, G. M. L.; Moore, R. E.; Mooberry, S. L.; Corbett, T. H.; Valeriote, F. A. *J.*

*Am. Chem. Soc.* **1995**, *117*, 12030-12049 (b) Subbaraju, G. V.; Golakoti T.; Patterson, G. M.; Moore, R. E. *J. Nat. Prod.* **1997**, *60*(3), 302-305 (c) Chaganty, S.; Golakoti, T.; Heltzel, C.; Moore, R. E.; Yoshida, W. Y. *J. Nat. Prod.* **2004**, *67*, 1403-1406

(13) Adapted from <http://203.90.127.50/~pkfdb/polyketide.html>

(14) For an in depth review on NRPS systems see Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*(7), 2651-2674 and references therein

For an in depth review on PKS systems see Hopwood, D. A. *Chem. Rev.* **1997**, *97*(7), 2465-2498 and references therein

(15) Becker, J. E.; Moore, R. E.; Moore, B. S. *Gene* **2004**, *325*, 35-42

(16) Dittmann, E.; Neilan, B. A.; Erhard, M.; von Dohren, H.; Borner, T. *Mol. Microbiol.* **1997**, *26*(4), 779-787

(17) Christiansen, G.; Fastner, J.; Erhard, M.; Borner, T.; Dittmann, E. *J. Bacteriol.* **2003**, *185*(2), 564-572

(18) Rouhiainen, L.; Vakkilainen, T.; Siemer, B. L.; Buikema, W.; Haselkorn, R.; Sivonen, K. *Appl. Environ. Microbiol.* **2004**, *70*(2), 686-692

(19) Azevedo, S. M.; Carmichael, W. W.; Jochimsen, E. M.; Rinehart, K. L.; Lau, S.; Shaw, G. R.; Eaglesham, G. K. *Toxicology* **2002**, *27*(181-182), 441-446

(20) Nishiwaki-Matsushima, R.; Ohta, T.; Nishiwaki, S.; Suganuma, M.; Kohyama, K.; Ishikawa, T.; Carmichael, W. W.; Fujiki, H. *J. Cancer Res. Clin. Oncol.* **1992**, *118*(6), 420-424

(21) Chen G.; Yu S.; Wei G. *Zhonghua Yu Fang Yi Xue Za Zhi* **1996**, *30*(1), 6-9

(22) Taken from

[http://www.who.int/water\\_sanitation\\_health/dwq/chemicals/microcystin/en/](http://www.who.int/water_sanitation_health/dwq/chemicals/microcystin/en/)

- (23) Goldberg, J.; Huang, H. B.; Kwon, Y. G.; Greengard, P.; Nairn, A. C.; Kuriyan, J. *Nature* **1995**, *376*(6543), 745-753
- (24) Prinsep, M. R.; Caplan, F. R.; Moore, R. E.; Patterson, G. M. L.; Honkanen, R. E.; Boynton, A. L. *Phytochemistry* **1992**, *31*(4), 1247-1248
- (25) Oksanen, I.; Jokela, J.; Fewer, D. P.; Wahlsten, M.; Rikkinen, J.; Sivonen, K. *Appl. Environ. Microbiol.* **2004**, *70*(10), 5756-5763
- (26) <sup>1</sup> Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning, a Laboratory Manual*; 2<sup>nd</sup> ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989
- (27) 7.5g Bacto<sup>®</sup> yeast extract and 8.0g Bacto<sup>®</sup> nutrient broth were dissolved in 900 ml dH<sub>2</sub>O. The Ph was then adjusted to 7.5 with NaOH and the volume adjusted to 1 L. The media was autoclaved for 20 min at 15 lbs./sq. in.
- (28) Neilan, B. A.; Jacobs, D.; Goodman, A. *Appl. Envir. Micro.* **1995**, *61*, 3875-3883
- (29) William, O. O. *MS Thesis ES 04.03*, **2004**
- (30) Obtained from the Culture Collection of Algae and Protozoans.  
<http://www.ccap.ac.uk/media/recipes.htm#>
- (31) Moore, R. E.; Cheuk, C.; Yang, X. -Q. G.; Patterson, G. M. L.; Bonjouklian, R.; Smitka, T. A.; Mynderse, J. S.; Foster, R. S.; Jones, H. D.; Swartzendruber, J. K.; Deeter, J. B. *J. Org. Chem.* **1987**, *52*(6), 1036-1043
- (32) Corpet, F. *Nucl. Acids Res.* **1998**, *16*(22), 10881-10890
- (33) Vector NTI was graciously provided by HS-BRIN through grant RR-16467
- (34)  
<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfind>
- b



- (35) Moffitt, M. C.; Neilan, B. A. *Appl. Environ. Microbiol.* **2004**, *70*(11), 6353-6362
- (36) Stachelhaus, T.; Mootz, H. D.; Marahiel, M. A. *Chem. Biol.* **1999**, *6*, 493-505
- (37) Marahiel, M. A.; Stachelhaus, T.; Mootz, H. D. *Chem. Rev.* **1997**, *97*, 2651-2673
- (38) Tanabe Y.; Kaya, K.; Watanabe, M. M. *J. Molec. Evol.* **2004**, *58*, 633-641
- (39) Mikalsen, B.; Borson, G.; Skulberg, O. M.; Fastner, J.; Davies, W.; Gabrielsen, T. M.; Rudi, K.; Jakobson, K. S. *J. Bacteriol.* **2003**, *185*, 2774-2785
- (40) Challis, G. L.; Ravel, J.; Townsend. C. A. *Chemistry & Biology* **2000**, *7*, 211-224
- (41) Liberles, J. S.; Thorolfsson M.; Martinez, A. *Amino Acids* **2005**, *28*(1), 1 – 12 and references therein
- (42) Koponen, O.; Tolonen, M.; Qiao, M.; Wahlström, G.; Helin, J.; Saris, P. E. *J. Microbiol.* **2002**, *148*, 3561 - 3568
- (43) Chatterjee, C.; Miller, L. M.; Leung, Y. L.; Yi, M.; Kelleher, N. L.; van der Donk, W. A. *J. Am. Chem. Soc.* **2005**, *127* (44), 15332 - 15333
- (44) Yeh, E.; Kohl, R. M.; Bruner, S. D.; Walsh, C. T. *ChemBioChem* **2004**, *5*, 1290-1293
- (45) Parker, R. C.; Watson, R. M.; Vinograd, J. *PNAS* **1977**, *74*(3), 851-855
- (46) White, K. N.; Konopelski, J. P. *Org. Lett.* **2005**, *7*(19), 4111-4112
- (47) (a) Ohafune, Y.; Kurokawa, N.; Higuchi, N.; Saito, M.; Hashimoto, M.; Tanaka, T. *Chem. Lett.* **1984**, 441-444; (b) McDermott, J. R.; Benoiton, N. L. *Can. J. Chem.* **1973**, *51*, 2555-2561; (c) Cheung, S. T.; Benoiton, N. L. *Can. J. Chem.* **1977**, *55*, 916-921
- (48) Biron, E.; Kessler, H. *J. Org. Chem.* **2005**, *70*, 5183-5189
- (49) Fujii, K.; Ikai, Y.; Mayumi, T.; Oka, H.; Suzuki, M.; Harada, K-I. *Anal. Chem.* **1997**, *69*, 3346-3352