Cyanobactins - ribosomally synthesized and post-translationally modified peptides produced by cyanobacteria

Niina Leikoski

Division of Microbiology and Biotechnology
Department of Food and Environmental Sciences
Faculty of Agriculture and Forestry and
Viikki Doctoral Programme in Molecular Biosciences
University of Helsinki

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Agriculture and Forestry of the University of Helsinki, for public examination in auditorium 1041, at Viikki Biocenter 2, Viikinkaari 5, on August 16th, at 12 o'clock noon.

HELSINKI 2013

Supervisors: Principal supervisor
Professor Kaarina Sivonen
Department of Food and Environmental Sciences
University of Helsinki, Finland

Docent David P. Fewer
Department of Food and Environmental Sciences
University of Helsinki, Finland

Docent Jouni Jokela
Department of Food and Environmental Sciences
University of Helsinki, Finland

Reviewers: Professor Marcel Jaspars
Marine Biodiscovery Centre
Department of Chemistry
University of Aberdeen,
Scotland, UK

Docent Jarmo Niemi
Department of Biochemistry and Food Chemistry
University of Turku, Finland

Opponent: Dr Paul D. Cotter
Teagasc Food Research Centre
Alimentary Pharmabiotic Centre
Moorepark, Fermoy
Cork, Ireland

Front cover: *Anabaena* sp. 90 -cyanobacterium (photo by Kaisa Rantasärkkä) and the structure of anacyclamide A10.

ISSN 1799-7372
ISBN 978-952-10-9035-6 (paperpack)
ISBN 978-952-10-9036-3 (PDF)

Unigrafia
Helsinki 2013

Table of contents

List of original publications 4

Abbreviations 5

Abstract 6

Tiivistelmä 8

1. Introduction 10

1.1. Cyanobacteria 10

1.2. Ribosomally synthesized and post-translationally-modified peptides 10

1.3. Cyanobactins 11

1.4. Cyanobactin producers 11

1.5. Chemical diversity of cyanobactins 12

1.5.1. Bioactivities 23

1.5.2. Ecological role of cyanobactins 23

1.6. Biosynthesis of cyanobactins 24

1.6.1. Biosynthetic genes 24

1.6.2. Precursor peptide 25

1.6.3. Cyclization and cleavage 28

1.6.4. Heterocyclization 28

1.6.5. Prenylation 29

1.7. Analogous biosynthetic pathways in cyanobacteria 29

1.8. Significance and biotechnological potential 30

2. Aims of the study 32

3. Summary of materials and methods 33

4. Summary of results and discussion 34

4.1. Distribution of cyanobactin biosynthetic genes (I) 34

4.2. Anacyclamides – novel cyanobactins in the strains of *Anabaena* (II) 34

4.3. Piricyclamides – novel cyanobactins in the strains of *Microcystis* (III) 38

4.4. Genome mining of cyanobactin pathways (IV) 41

4.5. The discovery of linear cyanobactins (IV) 45

4.6. Cyanobactins in the strains of *M. aeruginosa* (III and IV) 49

4.7. Evolution of the cyanobactin pathway 50

5. Conclusions 52

6. Acknowledgements 54

7. References 55

List of original publications

This thesis is based on the following publications:

I Leikoski N, Fewer DP, Sivonen K (2009) Widespread occurrence and lateral transfer of the cyanobactin biosynthesis gene cluster in cyanobacteria. Applied and Environmental Microbiology 75: 853-857.

II Leikoski N, Fewer DP, Jokela J, Wahlsten M, Rouhiainen L, Sivonen K (2010) Highly diverse cyanobactins in strains of the genus *Anabaena*. Applied and Environmental Microbiology 76: 701–709.

III Leikoski N, Fewer DP, Jokela J, Alakoski P, Wahlsten M, Sivonen K (2012) Analysis of an inactive cyanobactin biosynthetic gene cluster leads to discovery of new natural products from strains of the genus *Microcystis*. PLoS One 7: E43002.

IV Leikoski N, Liu L, Jokela J, Wahlsten M, Gugger M, Calteau A, Permi, P, Kerfeld CA, Sivonen K, Fewer DP (2013) Genome mining expands the chemical diversity of the cyanobactin family to include highly modified linear peptides. Chemistry and Biology: in press.

The author's contribution

I Niina Leikoski participated in the design of the study. She performed the experiments, interpreted the results and wrote the manuscript.

II Niina Leikoski participated in the design of the study. She performed the experiments and participated in the analysis of the results. She wrote the manuscript.

III Niina Leikoski participated in the design of the study. She participated in the experimental work. She contributed to the analysis of the results and wrote the manuscript.

IV Niina Leikoski participated in the design of the study, experimental work and she contributed to the analysis of the results and wrote the manuscript.

Abbreviations

Proteinogenic amino acids:

Ala	(A)	Alanine
Arg	(R)	Arginine
Asn	(N)	Asparagine
Asp	(D)	Aspartic acid
Cys	(C)	Cysteine
Gln	(Q)	Glutamine
Glu	(E)	Glutamic acid
Gly	(G)	Glycine
His	(H)	Histidine
Ile	(I)	Isoleusine
Leu	(L)	Leusine
Lys	(K)	Lysine
Met	(M)	Methionine
Phe	(F)	Phenylalanine
Pro	(P)	Proline
Ser	(S)	Serine
Thr	(T)	Threonine
Trp	(W)	Tryptophan
Tyr	(Y)	Tyrosine
Val	(V)	Valine

bp	base pair
LC	liquid chromatography
MS	mass spectrometry
NMR	nuclear magnetic resonance
NRPS	non-ribosomal peptide synthetase
PCR	polymerase chain reaction
PRPS	post-ribosomal peptide synthesis
RiPPs	ribosomally synthesized post-translationally modified peptides
patA-G	genes encoding patellamide biosynthetic genes
PatA-G	proteins encoded by the *pat* genes
sp.	species

Abstract

Ribosomal peptides are produced through the post-translational modification of short precursor peptides. The chemical diversity of ribosomal peptides is enormous and they are found in all domains of life. Cyanobactins are a new family of ribosomal peptides produced by cyanobacteria. A great variety of cyclic peptides that can be assigned to the cyanobactin family are described from cyanobacteria or from marine animals that harbour cyanobacteria. Most cyanobactins have been found in screening programs to identify anti-cancer compounds and the first cyanobactin was reported from a marine tunicate in 1980. The biosynthetic origins of cyanobactins remained unclear until 2005 when they were shown to be produced through the post-translation modification of a short precursor peptide by the symbiotic cyanobacteria *Prochloron*. The precursor peptide is cleaved, N-to-C macrocyclized, and in some cases the amino acids can be heterocyclized or prenylated. As a family, the cyanobactins possess cytotoxicity, multidrug-resistance reversing activity, anti-malarial, antimicrobial, and allelopathic activities.

The aim of this study was to explore the distribution of cyanobactin pathways and determine the chemical diversity of the peptides they encode. Cyanobacterial strains were studied using a combination of bioinformatics, molecular biology, evolutionary biology, and structural chemistry.

A PCR-based approach demonstrated that the cyanobactin genes are common and sporadically distributed in bloom-forming cyanobacteria. The gene encoding the PatA protease was found in 48 of 132 cyanobacterial strains studied. A cyanobactin gene cluster was detected in the draft genome of bloom-forming cyanobacterium *Anabaena* sp. 90. The product of the pathway was identified as a cyclic peptide, consisting of proteinogenic amino acids, and was named anacyclamide. Diverse anacyclamides were identified from 27 strains of the genus *Anabaena*. Anacyclamides are cyclic peptides consisting of 7-20 proteinogenic amino acids which were prenylated or geranylated in some strains. Genome mining also identified an inactive cyanobactin gene cluster in *M. aeruginosa* NIES843. Based on genetic similarity, an active counterpart was found, and this led to the discovery of functional cyanobactin gene clusters in 6 strains of the genus *Microcystis* which produced new cyanobactins. These peptides, which were named piricyclamides, are similar to anacyclamides. Piricyclamides have 7 to 17 amino acids and like anacyclamides the only conserved amino acid is a single proline.

Bioinformatic analysis of the cyanobactin pathway in 126 cyanobacterial genomes identified 31 cyanobactin gene clusters. Surprisingly, two of the pathways identified in this study were shown to produce novel linear cyanobactins. These linear peptides contain a thiazole and an N-terminus protected with a prenyl group and methylated C-terminus. The linear cyanobactins, consisting of just 3-5 amino acids, were named aeruginosamides B, C, and viridisamide A. Genome mining demonstrated

that cyanobactin pathways were widely distributed among all cyanobacteria. Phylogenetic analyses subsequently showed that the cyanobactin pathway has a complex evolutionary history which differs from the evolutionary history of cyanobacteria.

The cyanobactin family was originally defined as being cyclic peptides containing heterocyclized amino acids and/or prenylated amino acids. In this study, altogether three types of cyanobactins were described, cyclic anacyclamides, piricyclamides, and the linear cyanobactins aeruginosamides and viridisamide A. This work broadened the cyanobactin definition to include unmodified cyclic peptides, and expanded the length variation of cyanobactins setting new upper and lower limits, 3–20, for the length of the peptides in the cyanobactin family. The definition of cyanobactin was further refined to include linear peptides with termini protected by prenylation and methylation. Overall, this study has shown that the cyanobactin pathway is common in cyanobacteria and able to produce a range of chemically diverse peptides.

Tiivistelmä

Bakteerit tuottavat paljon erilaisia pieniä ribosomaalisia peptidejä, jotka on tuotettu suuremmasta esipeptidistä muokkaamalla. Ribosomaalisia peptidien tuottajia löytyy kaikista eliökunnan kolmesta pääryhmästä. Syanobaktiinit ovat syanobakteerien tuottamia ribosomaalisia rengaspeptidejä, joita on kuvattu myös merieläimistä, joissa on syanobakteereita. Suurin osa syanobaktiineista on löydetty etsittäessä lääkeaineita syöpäsoluja vastaan. Ensimmäinen syanobaktiini kuvattiin vaippaeläimestä 1980. Syanobaktiinien alkuperä saatiin selville vasta 2005, kun biosynteesi kuvattiin vaippaeläimen symbioottisesta *Prochloron*-syanobakteerista. Syanobaktiinien biosynteesissä ensin tuotetaan esipeptidi, jota muokataan translaation jälkeen. Esipeptidistä katkaistaan lopullisen syanobaktiinin muodostava osa, josta muodostetuu rengas, kun aminohappoketjun amino- ja karboksyylipää yhdistetään. Joistakin syanobaktiinien aminohapoista on muodostettu heterosyklejä tai aminohappoja on prenyloitu. Syanobaktiinien on havaittu olevan soluille myrkyllisiä, antimikrobisia, lääkevastustuskyvyn alentavia ja malariaa vastaan toimivia vaikutuksia sekä niillä on havaittu myös olevan muiden eliöiden kasvuun vaikuttavia, allelopaattisia, ominaisuuksia.

Tässä työssä on tutkittu syanobaktiinien biosynteesigeenien esiintymistä syanobakteereissa sekä syanobaktiinien kemiallisten rakenteiden monimuotoisuutta. Tutkimus tehtiin syanobakteerikantoja tutkimalla ja yhdistelemällä bioinformatiikkaa, molekyylibiologiaa, evoluutiobiologiaa sekä kemiallisia menetelmiä.

Syanobaktiinien tuottamiseen tarvittavien geenien todettiin olevan yleisiä erityisesti massaesiintymiä eli kukintoja muodostavilla syanobakteereilla. Tutkituista 132 syanobakteerikannasta 48 löydettiin PCR-menetelmällä syanobaktiini biosynteesille oleellinen *patA*-geeni. *Anabaena* sp. 90 -syanobakteerin genomissa oli syanobaktiinien tuottamiseksi tarvittavat geenit. Geenien tuottama syanobaktiini löydettiin ja se oli rengaspeptidi, joka koostui tavallisista aminohapoista. Tämä uusi syanobaktiini nimettiin anasyklamidiksi. Anasyklamideja löydettiin 27 *Anabaena*-kannasta. Anasyklamidit ovat rengaspeptidejä, jotka koostuvat 7–20 tavallisesta aminohaposta ja toisinaan jokin aminohappo oli prenyloitu tai geranyloitu. *Microcystis aeruginosa* NIES 843 -kannan genomissa on toimimaton syanobaktiini geeniryhmä. Kuudesta *Microcystis*-kannasta löydettiin vastaavanlaisia, mutta toiminnallisia geenejä, jotka tuottivat uudenlaisia syanobaktiineja. Peptidit nimettiin pirisyklamideiksi. Pirisyklamideissa on 7–17 tavallista aminohappoa, joista viimeinen on aina proliini, kuten anasyklamideissakin. Myös pirisyklamideissa on prenyloituja tai geranyloituja aminohappoja. Joissakin pirisyklamideissa oli myös rikkisilta, joka yhdistää kaksi peptidirenkaassa olevaa kysteiiniä toisiinsa.

Bioinformaattisen analyysin avulla löydettiin 126 syanobakteerigenomista 31 syanobaktiinien tuottamiseen tarvittavaa geeniryhmää. Näistä kahden geeniryhmän havaittiin tuottavan lineaarisia syanobaktiineja, joissa on vain

3–5 aminohappoa. Näissä lineaarisissa syanobaktiineissa on kysteiinistä muokattu heterosyklinen tiatsolirengas, prenyyliryhmällä suojattu aminopää sekä metyloitu karboksyylipää. Uudet lineaariset syanobaktiinit nimettiin aeruginosamidi B:ksi ja C:ksi sekä viridisamidi A:ksi. Genomeja tutkimalla havaittiin, että syanobaktiinien tuottamiseksi tarvittavat geeniryhmät ovat laajalle levinneitä syanobakteerien joukossa. Fylogeneettisten analyysien avulla osoitettiin, että syanobaktiineja tuottavien geenien evolutiivinen historia on monimutkainen ja se eroaa syanobakteerien evolutiivisesta historiasta.

Aikaisemmin syanobaktiinit oli määritelty rengaspeptideiksi, joissa on heterosyklisiksi muokattuja tai prenyloituja aminohappoja. Tässä työssä kuvattiin kolmenlaisia syanobaktiineja; rengasmaiset anasyklamidit ja pirisyklamidit sekä lineaariset aeruginosamidit ja viridisamidi A. Työssä laajennettiin syanobaktiinien määritelmää käsittämään myös ainoastaan tavallisista aminohapoista koostuvia rengaspeptidejä sekä asetettiin uudet ääripituudet syanobaktiineille (3–20 aminohappoa). Syanobaktiinit määriteltiin nyt sisältämään myös lineaarisia peptidejä, joiden päät on suojattu prenyloimalla ja metyloimalla. Kaiken kaikkiaan työ osoitti, että syanobaktiinien tuottamiseksi tarvittavat geenit ovat yleisiä syanobakteereilla ja ne tuottavat suuren joukon erilaisia syanobaktiineja.

1. Introduction

1.1. Cyanobacteria

Cyanobacteria are aerobic photoautotrophs and require only light, carbon dioxide, water, and inorganic substances for growth. Cyanobacterial species can occupy terrestrial habitats and, more commonly, aqueous environments. They can survive extreme conditions, such as high salinity, high and low temperatures, and high levels of UV-radiation. The morphology of cyanobacteria is diverse with unicellular, colonial, and multicellular filamentous forms. Cyanobacteria form a distinct group in the phylogenetic tree based on bacterial 16S rRNA genes (Castenholz 2001). They are infamous for the production of toxins including the most studied group of secondary metabolites, the hepatotoxic microcystins (Sivonen and Jones 1999). Cyanobacteria are also rich in other secondary metabolites, some of which have interesting antimicrobial, anti-HIV, and anticancer activities (Burja et al. 2001, Welker and von Döhren 2006). The structural diversity of cyanobacterial secondary metabolites is tremendous, but their ecological role remains unclear (Burja et al. 2001, Welker and von Döhren 2006).

1.2. Ribosomally synthesized and post-translationally-modified peptides

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are produced in all three domains of life; Archaea, Bacteria and Eucarya (Arnison et al. 2013). These peptides are produced by a pathway now designated as post-ribosomal peptide synthesis (PRPS), analogous to non-ribosomal peptide synthetase (NRPS), enzymes that catalyse the biosynthesis of peptides by a modular pathway which acts independently of the ribosome (Arnison et al. 2013). In PRPS, an unmodified precursor peptide is produced by normal translation on the ribosome. This precursor peptide directly encodes the sequence that will form the end-product peptide (Figure 1). This region is termed the core sequence (Oman and van der Donk 2010 and Arnison et al. 2013). The precursor peptide is subsequently cleaved and modified to form the final end-product (Figure 1). A wide range of modifications has been described such as various cyclizations and linkages of the amino acid backbone, phosphorylation, hydroxylation, carboxylation, glycosylation, C-terminal amination, epimerization, lanthionine formation, prenylation, and heterocyclization (Arnison et al. 2013, Dunbar et al. 2013, McIntosh et al. 2009). Examples of RiPPs include bacterial lanthipeptides, microcins, lasso peptides in actinobacteria and proteobacteria, fungal amatoxins and phallotoxins, cyclotides and orbitides in plants, conopeptides in snails, and cyanobactins and microviridins in cyanobacteria (Oman and van der Donk 2010 and Arnison et al. 2013). RiPPs are mostly known for their antibacterial activities but other interesting bioactivities have been reported, such as anti-malarial, anticancer, and anti-HIV activities (Arnison et al. 2013). The ever increasing number of microbial genome sequences promotes the discovery of ribosomally synthesized natural products because

the product is directly linked to the precursor gene sequence (Velásques and van der Donk 2011).

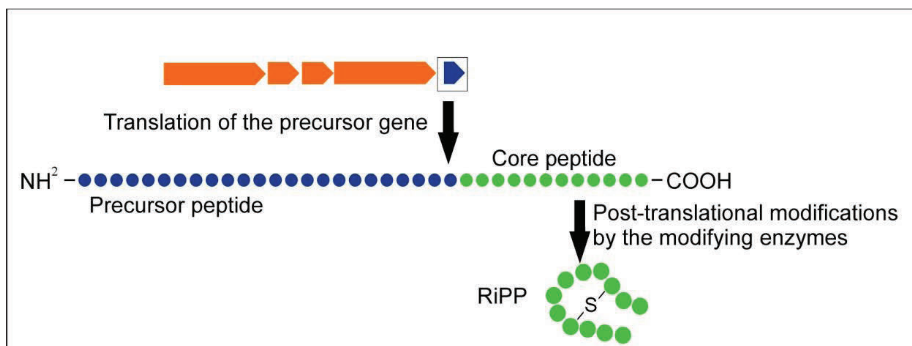


Figure 1. Schematic of post-ribosomal peptide synthesis (PRPS).

1.3. Cyanobactins

Cyanobactins are small cyclic peptides produced by many cyanobacteria (Donia et al. 2008a, Sivonen et al. 2010, Donia and Schmidt 2010). The first cyanobactins, ulicyclamides and ulithiacyclamides, were described in a tunicate *Lissoclinum patella* for the first time by Ireland and Scheuer (1980). Many cyanobactins are found from marine animals which harbour cyanobacteria (Donia and Schmidt 2010, Sivonen et al. 2010). The tunicate *L. patella* has an obligate symbiotic cyanobacterium, *Prochloron*, and it was hypothesised that the symbiont is the producer of the cyanobactins (Sings and Rinehart 1996). In fact, two separate but simultaneous studies proved that *Prochloron* synthesizes these metabolites, patellamides, by a PRPS pathway (Long et al. 2005, Schmidt et al. 2005). The biosynthetic pathway for patellamides was discovered based on genome sequencing of symbiotic *Prochloron* cells followed by heterologous expression (Schmidt et al. 2005) and in the concomitant study by shotgun cloning and heterologous expression of the clones directly (Long et al. 2005). In both of the studies, the PRPS pathway for patellamide biosynthesis from the unculturable *Prochloron* was expressed in a heterologous host demonstrating that patellamides are made by a PRPS instead of an NRPS pathway. The name cyanobactins was first introduced to this family of peptides by Donia et al. (2008a).

1.4. Cyanobactin producers

The vast majority of cyanobactins are described from marine animals, such as ascidians and sponges (Table 1), which often harbour cyanobacteria. In the ascidian *Lissoclinum patella* the obligate symbiont *Prochloron* has been proven to be the producer of cyanobactins (Schmidt et al. 2005). The biosynthetic genes have been described in the genome of *Prochloron* and it is highly likely that producers of all these closely related compounds in marine organisms (Table 1) are the cyanobacteria harbouring them (Schmidt et al.

2005, Donia and Schmidt 2010). Cyanobactins are also frequently found in free-living cyanobacteria from freshwater, oceanic, and terrestrial habitats (Table 2). The most frequent free-living producers of cyanobactins seem to be the strains of the genera *Microcystis* and *Oscillatoria* (Table 2). Based on a study of 126 cyanobacterial genomes cyanobactin gene clusters are found in 10 % of the genomes (Shih et al. 2013).

1.5. Chemical diversity of cyanobactins

The name cyanobactin refers to cyanobacterial, ribosomally-synthesised, cyclic peptides which may have prenylated or heterocyclized amino acids (Figure 2). Not all compounds that belong structurally to this group of peptides are proven to be synthesized by the cyanobactin pathway. Based on current knowledge, it is probable that all cyanobactin-like peptides (Tables 1 and 2) are biosynthetically related (Donia and Schmidt 2010, Sivonen et al. 2010). Cyanobactins consist of 7 to 20 amino acids, which are generally in the L-conformation, but D-amino acids are also occasionally found in these peptides. D-amino acids are often located next to a heterocyclized amino acid (Milne et al. 2006, McIntosh et al. 2009). Epimerisation in the cyanobactins is considered to be non-enzymatic (Milne et al. 2006, McIntosh et al. 2009). If the cyanobactin encodes a cysteine it is heterocyclized to thiazole or thiazoline, or forms di-sulfide bridge with another cysteine. Threonines and serines are often heterocyclized to oxazoles or oxazolines. The oxidation level of heterocycles varies but it seems that the oxidised forms are more common especially considering thiazoles. Sometimes the cyanobactins are prenylated or, much more rarely geranylated. The known prenylated amino acids are tyrosine, serine, threonine, and tryptophan (Table 1-2). In marine animals prenylated cyclic peptides are occasionally reported and these compounds are probably made by cyanobacteria. For example, myriastramides have *O*-prenylated tyrosine (Erickson et al. 2003) and mollamides have *O*-prenylated serine (Donia et al. 2008b). Some of the cyanobactins have di-sulfide bridges from cysteine to cysteine. The cyanobactin gene cluster does not encode an enzyme to carry out the formation of the disulfide bridge and it is most likely a spontaneous reaction (McIntosh et al. 2009). Biosynthetic pathways for halogenated cyanobactins have not been described, however, strains of the genus *Microcystis* produce cyclic peptides called aeruginazoles, which are proposed to be cyanobactins (Table 2) (Raveh and Carmeli 2010, Adiv et al. 2012). These cyanobactins are 12 to 17 amino acids with thiazoles and two of these contain chlorinated tyrosine (Adiv et al. 2012). Myriastramides are cyanobactin-like octapeptides detected from a sponge (Erickson et al. 2003). Myriastramide B contains a chlorinated *O*-prenyl moiety, which is attached to a tyrosine amino acid (Erickson et al. 2003).

Just a few of the cyanobactins consist solely of proteinogenic amino acids without any modifications beyond N-to-C macrocyclization. In addition to post-translational modifications, the huge diversity of cyanobactins is derived directly from the genetic variation of cyanobactin precursor peptide genes. The core region of the precursor gene is not conserved making the amino acid content of cyanobactins hypervariable. Linear compounds

structurally related to cyanobactins have also been described (Carrol et al. 1995, Nagatsu et al. 1995, Lawton et al. 1999, Table 3.). These linear tetrapeptides have methylated, decarboxylated, or prenylated C-terminus and the N-terminus is prenylated (Table 3 Carrol et al. 1995 Nagatsu et al. 1995, Lawton et al. 1999). The biosynthetic pathway of these metabolites has not been described, but the production of virenamide and aeruginosamide has been suggested to be ribosomal in origin (Schmidt and Donia 2010).

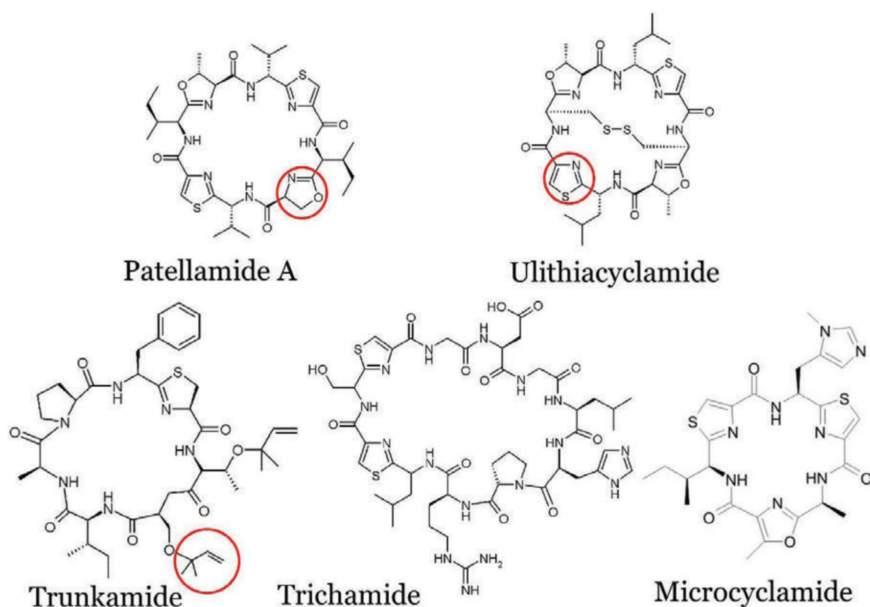


Figure 2. The chemical structure of selected cyclic cyanobactins. Patellamide, ulithiacyclamide, and trunkamide are produced by *Prochloron* (Schmidt et al. 2005, Donia et al. 2008a). Trichamide is produced by *Trichodesmium erythraeum* IMS101 (Sudek et al. 2006) and microcyclamide by *Microcystis aeruginosa* NIES 298 (Ishida et al. 2000, Ziemert et al. 2008b). The post-translational modifications; oxazoline, thiazole, and reverse *O*-prenyl are highlighted with circles.

Table 1. Examples of putative cyanobactins and their structural features characterized from marine animals. The peptides are organized alphabetically. The sequence of the peptide is presented in linear form so that the last amino acid is cysteine or proline. Number of each modification is given after the core peptide and the location of the prenyl is indicated by one letter amino acid abbreviation in the prenyl column.

Compound	Molecular weight	Number of amino acids	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	SS-bridge	Bioactivity	Source organism	Reference
Ascidiacyclamide	757.0	8	ITVCITVC		2	2				Cytotoxic	*Lissoclinum patella*	Ishida et al. 1988, Schmitz et al. 1989
Axinastatin 1	752.9	7	FVVPVNP							Cytostatic against cancer cells	*Axinella* sp.	Pettit et al. 1994
Axinastatin 2	766.9	7	FVLPVNP									
Axinastatin 3	781.0	7	FILPVNP									
Axinellin C	938.1	8	FPLTVPWP								*Stylotella aurantium*	Tabudravu et al. 2002a
Bistratamide A	570.7	6	ITACFC		1		2			Cytotoxic	*Lissoclinum bistratum*	Degnan et al. 1989b, Foster et al. 1992, Perez and Faulkner 2003
Bistratamide B	568.7	6	ITACFC		1	1	1					
Bistratamide C	503.6	6	VSACVC	1		2						
Bistratamide D	530.6	6	VTVSVC	1	1	1						
Bistratamide E	543.2	6	VTVCVC		1	2						
Bistratamide F	532.7	6	VTVSVC		2	1						
Bistratamide G	528.6	6	VTVSVC	2		1						
Bistratamide H	544.7	6	VTVCVC	1		2						
Bistratamide I	548.7	6	VTVSVC	1		1						
Bistratamide J	564.7	6	VTVCVC			2						
Comoramide A	668.9	6	ITFTAC		1		1	1T		Mild cytotoxicity against tumour cells	*Didemnum molle*	Rudi et al. 1998
Comoramide B	686.9	6	ITFTAC				1	1T				
Cyclodidemnamide	693.9	7	FTVPCVC		1	1	1			Weakly cytotoxic	*Didemnum molle*	Toske and Fenical 1995, Arrault et al. 2002
Cyclodidemnamide B	689.9	7	ITVPCLC			2						
Cyclonellin	963.1	8	YTANPRYP								*Axinella carteri*	Milanowski et al. 2004
Cycloxazoline	546.7	6	VTVTVT		3					Cytotoxic	*Lissoclinum bistratum*	Hambley et al. 1992
Didmolamide A	538.6	6	ACFTAC		1	2					*Didemnum molle*	Rudi et al. 2003
Didmolamide B	556.7	6	ACFTAC			2						

Compound	Molecular weight	Number of amino acids	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	SS-bridge	Bioactivity	Source organism	Reference
Dolastatin I	516.6	6	ITVSAC	1	1	1				Cytotoxic	*Dolabella auricularia* (sea hare)	Sone et al. 1997
Dolastatin E	490.6	6	ASACIC	1		1	1				*Dolabella auricularia*	Ojika et al. 1995
Haliclonamide A	857.0	8	PASYPTIP	1	1			1Y		Antifouling	*Haliclona* sp.	Guan et al. 2001, Sera et al. 2003
Haliclonamide B	788.9	8	PASYPTIP	1	1							
Haliclonamide D	806.9	8	PASYPTIP	1								
Haliclonamide E	875.0	8	PASYPTIP	1				1Y				
Hexamollamide	696.9	6	VVCTFP				1	1T		Moderate cytotoxicity against HeLa S3 cells	*Didemnum molle*	Teruya et al. 2008
Hymenamide A	880.1	7	VPFWRPP								*Hymeniacidon* sp. (sponge)	Kobayashi et al. 1993,
Hymenamide B	830.9	7	NFVEFPP									
Hymenamide C	826.9	7	FGPELWP							Immunomodulating activity		Napolitano et al. 2001,
Hymenamide D	769.9	7	YDPLAIP									
Hymenamide E	854.0	7	TTPYFFP									
Hymenamide F	765.0	7	AVMLRPP									Kobayashi et al. 1996, Tsuda et al. 1994
Hymenamide G	893.1	8	YVPLILPP									
Hymenamide H	904.1	8	LPWVPLTP									
Hymenamide J	1099.3	8	YDFWKVYP									
Hymenamide K	1007.2	8	YDFWKAVP									
Keenamide	621.3	6	IPGSLC				1	1S		Active against tumour cell lines	*Pleurobranchus forskalii*	Wesson et al. 1996
Leucamide	611.3	7	TASLPVC	2		1				Moderately cytotoxic against tumour cell lines	*Leucetta microraphis*	Kehraus et al. 2002
Lissoclinamide 1	705.9	7	ICFPTVC		1	2				Cytotoxic	*Lissoclinum patella**	Wasylyk et al. 1983, Degnan et al. 1989a, Schmitz et al. 1989, Donia et al. 2006, Hawkins et al. 1990, Morris et al. 2000
Lissoclinamide 2	679.9	7	ACFPTIC		1	1	1					
Lissoclinamide 3	679.9	7	ACFPTIC		1	1	1					
Lissoclinamide 4	741.9	7	FCFPTVC		1	1	1					
Lissoclinamide 5	739.9	7	FCFPTVC		1	2						
Lissoclinamide 6	741.9	7	FCFPTVC		1	1	1					
Lissoclinamide 7	743.9	7	FCFPTVC		1		2					
Lissoclinamide 8	741.9	7	FCFPTVC		1	1	1					

Compound	Molecular weight	Number of amino acids	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	SS-bridge	Bioactivity	Source organism	Reference
Lissoclinamide 9	707.9	7	VCFPTIC		1	1	1					
Lissoclinamide 10	726.9	7	ICFPTIC		1		2					
Mayotamide A	694.0	7	VPCICMC			1	2				*Didemnum molle*	Rudi et al. 1998
Mayotamide B	679.9	7	VPCVCMC			1	2					
Minimide	721.4	6	TLATIC				1	2S			*Didemnum molle*	Doniat et al. 2011
Mollamide	808.0	7	IPISFPC				1	1S		Cytotoxic	*Didemnum molle*	Donia et al. 2008b
Mollamide B	696.9	6	VFPTVC				1	1T		Cytotoxic against several cancer cell lines		
Mollamide C	618.8	6	IPGSLC			1		1S				
Myriastramide A	854.3	8	ITYPSAPP	2				Y			*Myriastra clavosa* (sponge)	Erickson et al. 2003
Myriastramide B	888.4	8	ITYPSAPP	2				Y†				
Myriastramide C	827.4	8	VSVWCVPP	1		1						
Nairaiamide A	802.1	7	VTIPIIP					1T			*Lissoclinum bistratum*	Foster and Ireland 1993
Nairaiamide B	816.1	7	ITIPIIP					1T				
Patellamide A	743.0	8	ISVCITVC		2	2				Cytotoxic antineoplastic, Patellamides B and C had multidrug reversing activity	*Lissoclinum patella**	Ireland et al. 1982, Fu et al. 1998, Schmitz et al. 1989, McDonald and Ireland 1992, Rashid et al. 1995, Schmidt et al 2005, Donia et al. 2006
Patellamide B	777.0	8	LTACITFC		2	2						
Patellamide C	763.0	8	VTACITFC		2	2						
Patellamide D	777.0	8	ITACITFC		2	2						
Patellamide E	791.0	8	VTVCITFC		2	2						
Patellamide F	763.0	8	VTVCITFC		2	2						
Patellamide G	795.0	8	ITACLTFC		1	2						
Patellin 2	733.0	6	TVPTLC				1	2T			*Lissoclinum patella**	Donia et al. 2008a, Zabriskie et al. 1990
Patellin 3	943.0	8	TLPVPTLC				1	2T		Moderately cytotoxic		
Patellin 6	963.2	8	TFPVPTVC				1	2T				
Phakellistatin 1	828.0	7	YPIPIFP							Cancer cell growth inhibitory	*Phakellia costata* (sponge) and *Stylotella aurantium*	Pettit et al. 1993a, b,
Phakellistatin 2	828.0	7	YPFPIIP							Cell growth inhibitory	*Phakellia carteri*	Tabudravu et al. 2002b
Phakellistatin 7	1109.4	10	YIPPIFALPP							Cancer cell growth inhibitory	*Phakellia costata* sponge	Pettit et al. 1995
Phakellistatin 8	1137.4	10	YIPPIFVLPP									
Phakellistatin 9	1123.4	10	YVPPIFVLPP									

Compound	Molecular weight	Number of amino acids	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	SS-bridge	Bioactivity	Source organism	Reference
Phakellistatin 12	1139.4	10	IFTLPPYIPP								*Phakellia* sp.	Pettit and Tan 2003,
Phakellistatin 13	798.9	7	TLWPFGP								*Phakellia fusca* (sponge)	Li et al. 2003
Stylisin 1	828.0	7	LPYPIFP								*Stylissa caribica* (sponge)	Mohammed et al. 2006
Stylisin 2	812.0	7	IPYPPFP									
Stylissamide A	845.0	7	VYPYKPP								*Stylissa caribica* (sponge)	Schmidt et al. 2007
Stylissamide B	812.0	7	IYPFPPP									
Stylissamide C	862.0	7	FIPYPFP									
Stylissamide D	828.0	7	FIPYPLP									
Tawicyclamide A	806.1	8	VCFCICVP			2	1			Weakly cytotoxic	*Lissoclinum patella*	McDonald et al. 1992
Tawicyclamide B	772.0	8	VCLCICVP			2	1					
Trunkamide A	838.0	7	TSIAPFC				1	2TS		Cytotoxic, multidrug reversing activity	*Lissoclinum patella**	Caba et al. 2001, Salvatella et al. 2003, Donia et al. 2008a
Ulicyclamide	677.8	7	FPTICAC		1	2				Not reported	*Lissoclinum patella*	Ireland and Scheuer 1980, Wasylyk et al. 1983
Ulithiacyclamide	763.0	8	CTLCCTLC		2	2			1	Cytotoxic	*Lissoclinum patella**	Ireland and Scheuer 1980,
Ulithiacyclamide B	797.0	8	CTFCCTLC		2	2			1			Ireland et al. 1982,
Ulithiacyclamide E	833.0	8	CTFCCTLC			2			1			Williams et al. 1989,
Ulithiacyclamide F	815.0	8	CTFCCTLC		1	2			1			Fu et al. 1998
Ulithiacyclamide G	815.0	8	CTFCCTLC		1	2			1			
Wainunuamide	745.9	7	GLFPHPP							Weakly cytotoxic	*Stylotella aurantium* (sponge)	Tabudravu et al. 2001
Waiakeamide	812.2	6	FPMPPMC				1			Antifouling	*Haliclona* (sponge)	Mau et al. 1996, Sera et al. 2003

*Compounds are confirmed to be produced by the *Prochloron* spp. symbiotic partners of ascidian (indicated with an asterisk after the source organism) and the biosynthetic genes have been described.

†Chlorinated prenyl

†† This table has been modified from the original table presented in Sivonen et al. 2010.

Table 2. The structural features of cyanobactins and putative cyanobactins. The peptides are organized according to the alphabetical order of the compound name. The sequence of the peptide is presented in linear form so that the last amino acid is cysteine, proline or threonine. Number of heterocyclized amino acids is given after the core peptide. This table includes also the cyanobactins described in the publications II and III.

Compound	Molecular weight	Number (aa)	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	Geranyl	S-S bridge	Bioactivity	Producer	References
Aerucyclamide A	534.7	6	ITGCIC		1	1	1				Toxic to fresh water crustacean Thamnocephalus platyurus, antimalarial	*Microcystis aeruginosa* (bloom-forming, planktonic)	Ziemert et al. 2008b, Revised structures and renaming: Portmann et al. 2008a, b
Aerucyclamide B	532.7	6	ITGCIC		1	2							
Aerucyclamide C	516.6	6	ATVSIC	1	1	1							
Aerucyclamide D (formerly microcyclamide 7806A and B)*	586.7	6	FTGCMC		1	1	1						
Aestuaramide 1–6*	794.0	7	VCMPCYP			2		1Y				*Lyngbya aestuarii* PCC 8106	Donia et al. 2008, McIntosh et al. 2013
Aestuaramide 7–12*	822.0	7	ACMPCYP			2		1Y					
Aeruginazole A	1155.4	12	GGGVFVCLCYNC			3					Antimicrobial	*Microcystis* sp. strain IL-323	Raveh and Carmeli 2010
Aeruginazole DA1497	1497.5	17	IAGFCGPLCVGPVCSAC			4					Antimicrobial (DA1497)	*Microcystis* bloom	Adiv et al. 2012
Aeruginazole DA1304	1304.4	13	GGSGFFICLCYNC			3							
Aeruginazole DA1274	1274.4	13	GGGGFFICLCYNC			3							
Aeruginazole DA1338	1338.4	13	GGSGFFICLCYNC †			3							
Aeruginazole DA1372	1372.4	13	GGSGFFICLCYNC †			3							
Agardhipeptin A	834.0	7	WGLHGWP								Plasmin inhibitory	*Oscillatoria agardhii* NIES204 (freshwater)	Shin et al. 1996
Agardhipeptin B	1036.3	8	WAPWVWLP										
Anacyclamide A10*	1053.2	10	TSQIWGSPVP									*Anabaena* e.g. 90	Article II
Anacyclamide B10*	1010.2	10	SSVIWGSPVP									*Anabaena* e.g. 1TU31S9	
Anacyclamide C10*	1115.2	10	SAQWQNFGVP									*Anabaena* e. g. 299A	
Anacyclamide D10*	1151.3	10	NAHWQNFGVP									*Anabaena* 1TU44S9	
Anacyclamide E10*	1087.3	10	YAPLQNFGVP									*Anabaena* PH256	
Anacyclamide E10P*	1155.4	10	YAPLQNFGVP					1				*Anabaena* PH256	
Anacyclamide A11*	1281.4	11	DNWLGEWIGIP									*Anabaena e.g. BIR260*	

Compound	Molecular weight	Number (aa)	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	Geranyl	S-S bridge	Bioactivity	Producer	References
Anacyclamide A15*	1690.9	15	HAFIGYDQDPTGKYP									*Anabaena* 37	
Anacyclamide A15G*	1827.1	15	HAFIGYDQDPTGKYP					1				*Anabaena* 37	
Anacyclamide A7*	948.1	7	RERFVYP									*Anabaena* 1TU39S8	
Anacyclamide F10P*	1191.3	10	YSNKPSDFSP					1				*Anabaena* 1TU32S11	
Anacyclamide A9P*	1114.3	9	YDDKLNLSP					1				*Anabaena* 1TU33S10	
Anacyclamide A20P*	2080.4	20	WGNGTGLDWKLLTGGISASP					1				*Anabaena* 202A1	
Anacyclamide A20PP*	2148.5	20	WGNGTGLDWKLLTGGISASP					2				*Anabaena* 202A1	
Anacyclamide A8P*	993.1	8	HQPWHAAP					1				*Anabaena* TR232	
Anacyclamide B8*	957.1	8	FSPDWRAP									*Anabaena* SYKE816	
Anacyclamide C8*	998.2	8	VIQHYLFP									*Anabaena* PH262	
Anacyclamide B7*	762.0	7	LIGIMHP									*Anabaena* SYKE844B	
Arthrospiramide A* Arthrospiramide B*	866.3 924.4	8 8	GLPGMPCFC GVLGMPCFC			2 2						*Arthrospira platensis* NIES 39 (Saline lake)	Donia and Schmidt 2011
Banyascyclamide A Banyascyclamide B Banyascyclamide C	538.6 522.7 556.7	6 6 6	ACFTAC ACLTAC ACFTAC		1 	2 2 2						*Nostoc* sp. TAU strain IL-235 (bloom-forming, planktonic)	Ploutno and Carmeli 2002
Cyanothecamide A* Cyanothecamide B* Cyanothecamide C*	830.4 828.4 729.3	7 7 7	VLYKDQC VLYKDQC KLGDSYC		 1	 1 	1 1					*Cyanothece* PCC 7425	Houssen et al. 2012
Dendroamide A Dendroamide B Dendroamide C	488.6 520.6 536.6	6 6 6	ACATVC ATACMC ATACMC	1 1 1		2 2 2					A - Multidrug-resistance reversing activity	*Stigonema dendroideum* (terrestrial)	Ogino et al. 1996
Dolastatin 3 Homodolastatin 3	660.8 674.8	7 7	VPLQCGC IPLQCGC			2 2						*Lyngbya majuscula*	Mitchell et al. 2000
Kawaguchipeptin A	1421	11	WLNGDNNWSTP					2 W				*Microcystis aeruginosa* NIES 88 (freshwater)	Ishida et al. 1996
Kawaguchipeptin B	1285	11	WLNGDNNWSTP								Antibacterial	*Microcystis aeruginosa* NIES 88	Ishida et al. 1997

Compound	Molecular weight	Number(aa)	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	Geranyl	S-S bridge	Bioactivity	Producer	References
Kororamide	969.2	7	LYCNPSLC			1	1					*Lyngbya majuscula* (Marine lake)	Mitchell et al. 2000
Microcyclamide*	582.7	6	HCATIC	1		2					Moderate cytotoxicity against P388 murine leukaemia cells (IC50 - 1.2 µg/ml	*Microcystis aeruginosa* NIES 298 (freshwater)	Ishida et al. 2000, Ziemert et al. 2008b
Microphycin AL828	829	8	EIGVYGLP									*Microcystis* sp. TAU strain IL-306	Gesner-Apter and Carmeli 2008
Nostocyclamide	474.6	6	ATGCVC	1		2					Anticyanobacterial and antialgal, toxic also to the rotifer Bracchionus calyciflorus	*Nostoc* sp. 31	Todorova et al. 1995
Nostocyclamide M	506.6	6	ATGCMC	1		2					Allelopathic anticyanobacterial	*Nostoc* sp. 31	Jüttner et al. 2001
Oscillacyclin	995	9	FTTSIAYNP									*Oscillatoria agardhii* 97 (freshwater)	Fujii et al. 2000
Planktocyclin	801	8	GLVMFGVP								Inhibitor of mammalian trypsin and α-chymotrypsin	*Planktothrix rubescens* (freshwater)	Baumann et al. 2007
Piricyclamide*	1162 1434 1368 1533	9 12 12 14	MSGVDYYNP NEFMQTGSYSGP TFCDLATKQCYP WILLADGTRPKNAP						1 1	 1		*Microcystis aeruginosa* PCC 7005	Article III
Piricyclamide*	1543	15	TLGCMNGTERCLGLP							1		*Microcystis* Izancya 36 and 41	Article III
Piricyclamide*	914 1397	8 12	FAIFLLLP ILGEGEGWNYNP					 1				*Microcystis* Izancya 42	Article III
Piricyclamide*	983	9	GTHLYTYTP									*Microcystis aeruginosa* SYKE864	Article III

Compound	Molecular weight	Number (aa)	Amino acids of the core peptide (N-to-C cyclic)	Oxazole	Oxazoline	Thiazole	Thiazoline	Prenyl	Geranyl	S-S bridge	Bioactivity	Producer	References
Prenylagaramide A	1081	9	YGTGEFFNP					Y				*Oscillatoria agardhii* NIES 205 and NIES 596 (freshwater)	Murakami et al. 1999
Prenylagaramide B* Prenylagaramide C*	929	7 9	INPYLYP QAYLGIPLP					Y Y				*Oscillatoria agardhii* NIES 596 (freshwater)	Murakami et al. 1999 Donia and Schmidt 2011
Raocyclamide A Raocyclamide B	550.6 568.7	6	FSASIC FSASIC	1 1	1	1 1					A- moderate toxicity against sea urchin embryos	*Oscillatoria raoi* (soil isolate)	Admi et al. 1996
Tenuecyclamide A* Tenuecyclamide B* Tenuecyclamide C* Tenuecyclamide D*	460.5 460.5 506.6 522.6	6 6 6 6	ATACAC ATACAC ATGCMC ATGCMC	1 1 1 1		2 2 2 2					Inhibited division of sea urchin embryos (B not tested)	*Nostoc spongiaeforme* var. *tenue* (litophytic)	Banker and Carmeli 1998, Donia et al. 2008
Trichamide*	1099.2	11	GDGLHPRLCSC			2					No effects found (tested for cytotoxicity, antifungal, antibacterial and antiviral activities)	*Trichodesmium erythraeum* IMS 101 (marine)	Sudek et al. 2006
Venturamide A Venturamide B	488.6 518.6	6 6	ATACVC ATTCVC	1 1		2 2					Strong antimalarial activity	*Oscillatoria* (marine)	Linington et al. 2007
Westiellamide	546.7	6	VTVTVT		3						Mildly cytotoxic	*Westiellopsis prolifica* (terrestrial)	Prinsep et al. 1992
Wewakazole	1141.3	11	GVTFSFPISAPP	3								*Lyngbya majuscula* (marine)	Nogle et al. 2003

* The cyanobactins which are shown to be products of ribosomal pathway are indicated with an asterisk after the cyanobactin.

† Aeruginazole DA1338 chlorinated at tyrosine (Y) and DA1372 di-chlorinated at Y

†† This table has been modified from the original table presented in Sivonen et al. 2010.

Table 3. Linear peptides, which resemble cyanobactins. This table includes also the cyanobactins described in the manuscript IV.

Compound	Molecular weight	Number of amino acids	Amino acids of the core peptide	C-terminus	N-terminus	Bioactivity	Producer	Reference
Aeruginosamide	561.3	4	IVPC	Methylated	Prenylated	Mildly cytotoxic to tumour and leukaemia cells	*Microcystis aeruginosa* bloom	Lawton et al. 1999
Aeruginosamide B Aeruginosamide C	574 675	4 5	FFPC FFPVC	Methylated Methylated	Prenylated Prenylated		*Microcystis aeruginosa* PCC 9432	Article IV
Virenamide A Virenamide B Virenamide C	469.3 427.2 566.3	4 4 4	FVVC FFVC FFFC	Decarboxylated Decarboxylated Decarboxylated	Di-prenylated Prenylated Prenylated	Moderate cytotoxicity	*Diplosoma virens* (Ascidian)	Carrol et al. 1995
Muscoride		4	VPTT	Prenylated	Prenylated	Weakly antibacterial against B. subtilis	*Nostoc muscorum* IAM M14	Nagatsu et al. 1995
Viridisamide A	444	3	FIC	Methylated	Prenylated		*Oscillatoria nigro viridis* PCC 7112	Article IV

1.5.1. Bioactivities

The cyclic peptide structure is promising in drug development (Wipf 1995, Driggers et al. 2008). The macrocyclic structures have good properties for drug candidates such as good solubility, lipophility, stability, and bioavailability (Driggers et al. 2008). Cyclic peptides lack ionized C- and N-termini, which makes the passage of such peptides across membranes easier, whilst increasing the resistance of the structure to enzyme degradation *in vivo*, as well as improving the bioavailability of the peptide (Wipf 1995). Other properties, which affect the bioactivity of the peptide, include the presence of heterocyclized amino acids and their level of oxidation. Minor changes in the oxidation level of heterocycles can change the bioactivity dramatically (Houssen and Jaspars 2010). Many cyanobactins have been discovered in the screening of potential bioactive compounds in different types of assays against various cancer cell lines (Table 1–3). As a consequence, most of the cyanobactins from marine animals have been shown to be cytotoxic (Donia and Schmidt 2010, Tables 1–3) in contrast to many other ribosomally-synthesised, modified peptides with antibacterial activities (Arnison et al. 2013). Studies on ulithiacyclamide, which has promising anticancer properties, were halted because of their high toxicity in animal models (Houssen and Jaspars 2010). Cyanobactins also have other interesting bioactivities, such as multi-drug resistance reversing properties, anti-malarial, anti-viral, and allelopathic activities (Tables 1–3). In many cases there is no bioactivity reported. The evaluation of the bioactivity of cyanobactins is heavily dependent on the amount of the compound produced by the cyanobacterium. Many novel compounds await the development of a synthetic route for bioactivity investigations to commence (Wipf 1995). New developments in combinatorial biosynthesis and heterologous expression systems for cyanobactins may enable more bioactivity studies when the production levels become sufficient.

1.5.2. Ecological role of cyanobactins

The ecological role of cyanobactins is not understood. In contrast to traditional ribosomal peptides such as bacteriocins, which usually have antibacterial activities (Arnison et al. 2013), the cyanobactins have often been found in biomedical screenings (Tables 1-3). Cyanobactins have many described bioactivities, but allelopathy against competing cyanobacteria and antibacterial activities are described for only a few cyanobactins (Todorova et al. 1995, Ishida et al. 1997, Jüttner et al. 2001, Raveh and Carmeli 2010, Adiv et al. 2012). These activities may yield information about the ecological function of cyanobactins, unlike the bioactivities found by biomedical screening. The cyanobactins found from immobile marine animals have been shown to have antifouling properties to protect from fouling organisms (Sera et al. 2002 and 2003). The metal binding properties of *Lissoclinum patella* metabolites have been studied and molecular shape seems to fit for metal binding e.g. Cu (II) and Zn (II) (Morris et al. 2001ab, Houssen and Jaspars 2010, Donia and Schmidt 2010). The metal binding propertyof these metabolites has not been shown to have a role in the ecology of *L. patella* and

Prochloron symbiosis, however, it has been reported in many studies (Houssen and Jaspars 2010). An increasing number of genome sequences has led to the discovery of a number of new ribosomal pathways in cyanobacteria and, in many of these, the bioactivity or the ecological role is unknown (Haft et al. 2010, Li et al. 2010, Wang et al. 2011, Tang and van der Donk 2012).

1.6. Biosynthesis of cyanobactins

Cyanobactins are made through PRPS (Post Ribosomal Peptide Synthesis) (Arnison et al. 2013). This means that a gene encoded precursor peptide is first transcribed. The precursor peptide is 50–150 amino acids long. The final peptide is modified and cleaved (Donia and Schmidt 2010, Sivonen et al. 2010). Cleavage of the cyanobactin precursor peptide takes place at a minimum of two sites. A hypervariable amino acid core sequence inside these will ultimately form the cyanobactin (Figure 3) (Donia et al. 2006, Donia et al. 2008). In addition to the cleavage, the cyanobactin precursor is N-to-C macrocyclized and some of the amino acids can be modified (Donia and Schmidt 2010, Oman and van der Donk 2010, Sivonen et al. 2010). Generally the cysteines, threonines, and serines are heterocyclized to thiazolines and oxazolines, which can be oxidised to thiazoles and oxazoles (Figure 2). Occasionally the core sequence is also prenylated (Figure 2) (Donia et al. 2008a, Donia and Schmidt 2011).

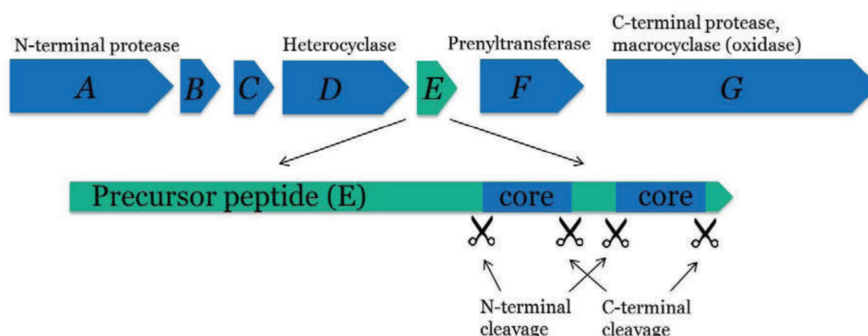


Figure 3. Schematic figure of the cyanobactin gene cluster and the structure of the precursor peptide. The functions of the genes are shown. The genes B and C are essential, but the function is unknown. The cyanobactin structures are formed from the core region shown in the precursor peptide.

1.6.1. Biosynthetic genes

A typical set of genes are generally present in the cyanobactin biosynthetic gene clusters (Figure 3). The gene cluster is approximately 10 kb and consists of the genes *A*-*G*, but the order of these genes can vary and cyanobactin gene clusters often encode an excess of additional short hypothetical proteins (Figure 4). The precursor gene is designated with the letter *E*. The

cyanobactin gene cluster encodes two protease genes, *A* and *G* (Donia and Schmidt 2010, Sivonen et al. 2010). These proteases are responsible for the cleavage of the precursor and N-to-C macrocyclization. The A-protease cleaves the N-terminal cleavage site of the core sequence and G-protease cleaves at the C-terminal cleavage site (Lee et al. 2009, Figure 2). The G-protease has a macrocyclase domain which N-to-C cyclizes the cyanobactin. It can also have a domain which oxidises the heterocycles (Agarwal et al. 2012, Koehnke et al. 2012). In a small number of cyanobactin gene clusters the oxidase is encoded in a separate gene (Sudek et al. 2008, Houssen et al. 2012). Both the A and G proteases also have a domain of unknown function (Agarwal et al. 2012). The D-protein is a heterocyclase which modifies cysteines, serines, and threonines to heterocycles (McIntosh et al. 2010, McIntosh and Schmidt 2010). The cyanobactin gene clusters encode PatF, that in some cases, has been shown to prenylate the cyanobactins (Donia and Schmidt 2011, McIntosh et al. 2011). However, the *patF*-gene is also present in the non-prenylating pathways. Additionally, the cyanobactin gene clusters encode *B* and *C* genes, but these genes have no assigned function (Donia and Schmidt 2010, Sivonen et al. 2010).

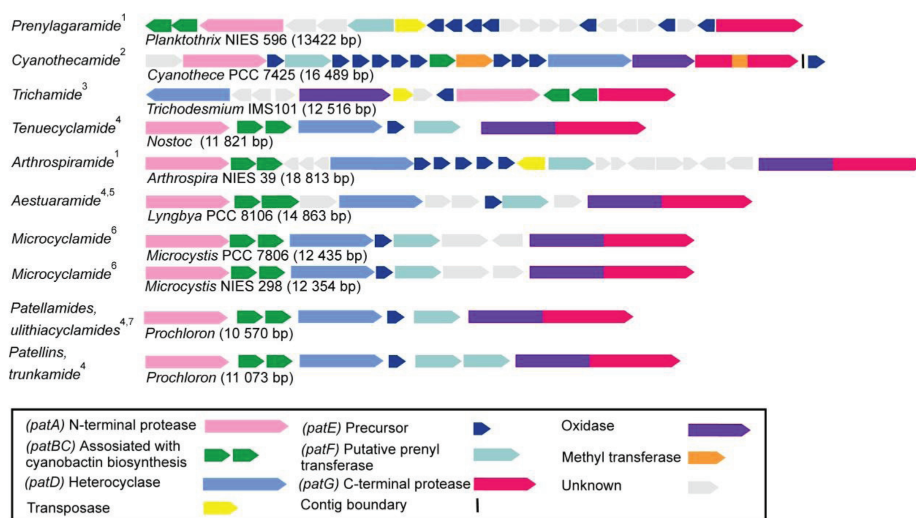


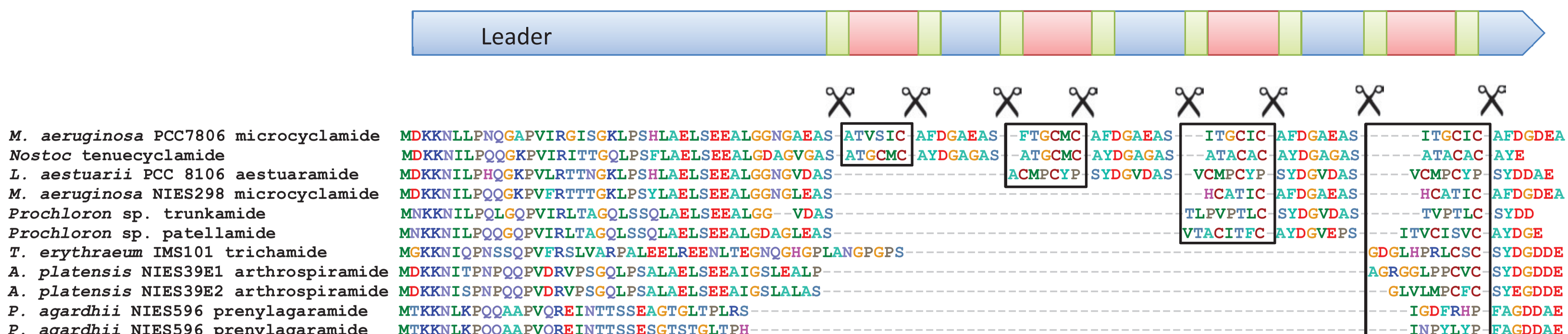
Figure 4. Cyanobactin gene clusters described in the literature. The length of the gene cluster is indicated. [1]Donia and Schmidt 2011, [2]Houssen et al. 2012, [3]Sudek et al. 2006, [4]Donia et al. 2008a, [5]McIntosh et al. 2013, [6]Ziemert et al. 2008b and [7]Schmidt et al. 2005.

1.6.2. Precursor peptide

The precursor peptide is an essential component in the ribosomally synthesised and post-translationally modified peptides (Arnison et al. 2013). The cyanobactin structure is directly encoded in the precursor peptide and the core determines the amino acid backbone of the cyanobactin (Figure 5) (Donia and Schmidt 2010 and Sivonen et al. 2010). The structure and the

sequence of the cyanobactin precursor peptide is conserved (Oman and van der Donk 2010, Houssen et al. 2010). The N-terminal part of the precursor peptide encodes a conserved leader sequence which is thought to be recognized by some of the modifying and cleaving enzymes (Houssen et al. 2010). The sites surrounding the core sequence are recognition sequences (Oman and van der Donk 2010, Arnison et al. 2013) (Figure 5). The number of core sequences varies from one to four. In the tenuecyclamide (Donia et al. 2008) and microcyclamide (Ziemert et al. 2008b) pathways, the precursor peptide has four cores in contrast to e.g. trichamide (Sudek et al. 2006), which has a single core sequence (Figure 5). The number of precursor genes varies from 1 to 10. Ten precursor genes have been described in the cyanothecamide biosynthetic cluster and one of the precursor genes is located on a plasmid (Houssen et al. 2012). Hence, there are two ways to enhance production of cyanobactins depending on the number of precursor genes or core sequences (Sivonen et al. 2010). The precursor peptide can be used for bioinformatic prediction of the cyanobactin product by aligning and comparison of the cleavage sites to known ones (Schmidt et al. 2005, Ziemert et al. 2008b, Donia et al 2008, Donia and Schmidt 2011). This information and knowledge of the enzymes carrying out post-translational modification in the biosynthetic pathway enables the prediction of the cyanobactin (Sudek et al. 2006, Ziemert et al. 2008b, Schmidt and Donia 2011, Houssen et al. 2012).

Figure 5. A schematic figure of a cyanobactin precursor peptide. The schematic precursor peptide has four core sequences (red) and recognition sequences (light green) surrounding the core sequences. Corresponding cleavage sites are shown in the cyanobactin precursor peptide alignment below. The producing strain and cyanobactin product are shown. The cleavage sites are indicated and the core sequences are highlighted with boxes.



1.6.3. Cyclization and cleavage

The cyanobactin gene clusters encode two proteases which both contain subtilisin-like serine protease domains (Schmidt and Donia 2009). Lee et al. (2009) demonstrated *in vitro* that the PatG protease from the *Prochloron* patellamide pathway cleaves the C-terminal cleavage site and catalyses the macrocyclization of natural and artificial peptides (Lee et al. 2009). They also demonstrated that the *Prochloron* PatA-protease recognizes the N-terminal cleavage site of the precursor (Lee et al. 2009). The substrate tolerance of the PatG protease has been studied with synthetic substrates and the PatG protease from *Prochloron* was shown to macrocyclise a broad array of synthetic substrates even with non-proteinogenic and D-amino acids, but at a lower rate (McIntosh et al. 2010). Substrate peptides with 6 to 11 amino acids were circularized. The recognition site for the cleavage was shown to be 4 to 5 amino acids (McIntosh et al. 2010). The crystal structure of the protease-macrocyclase domain of the PatG from the patellamide and prenylagaramide pathway was studied and the catalytic triad Asp548, His618, and Ser783 was shown (Agarwal et al. 2012, Koehnke et al. 2012). The crystal structure of highly homologous protease domain of the PatA in the patellamide pathway has also been determined (Agarwal et al. 2012). In the macrocyclization process, first the PatA protease removes the amino terminal next to the core sequence to yield a free amino terminal, and the PatG protease removes the carboxy terminal sequence flanking the core (Agarwal et al. 2012, Koehnke et al. 2012). The cleavage site is protected by the PatG protease preventing access to water, stopping the hydrolysis until the transamidation reaction is complete (Agarwal et al. 2012, Koehnke et al. 2012).

1.6.4. Heterocyclization

The heterocyclase (D) was shown to catalyze the heterocyclization of cysteines, threonines, and serines to thiazolines or oxazolines (Donia and Schmidt 2010, Sivonen et al. 2010). Heterocyclase D has been described in detail from the patellamide and trunkamide pathway (McIntosh and Schmidt 2010). The pathways which encode the heterocyclase also encode an oxidase domain fused with the G-protease or standalone oxidase which can oxidise the oxazolines and thiazolines to oxazoles and thiazoles (Donia and Schmidt 2010, Sivonen et al. 2010). It has been shown that the heterocyclase enzyme (D) alone is responsible for the hetrocyclization activity (McIntosh and Schmidt 2010). Heterocyclization was shown to take place before cleavage and macrocyclization for the precursor peptide as the heterocyclase enzyme needs the leader sequence to function in the patellamide and trunkamide pathway. The A-protease is capable of cleaving the unmodified precursor peptide before or after it is processed by heterocyclase but heterocyclization is dependent on the leader sequence (McIntosh and Schmidt 2010).

1.6.5. Prenylation

The prenylation of cyanobactins is carried out by the prenyl transferase (F). Whilst the gene encoding the prenyl transferase is present in all cyanobactin gene clusters, only a few cyanobactins are actually prenylated (Schmidt and Donia 2009, Sivonen et al. 2010). Prenylating cyanobactin pathways include the prenylagaramide pathway in *Planktothrix agardhii* NIES 596 (Donia and Schmidt 2011), the aestuaramide pathway in *Lyngbya aestuarii* PCC 8106, and the trunkamide pathway in *Prochloron* (Donia et al. 2006). Prenylagaramide has *O*-prenylated tyrosine (Donia and Schmidt 2011), and in the trunkamide has *O*-prenylated threonine and serine (Donia et al. 2006). LynF prenyltransferase from *Lyngbya aestuarii* PCC 8106 has been studied in detail *in vitro* and it was shown to prenylate using dimethylallyl pyrophosphate (DMAPP) on the oxygen atom of tyrosine. This reverse *O*-prenylated tyrosine undergoes spontaneous Claisen rearrangement yielding *C*-prenylated product (McIntosh et al. 2011). The cyanobactins in *L. aestuarii* PCC 8106 were already predicted in 2008 (Donia et al. 2008), but the cyanobactins were recently discovered and named aestuaramides (McIntosh et al. 2013). The structure was as predicted except for the oxidation level of the methionine and the form of the isoprenoid unit creating 12 variants from two core sequences (McIntosh et al. 2013). The oxidation of the methionine may be spontaneous and it may be increased during sample preparation, but the prenylation is made by LynF (McIntosh et al. 2011, McIntosh et al. 2013). In addition to unprenylated aestuaramides, two types of prenyls were found. It was shown that the aestuaramides are first reverse *O*-prenylated on tyrosine which can spontaneously undergo the Claisen reaction to *C*-prenyl as the *in vitro* studies had indicated (McIntosh eth al. 2011, McIntosh et al. 2013).

1.7. Analogous biosynthetic pathways in cyanobacteria

In addition to cyanobactins, cyanobacteria produce other post-translationally modified peptides such as bacteriocins and microviridins (Philmus et al. 2008, Ziemert et al. 2008a, Li et al. 2010). Microviridins are tricyclic peptides cross-linked by ester and amide bonds processed from a precursor peptide (Philmus et al. 2008, Ziemert et al. 2008a). Bacteriocin biosynthetic genes seem to be very common in cyanobacteria (Wang et al. 2011, Shih et al. 2013). The majority of cyanobacterial genomes were shown, in two distinct studies, to encode bacteriocin biosynthetic genes (Wang et al. 2011, Shih et al. 2013). Cyanobacterial bacteriocins are described in detail in *Prochlorococcus*, which can produce up to 29 different bacteriocins (Li et al. 2010, Tang and van der Donk 2012). In the prochlorosin pathway in *Prochlorococcus*, a single enzyme ProcM modifies the 29 linear precursor peptides to highly variable polycyclic peptides with thioether cross-links of lanthionine and methyllanthionine (Tang and van der Donk 2012). Multiple bacteriocin biosynthetic gene clusters appear in one cyanobacterial genome. In a comparative analysis of 58 genomes, 145 bacteriocin gene clusters were found encoding altogether 290 putative bacteriocin precursors (Wang et al.

2011). Another bioinformatic study showed the wide distribution of bacteriocin pathways in different bacterial phyla and in that study cyanobactin pathways were included (Lee et al. 2008). An enormous variety of RiPPs such as bacteriocins, microcins, and lantibiotics are produced by other bacteria (Arnison et al. 2013). Most of these RiPPs have been extensively studied because of their interesting antimicrobial activities and all are derived from a precursor peptide that is post-translationally modified (Nolan and Walsh 2008, Arnison et al. 2013).

Cyanobactins form a distinct group with shared and unique features in this enormous group of ribosomal natural products (Arnison et al. 2013). The unique feature of cyanobactins is in their biosynthesis. The precursor has multiple cleavage sites, which requires two proteases. Shared features of cyanobactins include amino acid heterocyclization, oxidation of heterocycles, prenylation, methylation, and epimerisation. A classification scheme has been suggested for cyanobactins and it was based on corresponding genotypes and cyanobactin structure (Schmidt and Donia 2011). Cyanobactins have also been classified based strictly on the structure like bacteriocins (Cotter et al. 2013). In the bacteriocin classification, cyanobactins are suggested to be divided into two subsections (Cotter et al. 2013). The modified cyanobactins are in the bacteriocin class I of modified bacteriocins including N-to-C-cyclic cyanobactins with heterocyclized or prenylated amino acids and the unmodified anacyclamides are in the group IIc of N-to-C-cyclic unmodified bacteriocins (Cotter et al. 2013). It is highly likely that new cyanobactins with novel post-translational modifications will be found in the future and the classification will become even more complicated.

1.8. Significance and biotechnological potential

The significance of natural product discovery in inspiring synthetic organic chemistry to develop methods for production of novel products cannot be underestimated and this has an important role in drug development (Dunbar and Mitchell 2013). Cyclic peptides are known for their stability and resistance against proteases and therefore are considered to be desired candidates in drug design (Wipf 1995, Driggers et al. 2008, Cascales and Craik 2010, Velásques and van der Donk 2011, Thorstholm and Craik 2012). The biotechnological potential of the cyanobactin pathway lies in the macrocyclases and the precursor peptide guided biosynthesis (Agarwal et al. 2012, Koehnke et al. 2012, Tianero et al. 2012). The core sequence of the precursor peptide directly encodes the resulting cyanobactin and this sequence can be changed easily by genetic engineering in heterologous hosts and recombinant peptides can be produced (Donia et al. 2006, Tianero et al. 2012). The enzymes in the cyanobactin pathway have very flexible substrate specificity for different amino acid sequences of the precursor peptide (Donia et al. 2006, Schmidt and Donia 2010, Tianero et al. 2012). The trunkamide pathway has been utilized to produce the artificial compound eptidemnamide by replacing the precursor gene with an artificial precursor gene (Donia et al. 2006). Each gene of the trunkamide gene cluster (*patABCDFG*) was cloned

independently to an expression vector in *E. coli* and co-expressed with the plasmid carrying the artificial precursor gene. The gene cluster, and the eptidemnamide encoded in the artificial precursor gene was expressed. The study proved the flexibility and utility of the pathways to produce designed peptides (Donia et al. 2006). The enzymes of natural product pathways can be utilized in a biotechnological approach to create novel compounds (Houssen and Jaspars 2010). The crystal structures of the proteases have been examined and the results enable studies to further our understanding and engineering systems that can efficiently produce cyclic peptides from linear precursor peptides (Agarwal et al. 2012, Koehnke et al. 2012). Bacterial expression systems have been developed but further research is required to expand the catalytic potential and improve the production efficiency (Tianero et al. 2012).

2. Aims of the study

The aim of this study was to explore the distribution and abundance of the cyanobactin pathway in cyanobacteria and to characterize the products of novel cyanobactin pathways based on genome mining in order to extend the known chemical diversity of cyanobactins.

Specific aims of the present study were:

- To study the frequency of the cyanobactin pathway and evolution of the protease gene (I)
- To discover the product of an unknown cyanobactin pathway in *Anabaena* sp. 90 (II)
- To discover the product of an inactive cyanobactin pathway in *M. aeruginosa* NIES 843 (III)
- To study the frequency, distribution, and structure of cyanobactin pathways in cyanobacteria (IV)
- To describe novel biosynthetic capacity and evolution of the cyanobactin pathway (IV)

3. Summary of materials and methods

The materials and methods used in this study are listed in Table 4. Detailed descriptions are in the original papers, I-IV. A schematic workflow of this study is shown in Figure 6.

Table 4. The methods used in this study. Roman numerals refer to the article in which the method was used.

Method	Article
Cultivation of cyanobacterial strains	I-IV
-with stable isotopes ^{14}N and ^{34}S	II-IV
DNA extraction	I-III
Primer design	I-III
PCR amplification	I-III
Cloning	II & III
Analysing the transformants	II & III
Phylogenetic analyses	I, II & IV
Purification of aeruginosamides and NMR	IV
Sample preparation for LCMS	II-IV
Synthetic reference compounds	II
LCMS	II-IV
NMR	IV
High accuracy UPLC-ESI-QTOF mass spectrometry	IV
Bioinformatic analyses of gene clusters	I-IV

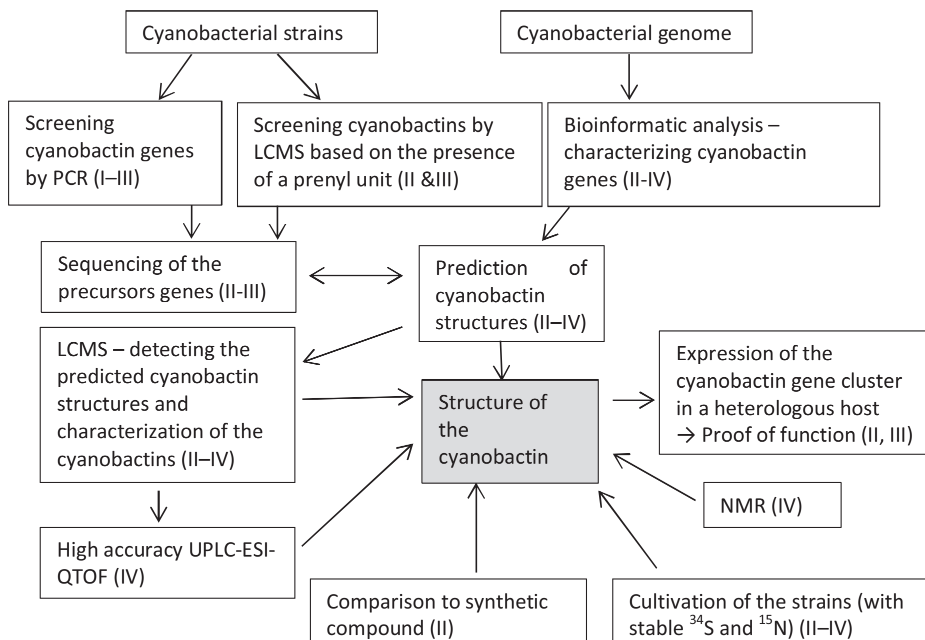


Figure 6. Schematic drawing of the workflow in studies I-IV.

4. Summary of results and discussion

4.1. Distribution of cyanobactin biosynthetic genes (I)

A novel method to identify cyanobactin pathways was developed in this study (I). Usually, cyanobactins are found through genome mining or in the screening of potential bioactive compounds (Velásques and van der Donk 2011, Donia and Schmidt 2010). In this study (I), distribution of the cyanobactin pathway was studied among planktonic bloom-forming cyanobacteria. A primer pair was designed to amplify the gene encoding PatA protease, which cleaves the precursor peptide at the N-terminal recognition site (Lee et al. 2009). The primers amplified a 1.4 kb section of the *patA* gene. The gene encoding the PatA protease was found in 48 of the 132 cyanobacterial strains studied. These cyanobacteria were taxonomically disparate and included unicellular, filamentous, and heterocyst-differentiating cyanobacteria. The *patA* gene was especially common in planktonic cyanobacteria. The results suggested that the cyanobactin pathway is widespread and sporadically distributed in cyanobacteria and hinted at the potential chemical diversity of cyanobactins encoded in this pathway. Prior to this study, just a few cyanobactin biosynthetic pathways had been described and most of those were found in the symbiotic *Prochloron* and marine cyanobacteria (Schmidt et al. 2005, Donia et al. 2006, Sudek et al. 2006, Donia et al. 2008, Ziemert et al. 2008b). To further explore the evolution of the PatA proteins, phylogenetic trees were constructed from the PatA protease and 16S rRNA sequences from 25 cyanobacterial strains. The trees were incongruent and the topologies were supported by high bootstrap-values. This observation suggests that the cyanobactin gene cluster has been transferred from one cyanobacterium to another and the incongruence is a result of horizontal gene transfer, which is supported by the high sequence diversity in the protease genes.

4.2. Anacyclamides – novel cyanobactins in the strains of *Anabaena* (II)

Bioinformatic analyses of the draft genome of *Anabaena* sp. 90 identified a cyanobactin gene cluster encoding an unknown cyanobactin. The gene cluster was annotated, however the automatic annotation missed the precursor gene that was found by manually searching. The gene cluster was different from those described previously for cyanobactins (Schmidt et al. 2005, Sudek et al. 2006, Ziemert et al. 2008b). Six genes were encoded similar to those found in other cyanobactin gene clusters (Figure 7); however, a heterocyclase responsible for the heterocyclization of serine, threonine and cysteine, was absent. The oxidase domain of the AcyG protease, responsible for the oxidation of thiazolines and oxazolines, was also absent. Based on these observations, it was obvious that there would be no heterocycles in these cyanobactins. In addition to this, the N-terminal cleavage site of the

precursor peptide was not conserved and the product could not be directly predicted.

Cysteines are common in cyanobactins (Schmidt et al. 2005, Sudek et al. 2006, Ziemert et al. 2008b). Therefore, in order to identify the product of this pathway, the precursor genes of other *Anabaena* strains were sequenced. The aim here was to identify a strain that encoded cysteine in the core region of the precursor peptide and discover the product of the pathway in this strain using a sulphur stable isotope labelling strategy. Primers were designed to amplify precursor genes in related *Anabaena* strains by PCR. A total of 26 precursor genes were amplified and sequenced, however, none of these precursor genes encoded cysteine in the core region. *Anabaena* sp. SYKE844B encoded a precursor peptide with methionine in the putative core region. This strain was cultivated on Z8 culture medium with $Mg^{34}SO_4$ as the sole sulphur source. A methanol extract of the cells was analysed by LCMS and compared to the cell extract, which was cultivated with $Mg^{32}SO_4$. By comparison of sum mass spectra of all of the ions from the chromatograms of labelled and unlabelled cell extracts, a peak was found with a mass shift of two units, which results from the mass difference between ^{32}S and ^{34}S atoms. By the above method the cyanobactin in *Anabaena* sp. SYKE844B was found to be a cyclic heptapeptide, Lys-Ile-Gly-Ile-Met-His-Pro, containing only proteinogenic amino acids and was later named anacyclamide A7. The structure analysis was based on the product ion spectrum and was confirmed by comparison to synthetic anacyclamide A7. Based on this finding, the cleavage sites were identified and applied to other anacyclamide precursors, and anacyclamides were detected in other 26 other *Anabaena* strains, including anacyclamide A10 from *Anabaena* sp. 90 (Figure 7). Altogether 18 anacyclamide variants were detected in 27 strains of *Anabaena*.

The anacyclamides in these *Anabaena* strains were cyclic peptides from 7 to 20 amino acids, containing only proteinogenic amino acids (Table 5). Some anacyclamides were prenylated, double prenylated or geranylated. The prenyl in a heteroatom is seen in the LCMS analysis as the neutral loss of 68 mass units, and geranyl as the loss of 136 mass units. The diversity of different anacyclamides resulting from the sequence variation in the precursor peptide was tremendous and just one amino acid, the proline at the C-terminus, was conserved. Characterization of anacyclamides was based primarily on the amino acid sequence of the core peptide, product ion analysis, and mass spectrometric and chromatographic comparison to synthetic anacyclamides. To confirm that the putative cyanobactin cluster in *Anabaena* sp. 90 is responsible for anacyclamide production, the *acy* gene cluster was amplified and cloned in *Escherichia coli* TOP10. The production of anacyclamide was detected by LCMS and this demonstrated that the *acy* genes are responsible for the anacyclamide production.

The discovery of anacyclamides here broadened the definition of cyanobactins to include unmodified cyclic peptides, which can have prenylated or geranylated amino acids. Other cyclic cyanobacterial peptides with only proteinogenic amino acids are nonapeptide oscillacyclin in *O. agardhii* (Fuji et al. 2000), octapeptides microphycin in *M. aeruginosa*

(Gesner-Apter and Carmeli 2008), planktocyclin in *P. agardhii* (Baumann et al. 2007), and dodecapeptide kawaguchipeptin B (Ishida et al. 1997). The discovery of anacyclamide pathway indicated that these peptides could be considered as potential products of the cyanobactin pathway. Based on the results in this study prenylagaramides (Murakami et al. 1999) in *P. agardhii* NIES 596 were also predicted to be products of this pathway. Subsequently, the prenylagaramides were shown to be made by the cyanobactin biosynthetic pathway (Donia and Schmidt 2011).

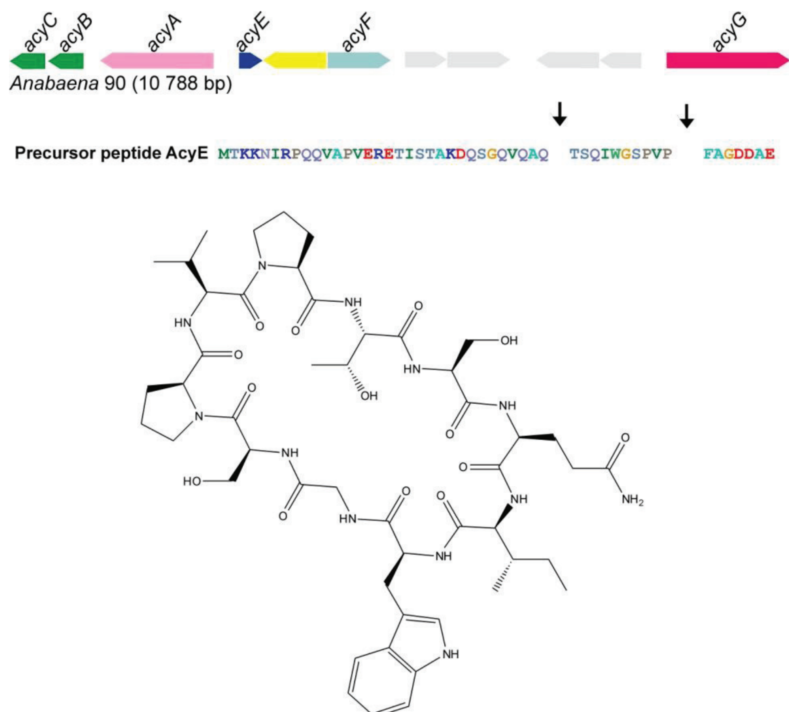


Figure 7. The biosynthetic pathway of anacyclamide in *Anabaena* sp. 90 and structure of anacyclamide A10. The genes in grey are ORFs encoding hypothetical proteins and the gene in yellow encodes a transposase. The cleavage sites in the AcyE precursor peptide are indicated with arrows.

Table 5. Cyanobactins discovered in this study (II-IV). The modified amino acid is indicated with one letter abbreviations of the amino acids if known in the modification column.

Cyanobactin	Core sequence	Amino acids	Modifications	Strain	Article
Anacyclamide A10	TSQIWGSPVP	10		*Anabaena* e.g. 90	II
Anacyclamide B10	SSVIWGSPVP	10		*Anabaena* e.g. 1TU31S9	
Anacyclamide C10	SAQWQNFGVP	10		*Anabaena* e. g. 299A	
Anacyclamide D10	NAHWQNFGVP	10		*Anabaena* 1TU44S9	
Anacyclamide E10	YAPLQNFGVP	10		*Anabaena* PH256	
Anacyclamide E10P	YAPLQNFGVP	10	*Y-prenyl*	*Anabaena* PH256	
Anacyclamide A11	DNWLGEWIGIP	11		*Anabaena* e.g. BIR260	
Anacyclamide A15	HAFIGYDQDPTGKYP	15		*Anabaena* 37	
Anacyclamide A15G	HAFIGYDQDPTGKYP	15	*Geranyl*	*Anabaena* 37	
Anacyclamide A7	RERFVYP	7		*Anabaena* 1TU39S8	
Anacyclamide F10P	YSNKPSDFSP	10	*Prenyl*	*Anabaena* 1TU32S11	
Anacyclamide A9P	YDDKLNLSP	9	*Y-Prenyl*	*Anabaena* 1TU33S10	
Anacyclamide A20P	WGNGTGLDWKLLTGGISASP	20	*Prenyl*	*Anabaena* 202A1	
Anacyclamide A20PP	WGNGTGLDWKLLTGGISASP	20	*2 Prenyls*	*Anabaena* 202A1	
Anacyclamide A8P	HQPWHAAP	8	*W-Prenyl*	*Anabaena* TR232	
Anacyclamide B8	FSPDWRAP	8		*Anabaena* SYKE816	
Anacyclamide C8	VIQHYLFP	8		*Anabaena* PH262	
Anacyclamide B7	LIGIMHP	7		*Anabaena* SYKE844B	
Piricyclamide	MSGVDYYNP NEFMQTGSYSGP TFCDLATKQCYP WILLADGTRPKNAP	9 12 12 14	*Geranyl* *Geranyl* *SS-bridge*	*Microcystis aeruginosa* PCC 7005	III
Piricyclamide	TLGCMNGTERCLGLP	15	*SS-bridge*	*Microcystis* Izancya 36 and 41	
Piricyclamide	FAIFLLLP ILGEGEGWNYNP	8 12	 *Prenyl*	*Microcystis* Izancya 42	
Piricyclamide	GTHLYTYTP	9	*Y-prenyl*	*Microcystis aeruginosa* SYKE864	
Piricyclamide	SQWGWRGLSDP	11		*Microcystis aeruginosa* SYKE764	
Aeruginosamide B Aeruginosamide C (LINEAR)	FFPC FFPVC	4 5	*Methylated C-terminus and prenylated N-terminus*	*Microcystis aeruginosa* PCC 9432	IV
Viridisamide A (LINEAR)	FIC	3	*Methylated C-terminus and prenylated N-terminus*	*Oscillatoria nigro-viridis* PCC 7112	

4.3. Piricyclamides – novel cyanobactins in the strains of *Microcystis* (III)

Bioinformatic examination of an inactive cyanobactin gene cluster in *M. aeruginosa* NIES 843 led to the discovery of an active functional gene cluster and a new group of cyanobactins, piricyclamides, in several strains of *M. aeruginosa*. A cyanobactin cluster was identified in the genome of *M. aeruginosa* NIES 843 (Donia and Schmidt 2011). The gene cluster was 19.7 kb, and it encoded many small hypothetical proteins and a short precursor gene. Re-examination of the cluster demonstrated that there were two more precursor genes in the inactive gene cluster, in addition to the described precursors. The gene encoding the PirG protease had a frameshift mutation, and the gene encoding the PirA protease had two insertion elements of 4.4 and 2.4 kb each. The *pir* cluster encoded all the essential cyanobactin genes except for the PatD heterocyclase and the oxidase domain of the PatG protease. This indicated that the active cluster would produce cyanobactins lacking heterocycles. The *pir* cluster encoded three precursor genes and the putative core regions of each could be recognised in an alignment of the precursor peptides together with known cyanobactin precursor peptides. Because this gene cluster was inactive, it was hypothesised that other strains of *M. aeruginosa* might encode an active counterpart.

A total of 74 *Microcystis* strains were thus screened by PCR and LCMS to find active producers of these types of cyanobactins. The LCMS screening was based on detecting prenyl or geranyl units at heteroatoms. This is seen as a neutral loss of the prenyl unit of 68 Da (geranyl unit of 136 Da) with the conditions used in the ion source. The neutral losses of the prenyl of geranyl units can be collected from all the recorded MS and MS^2 spectra resulting from one sample. In this way the molecular ions which carry a loosely attached prenyl or geranyl units can be located in the LCMS data. Cyanobactin candidates were detected in six strains by LCMS. In these strains, the precursor gene could also be amplified by PCR. The precursor genes were cloned and sequenced and putative cyanobactin predictions were made. Based on the predictions and LCMS analysis, including stable isotope labelling with ^{34}S and ^{14}N and derivatization experiments, novel cyanobactins were found and named as piricyclamides.

Piricyclamides contained only proteinogenic amino acids and some had prenylated or geranylated amino acids (Figure 8, Table 5). Three of the piricyclamide precursor peptide core sequences had a double cysteine pattern in the cyanobactin coding region. This pattern suggested that these piricyclamides could have a disulfide bridge. Labelling experiments with stable ^{34}S isotope indicated the presence of these two cysteines in three of the piricyclamide precursor peptides. A disulfide bridge was also identified and the structure verified by reduction and cysteine specific carboxyamidomethylation of the thiol groups (Neitz et al. 2011). In this method, the disulfide bridge is first reduced and the free thiols are carboxyamidomethylated. The disulphide bridge is seen as a shift of 116 Da of the molecular mass, whilst free reduced cysteine would shift 57 Da. The

molecular mass of piricyclamide with a disulphide bridge was seen after treatment in the LCMS analysis with 116 Da higher molecular mass, verifying the presence of a disulphide bridge.

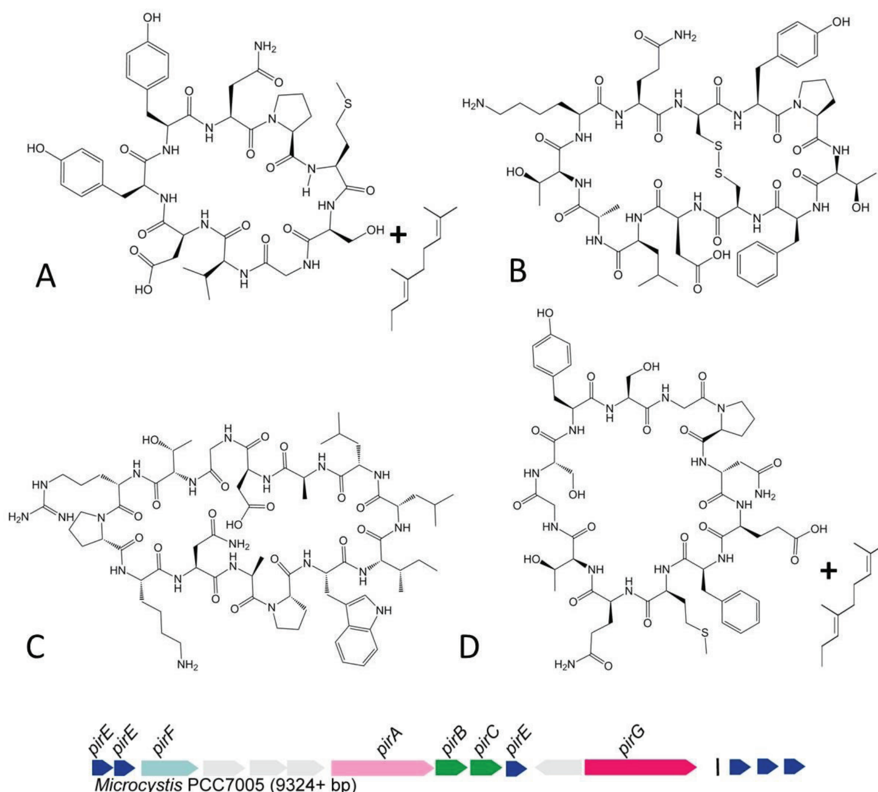


Figure 8. The piricyclamides gene cluster from *M. aeruginosa* PCC 7005. The detected and characterized piricyclamide structures are shown (A-D). Piricyclamides (A) and (D) have a geranyl unit in one of the amino acid side chain hetero atoms. The genes in grey are ORFs encoding hypothetical proteins and the precursor genes are separated from the cluster with a black bar showing that more precursor genes are encoded in the genome but could not be connected to the gene cluster, suggesting they are encoded at different locations in the genome of this strain.

In *M. aeruginosa* PCC 7005, four piricyclamides were found based on the amino acid sequences of the precursor peptides and LCMS analyses (Figure 8). Two of the piricyclamides were geranylated and one had a disulfide bridge (Figure 8, Table 5). The piricyclamide gene cluster was almost entirely amplified and sequenced from *M. aeruginosa* PCC 7005 (Figure 8). The primers were designed based on the interrupted gene cluster in *M. aeruginosa* NIES 843 and the precursor gene sequences obtained from *M. aeruginosa* PCC 7005. The gene cluster was cloned into *E. coli* to prove this cluster to be responsible for the production of piricyclamides. The gene

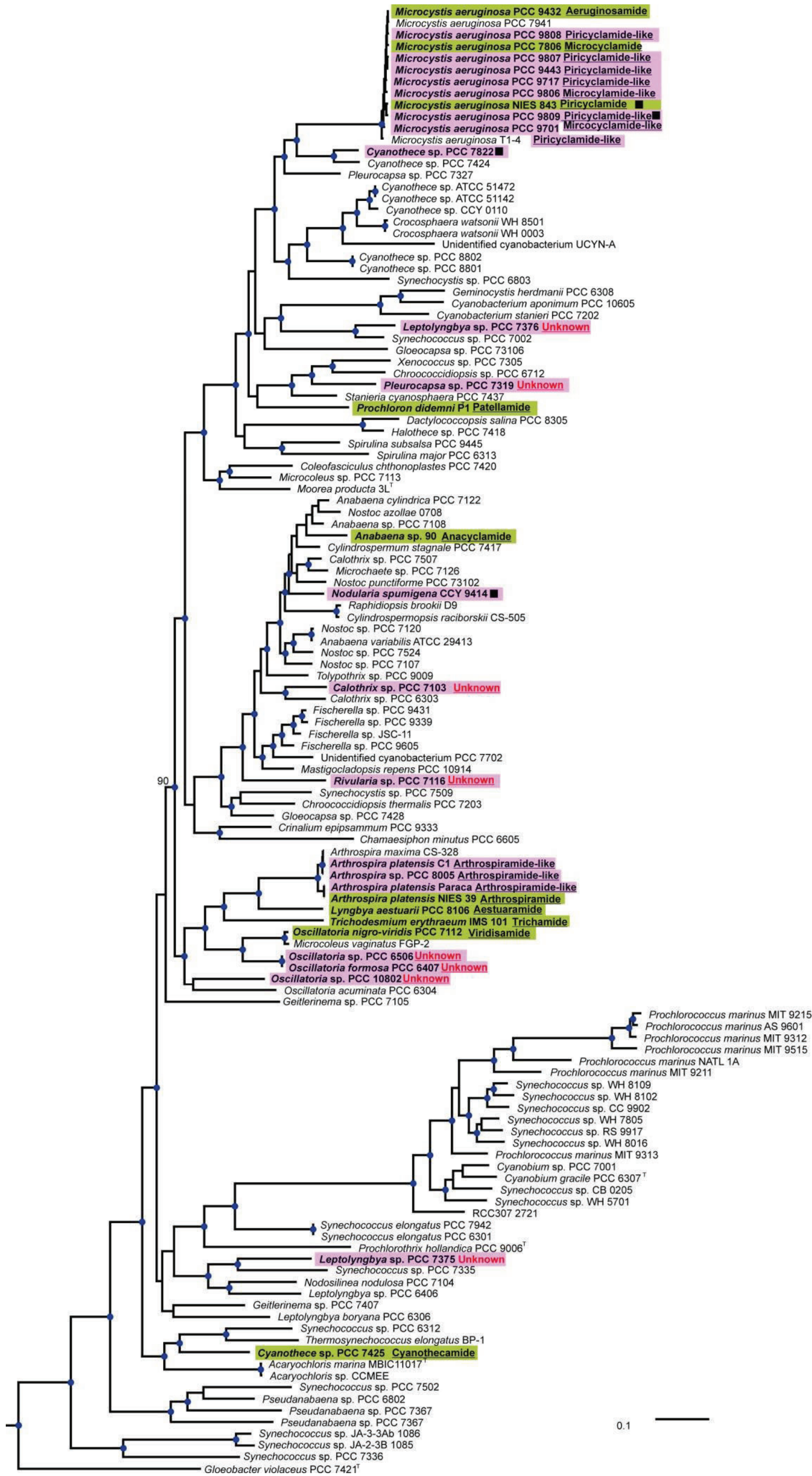
cluster was sequenced and shown to be 9.6 kb in length and much more compact in comparison to the inactive gene cluster. The *pir* cluster encoded a similar set of cyanobactin biosynthetic genes as the inactive gene cluster in *M. aeruginosa* NIES 843. The LCMS analysis of *E. coli* with the cloned piricyclamide gene cluster proved the functionality of the gene cluster but only one piricyclamide was detected. The cloned gene cluster lacked its own promoter and a few precursor genes. The only detected piricyclamide in *E. coli* was encoded in the precursor located in the middle of the *pir* gene cluster. The piricyclamide detected was TFCDLATKQCYP with a disulfide bridge. In addition, *M. aeruginosa* NIES 843, which carries an inactive cyanobactin pathway, was analysed with LCMS and no cyanobactins were detected.

In this study, the active counterpart for the inactive cyanobactin pathway in *M. aeruginosa* NIES 843 was described and it was shown to appear in several strains of *Microcystis*. Di-sulfide bridges and double cysteine patterns were seen in the piricyclamides precursor peptides, suggesting di-sulfide bridges. Among cyanobactins, the disulfide bridge is a rare post-translational modification as it appears in addition to piricyclamides only in ulithiacyclamides (Ireland and Scheuer 1980, Ireland et al. 1982, Williams et al. 1989, Fu et al. 1998). In other ribosomal peptides than cyanobactins, the disulfide bridge between two cysteines is more common and thought to stabilize the structure and make it more resistant to degradation (Arnison et al. 2013, McIntosh et al. 2009).

The diversity of piricyclamide precursors was examined in a natural bloom sample from Lake Tuusulanjärvi. A 16S rRNA clone library was constructed from the bloom sample and 97 % of the 16S rRNA genes sequenced could be attributed to the genus *Aphanizomenon*, while 3 % of the clones were assigned to the genus *Microcystis*. A *pirE* precursor gene library was constructed from the same sample and 122 clones were sequenced. The length of the core sequence varied from 7 to 22 amino acids and some showed the disulfide bridge pattern of cysteines, suggesting formation of a di-sulfide bridge. Extensive sequence variety was seen in the core region of the precursor peptides. Altogether 19 unique piricyclamide precursor peptides were found. Microviridins are another group of cyanobacterial ribosomal peptides. In comparison to the microviridin precursor gene library of 50 clones from a natural sample (Ziemert et al. 2010), the core sequence of microviridins had 7 invariable amino acids while in the piricyclamide amino acid sequence, the last proline is the only shared amino acid. This demonstrates the high sequence variation of piricyclamides and the flexibility of the pathway. The microviridin pathway has more restricted sequence diversity of the precursor peptides because the final structure is dependent on certain amino acids to obtain the cage-like structure (Ziemert et al. 2010). It seems that the piricyclamide pathway requires only the last amino acid of the core sequence to be a proline.

4.4. Genome mining of cyanobactin pathways (IV)

The frequency and distribution of the cyanobactin pathway was studied at the phylum level and on a genome scale. The presence of cyanobactin proteases PatA and PatG homologues in 126 cyanobacterial genome drafts was determined. The gene content of the cyanobactin gene clusters was analysed. In the set of 126 cyanobacterial genomes, 25 % encoded a cyanobactin pathway and 31 cyanobactin clusters were found (Figure 9) in 30 genomes. The cyanobactin pathway was sporadically distributed but seemed to be more common in late-branching cyanobacteria (Figure 9). The pathway was common in *Arthrospira*, *Oscillatoria*, and *Microcystis*, but not present in any of the *Prochlorococcus* or *Synechococcus* genomes (Figure 9). It has been speculated that the cyanobactin pathway is one of the major routes to small molecule diversity in cyanobacteria (Donia et al. 2008). The structural diversity of cyanobactins is constantly increasing (Arnison et al. 2013, Velásques and van der Donk 2011). However, the non-ribosomal and bacteriocin-type natural-product pathway appear to be more common than the cyanobactin pathway in cyanobacteria (Shih et al. 2013). Genome mining represents a starting point that shows the capacity of the genomes for production of ribosomal peptides.



Microcystis aeruginosa PCC 9432 Aeruginosamide
Microcystis aeruginosa PCC 7941
Microcystis aeruginosa PCC 9808 Piricyclamide-like
Microcystis aeruginosa PCC 7806 Microcyclamide
Microcystis aeruginosa PCC 9807 Piricyclamide-like
Microcystis aeruginosa PCC 9443 Piricyclamide-like
Microcystis aeruginosa PCC 9717 Piricyclamide-like
Microcystis aeruginosa PCC 9806 Microcylamide-like
Microcystis aeruginosa NIES 843 Piricyclamide
Microcystis aeruginosa PCC 9809 Piricyclamide-like
Microcystis aeruginosa PCC 9701 Mircocyclamide-like
Microcystis aeruginosa T1-4 Piricyclamide-like
Cyanothece sp. PCC 7822
Cyanothece sp. PCC 7424
Pleurocapsa sp. PCC 7327
Cyanothece sp. ATCC 51472
Cyanothece sp. ATCC 51142
Cyanothece sp. CCY 0110
Crocosphaera watsonii WH 8501
Crocosphaera watsonii WH 0003
Unidentified cyanobacterium UCYN-A
Cyanothece sp. PCC 8802
Cyanothece sp. PCC 8801
Synechocystis sp. PCC 6803
Geminocystis herdmanii PCC 6308
Cyanobacterium aponinum PCC 10605
Cyanobacterium stanieri PCC 7202
Leptolyngbya sp. PCC 7376 Unknown
Synechococcus sp. PCC 7002
Gloeocapsa sp. PCC 73106
Xenococcus sp. PCC 7305
Chroococcidiopsis sp. PCC 6712
Pleurocapsa sp. PCC 7319 Unknown
Stanieria cyanosphaera PCC 7437
Prochloron didemni P1 Patellamide
Dactylococcopsis salina PCC 8305
Halothece sp. PCC 7418
Spirulina subsalsa PCC 9445
Spirulina major PCC 6313
Coleofasciculus chthonoplastes PCC 7420
Microcoleus sp. PCC 7113
Moorea producta 3L
Anabaena cylindrica PCC 7122
Nostoc azollae 0708
Anabaena sp. PCC 7108
Anabaena sp. 90 Anacyclamide
Cylindrospermum stagnale PCC 7417
Calothrix sp. PCC 7507
Microchaete sp. PCC 7126
Nostoc punctiforme PCC 73102
Nodularia spumigena CCY 9414
Raphidiopsis brookii D9
Cylindrospermopsis raciborskii CS-505
Nostoc sp. PCC 7120
Anabaena variabilis ATCC 29413
Nostoc sp. PCC 7524
Nostoc sp. PCC 7107
Tolypothrix sp. PCC 9009
Calothrix sp. PCC 7103 Unknown
Calothrix sp. PCC 6303
Fischerella sp. PCC 9431
Fischerella sp. PCC 9339
Fischerella sp. JSC-11
Fischerella sp. PCC 9605
Unidentified cyanobacterium PCC 7702
Mastigocladopsis repens PCC 10914
Rivularia sp. PCC 7116 Unknown
90
Synechocystis sp. PCC 7509
Chroococcidiopsis thermalis PCC 7203
Gloeocapsa sp. PCC 7428
Crinalium epipsammum PCC 9333
Chamaesiphon minutus PCC 6605
Arthrospira maxima CS-328
Arthrospira platensis C1 Arthrospiramide-like
Arthrospira sp. PCC 8005 Arthrospiramide-like
Arthrospira platensis Paraca Arthrospiramide-like
Arthrospira platensis NIES 39 Arthrospiramide
Lyngbya aestuarii PCC 8106 Aestuaramide
Trichodesmium erythraeum IMS 101 Trichamide
Oscillatoria nigro-viridis PCC 7112 Viridisamide
Microcoleus vaginatus FGP-2
Oscillatoria sp. PCC 6506 Unknown
Oscillatoria formosa PCC 6407 Unknown
Oscillatoria sp. PCC 10802 Unknown
Oscillatoria acuminata PCC 6304
Geitlerinema sp. PCC 7105
Prochlorococcus marinus MIT 9215
Prochlorococcus marinus AS 9601
Prochlorococcus marinus MIT 9312
Prochlorococcus marinus MIT 9515
Prochlorococcus marinus NATL 1A
Prochlorococcus marinus MIT 9211
Synechococcus sp. WH 8109
Synechococcus sp. WH 8102
Synechococcus sp. CC 9902
Synechococcus sp. WH 7805
Synechococcus sp. RS 9917
Synechococcus sp. WH 8016
Prochlorococcus marinus MIT 9313
Cyanobium sp. PCC 7001
Cyanobium gracile PCC 6307
Synechococcus sp. CB 0205
Synechococcus sp. WH 5701
RCC307 2721
Synechococcus elongatus PCC 7942
Synechococcus elongatus PCC 6301
Prochlorothrix hollandica PCC 9006
Leptolyngbya sp. PCC 7375 Unknown
Synechococcus sp. PCC 7335
Nodosilinea nodulosa PCC 7104
Leptolyngbya sp. PCC 6406
Geitlerinema sp. PCC 7407
Leptolyngbya boryana PCC 6306
Synechococcus sp. PCC 6312
Thermosynechococcus elongatus BP-1
Cyanothece sp. PCC 7425 Cyanothecamide
Acaryochloris marina MBIC11017
Acaryochloris sp. CCMEE
Synechococcus sp. PCC 7502
Pseudanabaena sp. PCC 6802
Pseudanabaena sp. PCC 7367
Pseudanabaena sp. PCC 7367
0.1
Synechococcus sp. JA-3-3Ab 1086
Synechococcus sp. JA-2-3B 1085
Synechococcus sp. PCC 7336
Gloeobacter violaceus PCC 7421

Figure 9. Species tree of cyanobacterial genomes. The tree shows the distribution of the cyanobactin gene clusters. The maximum likelihood tree was constructed from 29 concatenated, conserved phylogenetic marker proteins. Nodes supported with a bootstrap value of ≥70 % are indicated with a blue dot. The strains highlighted in green have a complete cyanobactin pathway and the product is described. In the strains highlighted in pink, the product of the cyanobactin pathway has not been described or it is unknown. The strains with interrupted pathway are indicated with a square.

The gene content of the 31 cyanobactin pathways was typically *patA-patG* (Figure 10). The gene for the PatD heterocyclase and the oxidase domain of the PatG protease was absent from the clusters which encoded cyanobactins without heterocycles. The main difference between the gene clusters was the number of the short hypothetical proteins. Some clusters were very compact and others had many short hypothetical proteins making the clusters larger. Interrupted genes were found in the *Nodularia spumigena* CCY9414 cyanobactin cluster, which does not encode a precursor gene and has a truncated *patA* gene, suggesting it to be inactive (Voβ et al. 2013). In *Cyanothece* sp. PCC 7822, precursor genes could not be found. The cyanobactin gene cluster in *Calothrix* sp. PCC 7103 does not encode *patB* or *patC*, and *Leptolyngbya* sp. PCC 7376 does not encode *patC* or *patD*, even though there is an oxidase domain in *patG*. The cyanobactin gene cluster of *Leptolyngbya* sp. PCC 7375 did not encode a PatG homolog.

Some of the clusters were located on separate contigs. The number of precursor genes varied from 1–10 (Figure 10) in all cyanobactin gene clusters. Typically the last amino acid of the core region was cysteine or proline which was seen in the alignment of precursor peptides, except for the precursors which did not have the conserved C-terminal cleavage site and unclear last amino acids. The N-terminal leader sequences were conserved and could be divided into two groups containing either the LAELSEE or APVQR motif. Usually the LAELSEE-type precursor peptides had cysteine as the last amino acid of the core and proline in the APVQR-type. The pathways with the LAELSEE-type precursors generally showed the genetic capacity to produce heterocycles (Houssen et al. 2010). The leader sequence is thought to be important for the modifying enzymes (Houssen et al. 2010) and the observed division of the precursor leader sequences supports this finding for the presence of heterocyclase. An alignment of precursor peptides made it possible to directly predict the end-product of many of the cyanobactin pathways.

The precursor peptides which did not contain conserved cleavage sites, or where the cleavage sites were absent, made bioinformatic prediction of the product impossible. *Calothrix* sp. PCC 7103, *Leptolyngbya* sp. PCC 7376 and PCC 7375 gene clusters encoded novel types of precursor peptides, and the cyanobactins were impossible to predict. The putative core sequences in the precursor peptides of *Calothrix* sp. PCC 7103 had many cysteines, but the N-terminal cleavage site was unclear. The putative precursor peptide in *Leptolyngbya* sp. PCC 7376 and PCC 7375 showed no similarity to other precursor peptides at the C-terminal cleavage site. The cyanobactins that are unknown at the moment are interesting because they might have novel

structures and properties. Few of the clusters encoded proteins novel to the cyanobactin pathway. Methyltransferases were encoded in *Oscillatoria* sp. PCC 10802, *Cyanothece* sp. PCC 7425, *Cyanothece* sp. PCC 7822, *Microcystis aeruginosa* PCC 9432, *Oscillatoria nigro-viridis* PCC 7112 and *Leptolyngbya* sp. PCC 7376. Occasionally the methyltransferase domain was fused to other cyanobactin genes encoding PatA, PatG, or PatF homologs (Figure 10).

Often the studies on cyanobacterial genomes are not focused on the biosynthetic pathways of natural products but some genome articles report them (e. g. Wang et al. 2012, Voß et al. 2013, Frangeul et al. 2008, Kaneko et al. 2007). In this study, 126 cyanobacterial genomes in the NCBI database were included in the bioinformatic analysis and a general view of the occurrence of cyanobactin pathways and the genetic potential for cyanobactin production was characterized. This allows further studies to focus on the detection and characterization of potential natural products encoded in the genomes. One strategy to find the product of the pathway is to predict the structure of the compound based on the biosynthetic genes and precursor peptide, and to screen the predicted structures. This strategy has been successfully used in addition to this study in for example characterizing the product of the trichamide (Sudek et al. 2006), aesturamide (Donia et al. 2006, McIntosh et al. 2013), cyanothecamide (Donia and Schmidt 2011, Houssen et al. 2012) and arthrospiramide pathways (Donia and Schmidt 2011). Compared to the total of cyanobactin-like structures characterized from cyanobacteria (Tables 1–3), the number of the described biosynthetic gene clusters is very small. In this study, novel cyanobactin gene clusters have been characterized. The diversity of the post-translational modifications in cyanobactins is restricted compared to all ribosomally synthesised post-translationally modified peptides (Arnison et al. 2013), suggesting that more post-translational modifications await discovery in the cyanobactin family.

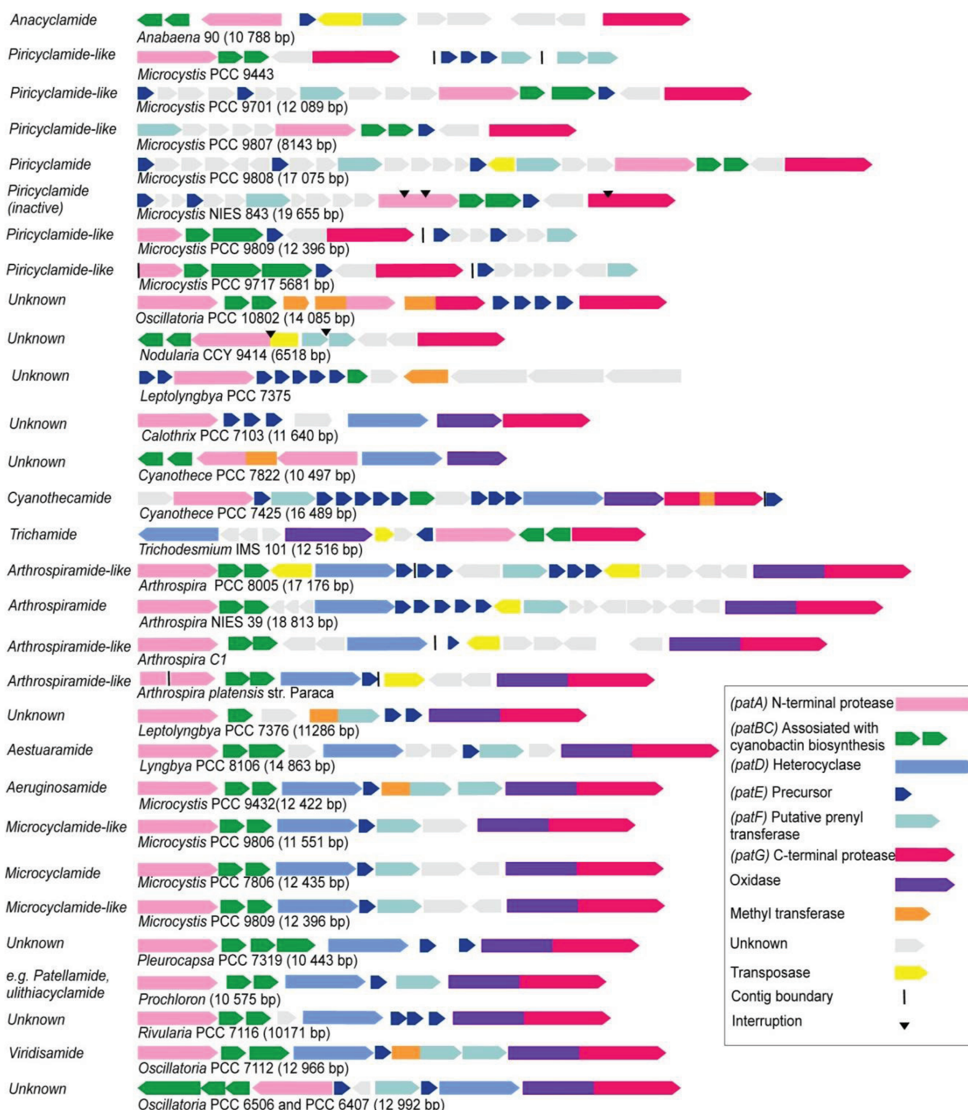


Figure 10. The 31 cyanobactin pathways found in the genomes of 126 cyanobacterial genomes. The end-product of the cyanobactin pathway is indicated where known.

4.5. The discovery of linear cyanobactins (IV)

An interesting gene encoding a bimodular protein with homology to methyltransferases and prenyltranferases (PatF) was found during the bioinformatic examination of the cyanobactin gene clusters from *Microcystis aeruginosa* PCC 9432, *Oscillatoria nigro-viridis* PCC 7112, and *Leptolyngbya* sp. PCC 7376 suggesting new post-translational modifications. The cyanobactins in *Leptolyngbya* sp. PCC 7376 could not be predicted

based on the precursor peptide because there was no conservation of the cleavage sites in the C-terminus. However, the precursor peptides of *M. aeruginosa* PCC 9432 and *O. nigro-viridis* PCC 7112 had conserved cleavage sites and the analysis suggested that the products of the cyanobactin pathway in these strains might be shorter than in the known cyanobactins. By aligning the *M. aeruginosa* PCC 9432 and *O. nigro-viridis* PCC 7112 precursor peptides with the similar known multicore precursors of microcyclamide (Ziemert et al. 2008b), patellamide, trunkamide, tenuecyclamide, and aestuaramide (Donia et al. 2008, McIntosh et al. 2013), these novel cyanobactins were predicted to be 3–5 amino acids. Predicted peptides in *M. aeruginosa* PCC 9432 were a tetrapeptide Phe-Phe-Pro-Cys and a pentapeptide Phe-Phe-Pro-Val-Cys. In *O. nigro-viridis* PCC 7112 the predicted cyanobactin was a tripeptide Phe-Ile-Cys. Candidates corresponding to the predictions were found by LCMS. Unexpectedly, the cyanobactins seemed to be linear peptides with a prenylated N-terminus and methylated C-terminus. Stable isotope labelling experiments with ^{15}N and ^{34}S, and high accuracy mass spectrometry supported this finding. Final confirmation of the pentapeptide Phe-Phe-Pro-Val-Cys structure from *M. aeruginosa* PCC 9432 was carried out with NMR. This conclusively demonstrated that the product of the cyanobactin pathway in this strain was a linear cyanobactin.

The structure was very similar to a cytotoxic compound, aeruginosamide, found in a bloom of *M. aeruginosa* (Lawton et al. 1999). Based on this similarity, the linear cyanobactins produced by *M. aeruginosa* PCC 9432 were named aeruginosamides B and C. The original aeruginosamide (Lawton et al. 1999) was considered to be aeruginosamide A. Instead of one prenyl, aeruginosamide A contains two prenyls on the N-terminus and the core sequence was slightly different (Figure 11B-D). In all the aeruginosamides the last amino acid is a heterocyclized cysteine. The tripeptide in *O. nigro-viridis* PCC 7112 was named viridisamide A (Figure 11A). The structures of the aeruginosamide B and viridisamide A were characterized based on LCMS product ion analyses and high accuracy UPLC-ESI-QTOF mass spectrometry. The bioinformatic examinations of cyanobactin pathways led to the discovery of linear cyanobactins, thus expanding the catalytic potential of the cyanobactin pathway.

When G-protease sequences were aligned, the comparison exposed the catalytic triad of the macrocyclase domain Asp-His- Ser (Agarwal et al. 2012, Koehnke et al. 2012). The aspartic acid is replaced by glutamic acid in G-protease of the linear cyanobactin producing strains in *M. aeruginosa* PCC 9432 and *O. nigro-viridis* PCC 7112, but histidine and serine are present. It can be speculated that the ability to macrocyclize the core peptide may be lost from these linear cyanobactin pathways. The G-protease is still needed, however, to act as a protease and the protease activity apparently remains since the precursor peptide requires cleavage at the C-terminus. Alternatively, the core sequences may be too short for the macrocyclase, or mutations in the macrocyclase have only resulted in the loss of the ability to macrocyclize.

The N-to-C cyclic structure of peptides stabilises and protects the structure from degradation (Wipf 1995, Craik and Cascales 2010, Montalbán-López et al. 2012). Most of the cyanobactins are cyclic but linear cyanobactins were discovered here (Study IV). The structure suggests that the function of the prenyl group is to protect the N-terminus and the C-terminus is protected by methylation, as end protection of free termini is not needed in N-to-C cyclic cyanobactins. Some highly modified linear peptides have been described from cyanobacteria (Van Wagoner et al. 2007). Modified peptides are usually thought to be nonribosomal but perhaps some of them are actually made by a ribosomal pathway. In addition to aeruginosamides, the linear peptides protected this way are rare, but there are those such as muscoride in *Nostoc muscorum* (Nagatsu et al. 1995) with the N-terminus prenylated and the C-terminus *O*-prenylated, and virenamide in *Diplosoma virens* (Carrol et al. 1996) with the N-terminus prenylated and a decarboxylated C-terminus. The discovery of linear cyanobactins suggests the pathway of these types of peptides is ribosomal.

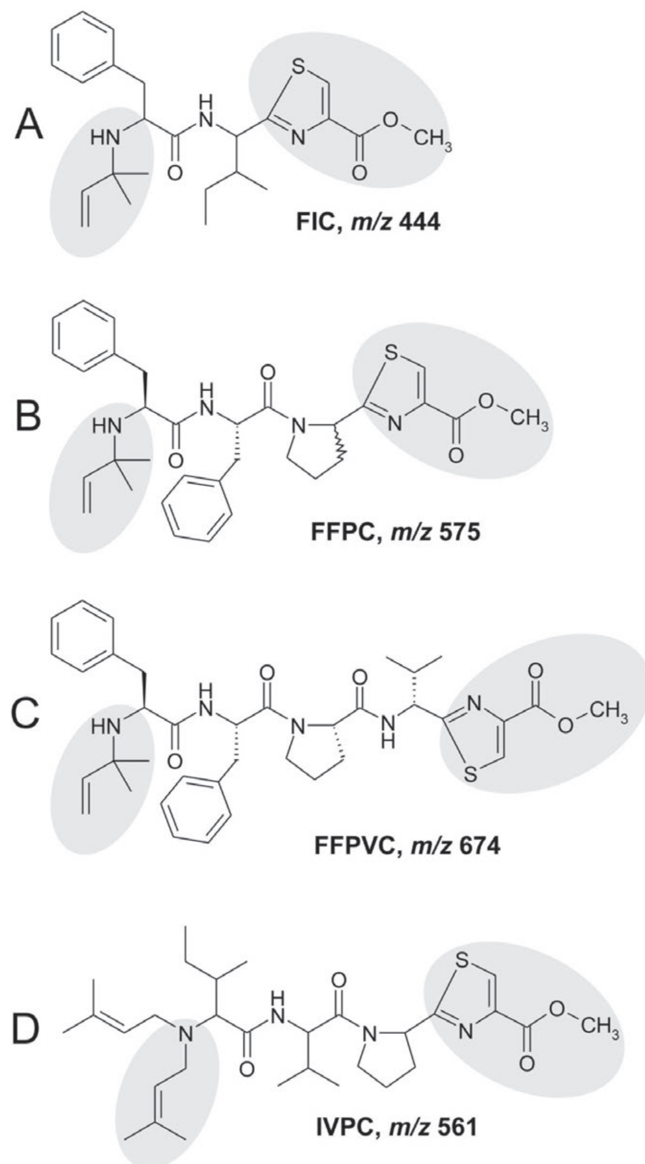


Figure 11. Linear cyanobactins. (A) Viridisamide A, (B) aeruginosamide B (C), aeruginosamide C, and (D) aeruginosamide A (Lawton et al. 1999). The amino acid sequence of the cyanobactin is shown with the mass to charge ratio of the protonated molecule. The N-prenylation and identical C-terminal methyl 1,3-thiazole-4-carboxylate structure are highlighted. Aeruginosamides C and D are reverse prenylated and the aeruginosamide A has a forward prenyl. The conformation of the prenyl in viridisamide is unclear.

4.6. Cyanobactins in the strains of *M. aeruginosa* (III and IV)

Microcystis is an ecologically and environmentally significant, toxic, and bloom-forming cyanobacterium (Kaneko et al. 2007, Frangeul et al. 2008). In addition to toxins and a range of other bioactive compounds, strains of *M. aeruginosa* can produce three different types of cyanobactins, each of which exhibits pronounced chemical variation. The microcyclamide pathway was first described in *M. aeruginosa* NIES 298 and PCC 7806 (Ziemert et al. 2008b). Microcyclamides are cyclic hexapeptides and they contain heterocyclized amino acids (Ishida et al. 2000, Ziemert et al. 2008b). A cryptic cyanobactin cluster interrupted by multiple hypothetical proteins and transposases was found in the genome of *M. aeruginosa* NIES 843 (Donia and Schmidt 2011). By re-analysing the cluster from this strain, this study showed it to be inactive due to insertions and interruption (III).

Piricyclamides are slightly different from the known *Microcystis* cyanobactins microcyclamides (Ziemert et al 2008). In piricyclamides, the amino acid content varied from 7–17 amino acids: occasionally piricyclamides were prenylated, geranylated, or had di-sulfide bridges. Phylogenetic analyses of concatenated A- and G-proteases were carried out to study how these two cyanobactin pathways in *M. aeruginosa* strains have evolved (III). The analysis showed the divergence of microcyclamide and piricyclamide pathways. The piricyclamide pathway grouped with trichamide and anacyclamide pathways, and the microcyclamide pathway was grouped with the other multicore precursor pathways, such as the patellamide (Schmidt et al. 2005), aestuaramide (Donia et al. 2008, McIntosh et al. 2013), and tenuecyclamide pathway (Donia et al. 2008).

In article IV, a third type of cyanobactin in *M. aeruginosa* strains was reported. The *M. aeruginosa* PCC 9432 was shown to produce linear cyanobactins of four and five amino acids. To date, three types of cyanobactin pathways are described in the strains of *Microcystis*, two of which are described in this study (II and IV). When the protease sequences from the aeruginosamide pathway were taken into account in the phylogenetic analysis of concatenated proteases, the aeruginosamide proteases are grouped with the microcyclamide proteases. This was already anticipated because the precursor sequence homology of the aeruginosamide precursor peptides to microcyclamide precursor peptides.

The cyanobactin diversity and frequency in the genus *Microcystis* is interesting. Nearly all strains of the genus *Microcystis* examined encoded cyanobactin genes. There are differences in the cyanobactin pathways in *Microcystis*. Piricyclamides are encoded in multiple precursor genes, while microcyclamides and aeruginosamides are encoded in single precursor genes with multiple core sequences. This suggests that the level of cyanobactin produced may be enhanced by an increase in the copy number of the core sequence. Another difference with the piricyclamide gene clusters from the other *Microcystis* cyanobactin clusters was that the piricyclamide gene clusters had long intergenic regions and encoded many hypothetical proteins.

The other cyanobactin gene clusters were compact with short intergenic regions between genes and lacking the hypothetical proteins. In addition to the aeruginosamide pathway, the *M. aeruginosa* PCC 9432 genome encoded piricyclamide precursor genes and a *pirF*, but no other piricyclamide genes were found. These precursor genes encoded similar piricyclamides detected in study III, but no piricyclamides were detected in this strain. This suggests that the pathway is inactive and piricyclamide genes in the genome of *M. aeruginosa* PCC 9432 are remnants of a functional gene cluster. In contrast *M. aeruginosa* PCC 9809 encoded two types of cyanobactin pathways; microcyclamide and piricyclamide pathway. However, the piricyclamide pathway was located in two separate contigs and the *patA* homolog was interrupted, suggesting it might be inactive. The piricyclamide pathways were, in many cases, fragmented on different contigs. The number of repetitive regions and plasticity is common in *Microcystis* genomes (Kaneko et al. 2007, Frangeul et al. 2008). This may lead to fragmented pathways on different contigs due to the difficult assembly of repeat rich regions.

Microcystis strains appear to be a prolific source of cyanobactins. In addition to the preceding cyanobactins, several cyanobactin-like compounds are described in *M. aeruginosa*, such as kawaguchipeptin (Ishida et al. 1996, 1997) and microphycin (Gesner-Apter and Carmeli 2008). Kawaguchipeptins are cyclic undecapeptides with proteinogenic amino acids and kawaguchipeptin A has a prenylated tryptophan fused to the amino acid backbone (Ishida et al. 1996, 1997). The structures suggest that these peptides are synthesised by the cyanobactin pathway. In addition, slightly different cyanobactin-like peptides have been characterized in *Microcystis*. Aeruginazoles consist of 12 to 17 amino acids with thiazoles (Raveh and Carmeli 2010, Adiv et al. 2012). Aeruginazoles can be chlorinated or di-chlorinated at tyrosine (Adiv et al. 2012). These peptides are thought to be made by the cyanobactin pathway (Raveh and Carmeli 2010, Adiv et al. 2012), but no halogenases have been described from the cyanobactin pathways to date.

4.7. Evolution of the cyanobactin pathway

The evolution of the cyanobactin proteins was examined using phylogenetic analyses in articles I and IV. The incongruence of 16S rRNA phylogeny with the phylogeny of part of PatA homologues suggested the cyanobactin gene clusters are exchanged between cyanobacteria by horizontal gene transfer (I). This same conclusion could be made based on the phylogenetic tree constructed from the sequences of the concatenated PatA and PatG homologs found in the bioinformatic study (IV) (Figure 12). The amino acid sequences of precursor peptides imply rapid evolution. The core sequences of precursor peptides vary considerably more than the rest of the precursor peptide. By alteration of a single nucleic acid in the core region, the cyanobactin products may gain new properties. The cyanobactin pathway seems to be a system developed for production of rapidly evolving peptides. According to the Screening Hypothesis (Firn and Jones 2000, 2003) the ability to create

chemical diversity is valuable to the producer. The flexibility of the enzymes in cyanobactin biosynthesis (Donia et al. 2006, Schmidt and Donia 2010, Tianero et al. 2012) supports the Screening Hypothesis.

The tree constructed from concatenated protease PatA and PatG sequences tree was strongly supported and it indicated that the pathways with multicore precursors formed a well-supported group, as the other pathways usually encoded multiple precursors with single core (Figure 12). This suggests that the single core precursor gene in multiple copies is an ancient feature and the multicore precursor gene in a single copy is a derived feature. Additionally, the pathways with the ability to heterocyclize amino acids were present throughout the tree, suggesting that this ability has been lost at least twice (Figure 12). This suggests that the inability to heterocyclize is derived feature and the ability to heterocyclize is an ancestral feature. The relationship of cyanobactins to other ribosomal peptides is unclear. Compared to the numerous ribosomal peptides (Arnison et al. 2013), cyanobactins have many similar genetic and structural features, but they have the distinct feature of requiring two proteases to process the precursor peptide. Questions remaining here include how this system evolved and what is the relationship other ribosomal peptides. Clearly the cyanobactin pathway is related to the pathways of other post-translationally modified peptides, but the missing link between these pathways is still to be found.

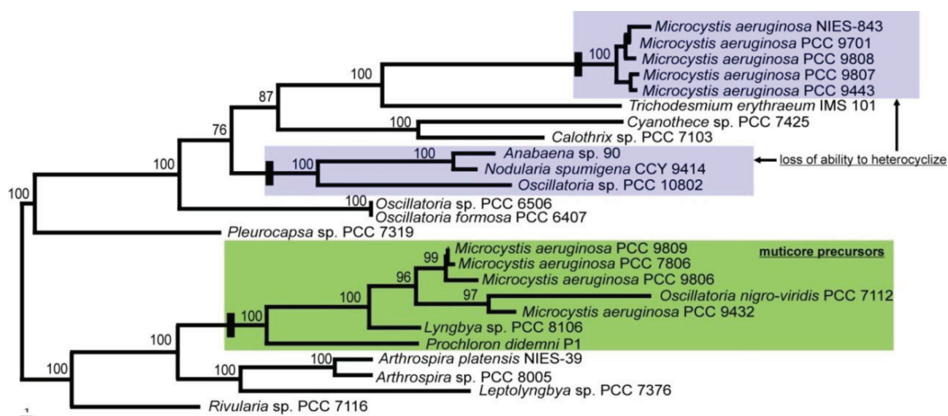


Figure 12. A maximum-likelihood tree illustrating the relationship between cyanobactin gene clusters. The tree was constructed from a concatenation of the cyanobactin protease amino acid sequences (PatA and PatG). The strains highlighted with blue show the cyanobactin pathways which lack the ability to heterocyclize amino acids. The strains indicated with green encode precursor genes with multiple core regions.

5. Conclusions

In this study, the frequency of the cyanobactin pathway has been described and shown to be widely but sporadically distributed in cyanobacteria. A PCR screening strategy was developed to find cyanobactin producing candidates. Horizontal gene transfer was demonstrated to be involved in the evolution of the cyanobactin pathway. This study has shown that in addition to *Prochloron*, the obligate symbiont of ascidians, the free-living cyanobacteria often have cyanobactin biosynthetic genes.

Novel cyanobactin pathways were described in the strains of *Anabaena* and *Microcystis*. The novel cyanobactins were named anacyclamides and piricyclamides. These cyanobactins were cyclic with solely proteinogenic amino acids, with occasionally some prenylated or geranylated amino acids, and rarely disulfide bridges occurring between two cysteines. These studies are the first and at the moment, only reports of the cyanobactin pathway producing unmodified cyanobactins, without heterocycles, consisting only of proteinogenic amino acids. These two families of cyanobactins have increased the length variation of cyanobactins. In these studies, a stable sulphur isotope labelling strategy was used and proven to be a useful method for locating cyanobactins and in LCMS data.

The result of a phylum-level bioinformatic analysis showed that 126 cyanobacterial genomes contained 31 cyanobactin pathways. Many of the pathways were new and contained novel genes and a new type of precursor peptides, suggesting the presence of novel cyanobactins. This information can be further utilized in the search of novel cyanobactins. This study focused on two pathways encoding a novel protein fusion of prenyltransferase and methyltransferase, encoded in the cyanobactin pathways of *Microcystis aeruginosa* PCC 9432 and *Oscillatoria* nigro-viridis PCC 7112. In addition, these pathways had interesting precursor peptides, which enabled prediction of cyanobactins consisting putatively of only 3-5 amino acids. Surprisingly, extensive chemical analysis showed that these cyanobactins were linear and were named aeruginosamides B, C, and viridisamide A. This work presents the novel finding that a cyanobactin pathway produces linear cyanobactins. These novel 3 to 5 amino acid linear cyanobactins have ends protected with N-prenyl and C-methyl. These types of linear peptides are rare in cyanobacteria. The naming of aeruginosamides was based on a highly similar known compound aeruginosamide described from *Microcystis* bloom (Lawton et al. 1999). The findings also suggest that the aeruginosamide is a cyanobactin and made ribosomally.

Many of the genomes examined were from *Microcystis* and this study has shown the capability of the strains of this genus to produce three types of cyanobactins; microcyclamides, piricyclamides, and aeruginosamides. Phylogenetic analyses showed that the piricyclamide and microcyclamide pathways are divergent and the aeruginosamide pathway is more closely related to the microcyclamide pathway than the piricyclamide pathway. The phylogenetic investigation of concatenated proteases from the genome study

showed that multicore precursor cyanobactin producers are clustered into one well-resolved clade. The other cyanobactin pathways generally have more than one precursor, each encoding one cyanobactin in single core. The ability to heterocyclize appears throughout the phylogenetic tree and it showed that the ability has been lost at least twice. The analysis suggested that the single core precursor and ability to heterocyclize would be ancient capabilities and derived features would include being a multicore precursor and loss of heterocyclase activity.

This study has redefined the genetic and chemical diversity of the cyanobactin family through genome mining and chemical analyses. Prior to this study, the distribution of cyanobactin pathways was not known and only a few pathways were described. In this study, the cyanobactin family is twice redefined: first to include cyanobactins consisting solely of unmodified amino acids, and second, linear cyanobactins. The results gained in this study have significantly influenced the cyanobactin field; demonstrating how novel natural products can be found with the combination of bioinformatics and chemical analyses. The rapidly expanding structural diversity of cyanobactins hints that novel post-translational modifications await discovery. The question still remaining is why cyanobactins are made. Future studies are needed to shed light on the ecology and purpose of these widespread cyanobacterial peptides.

6. Acknowledgements

This work has been carried out at Division of Microbiology and Biotechnology, Department of Food and Environmental Sciences, University of Helsinki. The study was funded by Viikki Doctoral Programme in Molecular Biosciences (VGSB) and The Academy of Finland (Center of Excellence: Interactive Photosynthesis and Bioactive Compound Research at Systems Biology level).

I wish to warmly thank my supervisors Kaarina Sivonen, David Fewer, and Jouni Jokela. I express the deepest gratitude for Kaarina that she took me in her group to do my Master's thesis; since then I have been privileged to work in the wonderful cyanogroup. Special thanks to David for the practical supervision and endless discussions and advice. I thank Jouni especially for the guidance with the chemical methods. I thank the members of my followup group: Natalia Battchikova and Mirja Salkinoja-Salonen for their ideas and advice.

I want to thank Professor Marcel Jaspars and Docent Jarmo Niemi for pre-examination of this thesis and giving suggestions and comments to improve it.

I want to express my gratitude to all the co-authors. I thank Lyudmila and Matti for the technical assistance. I wish to thank everyone in the cyanogroup, personnel in the division, and especially friends at the coffee table, which has made working here enjoyable and fun.

Very special thanks to family and friends for support and being there.

7. References

Admi V, Afek U, Carmeli S (1996) Raocyclamides A and B, novel cyclic hexapeptides isolated from the cyanobacterium *Oscillatoria raoi*. J Nat Prod 59: 396-399.

Adiv S, Ahronov-Nadborny, Carmeli S (2012) New aeruginazoles, a group of thiazole containing cyclic peptides from *Microcystis aeruginosa* blooms. Tetrahedron 68: 1367–1383.

Agarwal V, Pierce E, McIntosh J, Schmidt EW, Nair SK (2012) Structures of cyanobactin maturation enzymes define a family of transamidating proteases. Chem Biol 19: 1411-22.

Arnison PG, Bibb MJ, Bierbaum G, Bowers AA, Bugni TS Bulaj G, Camarero JA, Campopiano DJ, Challis GL, Clardy J, Cotter PD, Craik DJ, Dawson M, Dittmann E, Donadio S, Dorrestein PC, Entian KD, Fischbach MA, Garavelli JS, Göransson U, Gruber CW, Haft DH, Hemscheidt TK, Hertweck C, Hill C, Horswill AR, Jaspars M, Kelly WL, Klinman JP, Kuipers OP, Link AJ, Liu W, Marahiel MA, Mitchell DA, Moll GN, Moore BS, Müller R, Nair SK, Nes IF, Norris GE, Olivera BM, Onaka H, Patchett ML, Piel J, Reaney MJ, Rebuffat S, Ross RP, Sahl HG, Schmidt EW, Selsted ME, Severinov K, Shen B, Sivonen K, Smith L, Stein T, Süssmuth RD, Tagg JR, Tang GL, Truman AW, Vederas JC, Walsh CT, Walton JD, Wenzel SC, Willey JM, van der Donk WA (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep 30:108-60.

Arrault A, Witczak-Legrand A, Gonzalez P, Bontemps-Subielos N, Banaigs B (2002) Structure and total synthesis of cyclodidemnamide B, a cycloheptapeptide from the ascidian *Didemnum molle*. Tetrahedron Lett 43: 4041-4044.

Banker R, Carmeli S (1998) Tenuecyclamides A-D, cyclic hexapeptides from the cyanobacterium *Nostoc spongiaeforme* var. *tenue*. J Nat Prod 61: 1248-1251.

Baumann HI, Keller S, Wolter, FE, Nicholson GJ, Jung G, Susmuth RD, Jüttner F (2007) Planktocyclin, a cyclooctapeptide protease inhibitor produced by the freshwater cyanobacterium *Planktothrix rubescens*. J Nat Prod 70: 1611-1615.

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

Caba JM, Rodriguez IM, Manzanares I, Giralt E, Albericio F (2001) Solid-phase total synthesis of trunkamide A1. J Org Chem 66: 7568-7574.

Carrol AR, Feng Y, Bowden BF (1996) Studies of Australian ascidians. 5. Virenamides A–C, New cytotoxic peptides from the colonial didemnid ascidian *Diplosoma virens*. J Org Chem 61:4059–4061.

Cascales L, Craik DJ (2010) Naturally occurring circular proteins: distribution, biosynthesis and evolution. Org Biomol Chem 21: 5035–5047.

Castenholz RW (2001) Phylum BX. Cyanobacteria, oxygenic photosynthetic bacteria. p. 473–599. In: Boone DR, Castenholz RW (ed.) Bergey's Manual of Systematic Bacteriology Part 1, 2nd edition. SpringerVerlag, New York, USA.

Cotter PD, Ross, RP and Hill C (2013) Bacteriocins – a viable alternative to antibiotics. Nature Rev Microbiol 11: 95–105.

Degnan BM, Hawkins CJ, Lavin MF, McCaffrey EJ, Parry DL, van den Brenk AL, Watters DJ (1989a) New cyclic peptides with cytotoxic activity from the ascidian *Lissoclimum patella*. J Med Chem 32:1349–54.

Degnan BM, Hawkins CJ, Lavin MF, McCaffrey EJ, Parry DL, Watters DJ (1989b) Novel cytotoxic compounds from the ascidian *Lissoclinum bistratum*. J Med Chem 32:1354-1359.

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

Donia MS, Ruffner DE, Cao S, Schmidt EW (2011) Accessing the hidden majority of marine natural products through metagenomics. ChemBioChem 12:1230–1236.

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

Donia MS, Schmidt EW. (2010) Cyanobactins – Ubiquitous Cyanobacterial Ribosomal Peptide Metabolites. p. 539–558. In: Comprehensive Natural Products II Chemistry and Biology. Eds. Moore BS and Crews P.

Donia MS, Schmidt EW (2011) Linking chemistry and genetics in the growing cyanobactin natural products family. Chem Biol 18:508-19.

Donia MS, Wang B, Dunbar DC, Desai PV, Patny A, Avery M, Hamann MT (2008b) Mollamides B and C, Cyclic Hexapeptides from the Indonesian Tunicate *Didemnum molle*. J Nat Prod 27:941–945.

Driggers EM, Halel SP, Lee J, Terret NK (2008) The exploration of macrocycles for drug discovery — an underexploited structural class. Nat Rev Drug Disc 7:608–624.

Dunbar KL, Mitchell DA (2013) Revealing nature's synthetic potential through the study of ribosomal natural product biosynthesis. ACS Chem Biol 8:473–487.

Erickson KL, Gustafson KR, Milanowski DJ, Pannel LK, Klose JR, Boyd MR (2003) Myriastramides A–C, new modified cyclic peptides from the Philippines marine sponge *Myriastra clavosa*. Tetrahedron 59: 10231–10238.

Foster MP, Concepción GP, Caraan GB, Ireland CM (1992) Bistratamides C and D. Two new oxazole-containing cyclic hexapeptides isolated from a Philippine *Lissoclinum bistratum* ascidian. J Org Chem 57:6671-6675.

Foster MP, Ireland CM (1993) Nairaiamide A and nairaiamide B - 2 novel di-proline heptapeptides isolated from a Fijian *Lissoclinum-Bistratum* ascidian. Tetrahedron Lett 34:2871-2874.

Frangeul L, Quillardet P, Castets AM, Humbert J-F, Matthijs HCP, Cortez D, Tolonen A, Zhang C-C, Gribaldo S, Kehr J-C, Zilliges Y, Ziemert N, Becker S, Talla E, Latifi A, Billault A, Lepelletier A, Dittmann E, Bouchier C, Tandeu de Marsac N (2008) Highly plastic genome of *Microcystis aeruginosa* PCC 7806, a ubiquitous toxic freshwater cyanobacterium. BMC Genomics 9:274.

Firn RD, Jones CG (2000) The evolution of secondary metabolism - a unifying model. Mol Mic 37:989-994.

Firn RD, Jones CG (2003) Natural products - a simple model to explain chemical diversity. Nat Prod Rep 20:382-391.

Fu X, Do T, Schmitz FJ, Andrusevich V, Engel MH (1998) New cyclic peptides from the ascidian *Lissoclinum patella*. J Nat Prod 61:1547-1551.

Fujii K, Sivonen K, Naganawa E, Harada KI (2000) Non-toxic peptides from toxic cyanobacteria, *Oscillatoria agardhii*. Tetrahedron 56:725-733.

Gesner-Apter S, Carmeli S (2008) Three novel metabolites from a bloom of the cyanobacterium *Microcystis* sp. Tetrahedron 64:6628-6634.

Guan LL, Sera Y, Adachi K, Nishida F, Shizuri Y (2001) Isolation and evaluation of nonsiderophore cyclic peptides from marine sponges. Biochem Biophys Res Commun 283:976-981.

Haft DH, Basu MK, Mitchell DA (2010) Expansion of ribosomally produced natural products: a nitrile hydrates- and Nif11-related precursor family. BMC Biol 8:70.

Hambley TW, Hawkins CJ, Lavin MF, van den Brenk A, Watters DJ (1992) Cycloxazoline: a cytotoxic cyclic hexapeptide from the ascidian *Lissoclinum bistratum*. Tetrahedron 48:341-348.

Hawkins CJ, Lavin MF, Marshall KA, van den Brenk AL, Watters DJ (1990) Structure-activity relationships of the lissoclinamides: cytotoxic cyclic peptides from the ascidian *Lissoclinum patella*. J Med Chem 33:1634-1638.

Houssen WE, Jaspars M (2010) Azole-based cyclic peptides from the sea squirt *Lissoclinum patella:* Old scaffolds, new avenues. ChemBioChem 11:1803–1815.

Houssen WE, Koehnke J, Zollman D, Vendome J, Raab A, Smith MCM, Naismith JH, Jaspars M (2012) The Discovery of New Cyanobactins from *Cyanothece* PCC 7425 Defines a New Signature for Processing of Patellamides. ChemBioChem 13: 2683–2689.

Houssen WE, Wright SH, Kalverda AP, Thompson GS, Kelly SM, Jaspars M (2010) Solution structure of the leader sequence of the patellamide precursor protein, PatE1-34. ChemBioChem. 11:1867-73.

Ireland CM, Durso Jr AR, Newman RA, Hacker MP (1982) Antineoplastic cyclic peptides from the marine tunicate *Lissoclinum patella*. J Org Chem 47:1807-1811.

Ishida K, Matsuda H, Murakami M, Yamaguchi K (1996) Kawaguchipeptin A, a novel cyclic undekapeptide from cyanobacterium *Microcystis aeruginosa* (NIES-88). Tetrahedron 52: 9025-9030.

Ishida K, Matsuda H, Murakami M, Yamaguchi K (1997) Kawaguchipeptin B and antibacterial cyclic undekapeptide from the cyanobacterium *Microcystis aeruginosa*. J Nat Prod 60:724–726.

Ishida K, Nakagawa H, Murakami M (2000) Microcyclamide, a cytotoxic cyclic hexapeptide from the cyanobacterium *Microcystis aeruginosa*. J Nat Prod 63:1315-1317.

Ireland CM, Scheuer PJ (1980) Ulicyclamide and ulithiacyclamide, two new small peptides from a marine tunicate. J Am Chem Soc 102:5688-5691.

Ishida T, Tanaka M, Nabae M, Inoue M (1988) Solution and solid-state conformations of ascidiacyclamide, a cytotoxic cyclic peptide from ascidian. J Org Chem 53:107-112.

Jüttner F, Todorova AK, Walch N, von Philipsborn W (2001) Nostocyclamide M: a cyanobacterial cyclic peptide with allelopathic activity from *Nostoc* 31. Phytochemistry 57:613-619.

Kaneko T, Nakajima N, Okamoto S, Suzuki I, Tanabe Y, Tamaoki M, Nakamura Y, Kasai F, Watanabe A, Kawashima K, Kishida Y, Ono A, Shimizu Y, Takahashi C, Minami C, Fujishiro T, Kohara M, Katoh M, Nakazaki N, Nakayama S, Yamada M, Tabatai S, Watanabe MM (2007) Complete genomic structure of the bloom-forming toxic cyanobacterium *Microcystis aeruginosa* NIES-843. DNA Res 14:247–256.

Kehraus S, Konig GM, Wright AD, Woerheide G (2002) Leucamide A: A new cytotoxic heptapeptide from the Australian sponge *Leucetta microraphis*. J Org Chem 67:4989–4992.

Kobayashi J, Nakamura T, Tsuda M (1996) Hymenamide F, new cyclic heptapeptide from marine sponge *Hymeniacidon* sp. Tetrahedron 18:6355–6360.

Kobayashi J, Tsuda M, Nakamura T, Mikami Y, Shigemori H (1993) Hymenamide A and Hymenamide B, new proline-rich cyclic heptapeptides from the Okinawan marine sponge *Hymeniacidon* sp. Tetrahedron 49: 2391-2402.

Koehnke J, Bent A, Houssen WE, Zollman D, Morawitz F, Shirran S, Vendome J, Nneoyiegbe AF, Trembleau L, Botting CH, Smith MC, Jaspars M, Naismith JH (2012) The mechanism of patellamide macrocyclization revealed by the characterization of the PatG macrocyclase domain. Nat Struct Mol Biol 8:767-72.

Lawton L, Morris L, Jaspars M (1999) A bioactive modified peptide, aeruginosamide, isolated from the cyanobacterium *Microcystis aeruginosa*. J Org Chem 64:5329-5332.

Lee J, McIntosh J, Hathaway BJ, Schmidt EW (2009) Using marine natural products to discover a protease that catalyses peptide macrocyclization of diverse substrates. J Am Chem Soc 131:2122-2124.

Lee SW, Mitchell DA, Markley AL, Hensler ME, Gonzalez D, Wohlrab A, Dorrestein PC, Nizet V, Dixon JE (2008) Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA 105:5879-5884.

Li B, Sher D, Kelly L, Shi Y, Huang K, Knerr PJ, Joewong I, Rusch D, Chisholm SW, van der Donk (2010) Catalytic promiscuity in the biosynthesis of cyclic peptide secondary metabolites in planktonic marine cyanobacteria. Proc Natl Acad Sci USA107:10430–10435.

Li WL, Yi YH, Wu HM, Xu QZ, Tang HF, Zhou DZ, Lin HW, Wang ZH (2003) Isolation and structure of the cytotoxic cycloheptapeptide phakellistatin 13. J Nat Prod 66:146-148.

Linington RG, Gonzàles J, Ureña L-D, Romero LI, Ortega-Barria E, Gerwick WH (2007) Venturamides A and B: Antimalarial constituents of the Panamanian marine cyanobacterium *Oscillatoria* sp. J Nat Prod 70:397-401.

Long PF, Dunlap WC, Battershill CN, Jaspars M (2005) Shotgun cloning and heterologous expression of the patellamide gene cluster as a strategy to achieving sustained metabolite production. ChemBioChem 6:1760-5.

Mau CMS, Nakao Y, Yoshida WY, Scheuer PJ (1996) Waiakeamide, a cyclic hexapeptide from sponge *Ircinia dendroides*. J Org Chem 61:6302–6304.

McDonald LA, Foster MP, Phillips DR, Ireland CM, Lee AY, Clardy J (1992) Tawicyclamides A and B, new cyclic peptides from the ascidian *Lissoclinum patella*: studies on the solution- and solid-state conformations. J Org Chem 57:4616-4624.

McDonald LA, Ireland CM (1992) Patellamide E: a new cyclic peptide from the ascidian *Lissoclinum patella*. J Nat Prod 55:376-379.

McIntosh JA, Donia MS, Nair SK, Schmidt EW (2011) Enzymatic basis of ribosomal peptide prenylation in cyanobacteria. J Am Chem Soc 133: 13698-705.

McIntosh JA, Donia MS, Schmidt EW (2009) Ribosomal peptide natural products: Bridging the ribosomal and nonribosomal worlds. Nat Prod Rep 26:537–559.

McIntosh JA, Donia MS, Schmidt EW (2010) Insights into heterocyclization from two highly similar enzymes. J Am Chem Soc. 132:4089-91.

McIntosh JA, Lin Z, Tianero MD, Schmidt EW (2013) Aestuaramides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived Claisen rearrangements. ACS Chem Biol 8:877–883.

McIntosh JA, Schmidt EW (2010) Marine molecular machines: heterocyclization in cyanobactin biosynthesis. ChemBioChem 11:1413-21.

Milanowski DJ, Rashid MA, Gustafson KR, O'Keefe BR, Nawrocki JP, Pannell LK, Boyd MR (2004) Cyclonellin, a new cyclic octapeptide from the marine sponge *Axinella carteri*. J Nat Prod 67:441-444.

Milne BF, Long PF, Starcevic A, Hranueli D, Jaspars M (2006) Spontaneity in the patellamide biosynthetic pathway. Org Biomol Chem 4:631-8.

Mitchell SS, Faulkner DJ, Rubins K, Bushman FD (2000) Dolastatin 3 and two novel cyclic peptides from a Palauan collection of *Lyngbya majuscula*. J Nat Prod 63: 279-282.

Mohammed R, Peng J, Kelly M, Hamann MT (2006) Cyclic heptapeptides from the Jamaican sponge *Stylissa caribica*. J Nat Prod 69:1739-1744.

Montalbán-López M, Sánches-Hidalgo M, Cebrián R and Maqueda M (2012) Discovering the bacterial circular proteins: Bacteriocins, cyanobactins, and pilins. J Biol Chem 287:27007–27013.

Morris LA, van den Bosch JJK, Versluis K, Thompson GS, Jaspars M (2000) Structure determination and MSn analysis of two new lissoclinamides isolated from the Indo-Pacific ascidian *Lissoclinum patella*: NOE restrained molecular dynamics confirms the absolute stereochemistry derived by degradative methods. Tetrahedron 56:8345-8353.

Morris LA, Jaspars M, Kettenes-van den Bosch JJ, Versluis K, Heck AJR, Kelly SM, Price NC (2001a) Metal binding of *Lissoclinum patella* metabolites. Part 1: Patellamides A, C and ulithiacyclamide A. Tetrahedron 57: 3185–3197.

Morris LA, Milne BF, Jaspars M, Kettenes-van den Bosch JJ, Versluis K, Heck AJR, Kelly SM, Price NC (2001b) Metal binding of *Lissoclinum patella* metabolites. Part 2: Lissoclinamides 9 and 10. Tetrahedron 57:3199–3207.

Murakami M, Itou Y, Ishida K, Shin HJ (1999) Prenylagaramides A and B, new cyclic peptides from two strains of *Oscillatoria agardhii*. J Nat Prod 62:752-755.

Nagatsu A, Kajitani H, Sakakibara J (1995) Muscoride A: A New oxazole peptide alkaloid from freshwater cyanobacterium *Nostoc muscorum*. Tedrahedron Lett 36(23):4097–4100.

Napolitano A, Bruno I, Rovero P, Lucas R, Peris MP, Gomez-Paloma L, Riccio R (2001) Synthesis, structural aspects and bioactivity of the marine cyclopeptide hymenamide C. Tetrahedron 57: 6249-6255.

Neitz S, Jurgens M, Kellmann M, Schulz-Knappe P, Schrader M (2001) Screening for disulfide-rich peptides in biological sources by carboxyamidomethylation in combination with differential matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrometry 15: 1586–1592.

Nogle LM, Marquez BL, Gerwick WH (2003) Wewakazole, a novel cyclic dodecapeptide from a Papua New Guinea *Lyngbya majuscula*. Org Lett 5:3-6.

Nolan EM, Walsh CT. How nature morphs peptide scaffolds into antibiotics. (2008) ChemBioChem 10:34–53.

Ogino J, Moore RE, Patterson GML, Smith CD (1996) Dendroamides, new cyclic hexapeptides from blue-green alga. Multidrug-resistance reversing activity of dendroamide A. J Nat Prod 59:581-586.

Ojika M, Nemoto T, Nakamura M, Yamada K (1995) Dolastatin E, a new cyclic hexapeptide isolated from the sea hare *Dolabella-auricularia*. Tetrahedron Lett 36: 5057-5058.

Oman TJ, van der Donk WA (2010) Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol. 6:9-18.

Perez LJ, Faulkner DJ (2003) Bistratamides E-J, modified cyclic hexapeptides from the Philippines ascidian *Lissoclinum bistratum*. J Nat Prod 66: 247-250.

Pettit GR, Cichacz Z, Barkoczy J, Dorsaz AC, Herald DL, Williams MD, Doubek DL, Schmidt JM, Tackett LP, Brune DC, Cerny RL, Hooper JNA, Bakus GJ (1993a) Antineoplastic Agents. 249. Isolation and structure of the marine sponge cell-growth inhibitory cyclic peptide phakellistatin 1. J Nat Prod 56:260-267.

Pettit GR, Gao F, Cerny RL, Doubek DL, Tackett LP, Schmidt JM, Chapuis JC (1994) Antineoplastic Agents. 278. Isolation and structure of axinastatin 2 and axinastatin 3 from a western Caroline island marine sponge. J Med Chem 37:1165-1168

Pettit GR, Tan R, Williams MD, Tackett L, Schmidt JM, Cerny RL, Hooper JNA (1993b) Isolation and structure of phakellistatin 2 from the eastern indian-ocean marine *Sponge Phakellia carteri*. Bioorg Med Chem Lett 3:2869-2874.

Pettit GR, Tan R (2003) Antineoplastic agents 390. Isolation and structure of phakellistatin 12 from a chuuk archipelago marine sponge. Bioorg Med Chem Lett 13:685-688.

Pettit GR, Xu JP, Dorsaz AC, Williams MD, Boyd MR, Cerny RL (1995) Antineoplastic Agents Series Contribution. 324. Isolation and structure of the human cancer cell-growth inhibitory cyclic decapeptides phakellistatin 7, phakellistatin 8 and phakellistatin 9. Bioorg Med Chem Lett 5:1339-1344.

Philmus B, Christiansen G, Yoshida WY, Hemscheidt TK (2008) Post-translational modification in microviridin biosynthesis. ChemBioChem 9:3066-3073.

Ploutno A, Carmeli S (2002) Modified peptides from a water bloom of the cyanobacterium *Nostoc* sp. Tetrahedron 58:9949-9957.

Portmann C, Blom JF, Gademann K, Jüttner F (2008a) Aerucyclamides A and B: Isolation and synthesis of toxic ribosomal heterocyclic peptides from the cyanobacterium *Microcystis aeruginosa* PCC 7806. J Nat Prod 71:1193-1196.

Portmann C, Blom JF, Kaiser M, Brun R, Jüttner F, Gademann K (2008b) Isolation of aerucyclamides C and D and structure revision of microcyclamide 7806A: Heterocyclic ribosomal peptides from *Microcystis aeruginosa* PCC 7806 and their antiparasite evaluation. J Nat Prod 71:1891-1896.

Prinsep MR, Moore RE, Levine IA, Patterson GML (1992) Westiellamide, a bistratamide-related cyclic peptide from the blue-green alga *Westiellopsis prolifica*. J Nat Prod 55:140-142.

Raveh A, Carmeli S (2010) Aeruginazole A, a novel thiazole-containing cyclopeptide from the cyanobacterium *Microcystis* sp. Org Lett 12:3536–3539.

Rashid MA, Gustafson KR, Cardellina JH II, Boyd MR (1995) Patellamide F, a new cytotoxic cyclic peptide from the colonial ascidian *Lissoclinum patella*. J Nat Prod 58:594-597.

Rudi A, Aknin M, Gaydou EM, Kashman Y (1998) Four new cytotoxic cyclic hexa- and heptapeptides from the marine ascidian *Didemnum molle*. Tetrahedron 54: 13203-13210.

Rudi A, Chill L, Aknin M, Kashman Y (2003) Didmolamide A and B, two new cyclic hexapeptides from the marine ascidian *Didemnum molle*. J Nat Prod 66:575-577.

Salvatella X, Caba JM, Albericio F, Giralt E (2003) Solution structure of the antitumor candidate trunkamide A by 2D NMR and restrained simulated annealing methods. J Org Chem 68:211-215.

Schmidt EW, Donia MS (2009) Cyanobactin ribosomally synthesized peptides - a case of deep metagenome mining. Meth Enzymol 458:575-596.

Schmidt EW, Donia MS (2010) Life in cellulose houses: symbiotic bacterial biosynthesis of ascidian drugs and drug leads. Curr Opin Biotechnol 21:827-833.

Schmidt G, Grube A, Kock M (2007) Stylissamides A-D - new proline-containing cyclic heptapeptides from the marine sponge Stylissa caribica. European J Org Chem 24:4103-4110.

Schmitz FJ, Ksebati MB, Chang JS, Wang JL, Hossain MB, van der Helm D (1989) Cyclic peptides from the ascidian Lissoclinum patella: conformational analysis of patellamide D by X-ray analysis and molecular modelling. J Org Chem 54:3463-3472.

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

Sera Y, Adachi K, Fujii K, Shizuri Y (2003) A New antifouling hexapeptide from a Palauan Sponge, *Haliclona* sp. J Nat Prod 66: 719–721.

Shih PM , Wu D, Latifi A, Axen SD, Fewer DP, Talla E, Calteau A, Cai F, Tandeau de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin LA, Nolan M, Davenport KW, Han CS, Rubin E, Eisen JA, Woyke T, Gugger M, Kerfeld CA (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc Natl Acad Sci USA: 110:1053–1058.

Shin HJ, Matsuda H, Murakami M, Yamaguchi K (1996) Agardhipeptins A and B, two new cyclic hepta- and octapeptide, from the cyanobacterium *Oscillatoria agardhii* (NIES-204). Tetrahedron 52:13129-13136.

Sings HL, Rinehart KL (1996) Compounds produced from potential tunicate-blue-green algal symbiosis: A review. J Ind Microbiol Biotechnol 17:385-396.

Sivonen K, Jones G (1999) Cyanobacterial toxins, p. 41–111. In: Chorus, I. and Bartram, J (ed.), Toxic cyanobacteria in water. A guide to their public health consequences, monitoring and management, E & FN Spon, London, England.

Sivonen K, Leikoski N, Fewer DP, Jokela J (2010) Cyanobactins-ribosomal cyclic peptides produced by cyanobacteria. Appl Microbiol Biotechnol 86:1213-1225.

Sone H, Kigoshi H, Yamada K (1997) Isolation and stereostructure of dolastatin I, a cytotoxic cyclic hexapeptide from the Japanese sea hare *Dolabella auricularia*. Tetrahedron 53: 8149-8154.

Sudek S, Haygood MG, Youssef DTA, Schmidt EW (2006) Structure of trichamide, a cyclic peptide from the bloom-forming cyanobacterium *Trichodesmium erythraeum*, predicted from the genome sequence. Appl Environ Microbiol 72:4382-4387.

Tabudravu J, Morris LA, Kettenes-van den Bosch JJ, Jaspars M (2001) Wainunuamide, a histidine-containing proline-rich cyclic heptapeptide isolated from the Fijian marine sponge *Stylotella aurantium*. Tetrahedron Lett 42:9273-9276.

Tabudravu JN, Morris LA, Kettenes-van den Bosch JJ, Jaspars M (2002a) Axinellin C, a proline-rich cyclic octapeptide isolated from the Fijian marine sponge *Stylotella aurantium*. Tetrahedron 58: 7863-7868.

Tabudravu JN, Jaspars M, Morris LA, Kettenes-van den Bosch JJ, Smith N (2002b) Two distinct conformers of the cyclic heptapeptide phakellistatin 2 isolated from the Fijian marine sponge *Stylotella aurantium*. J Org Chem 67:8593-8601.

Tang W, van der Donk WA (2012) Structural characterization of four prochlorosins: A novel class of lantipeptides produced by planktonic marine cyanobacteria. Biochemistry 51:4271–4279.

Teruya T, Sasaki H, Suenaga K (2008) Hexamollamide, a hexapeptide from an Okinawan ascidian *Didemnum molle*. Tetrahedron Lett 49:5297-5299.

Thorstholm L, Craik DJ (2012) Discovery and applications of naturally occurring cyclic peptides. Drug Discovery Today: Technologies 9:1:13.

Tianero MD, Donia MS, Young TS, Schultz PG, Schmidt EW (2012) Ribosomal route to small-molecule diversity. J Am Chem Soc. 134:418-25.

Todorova AK, Jüttner F, Linden A, Plüss T, von Philipsborn W (1995) Nostocyclamide: a new macrocyclic, thiazole-containing allelochemical from *Nostoc* sp. 31 (cyanobacteria). J Org Chem 60:7891-7895.

Toske SG, Fenical W (1995) Cyclodidemnamide - a new cyclic heptapeptide from the marine ascidian *Didemnum molle*. Tetrahedron Lett 36:8355-8358.

Tsuda M, Sasaki T, Kobayashi J (1994) Hymenamide G, hymenamide H, hymenamide J, and hymenamide-K, 4 new cyclic octapeptides from the okinawan marine sponge *Hymeniacidon* sp. Tetrahedron 50:4667-4680.

Van Wagoner R, Drummond AK, Wright JLC (2007) Biogenetic diversity of cyanobacterial metabolites. Adv Appl Microbiol 61:89–217.

Velásquez JE, van der Donk WA (2011) Genome mining for ribosomally synthesized natural products. Curr Opin Chem Biol 15:11-21.

Voß B, Bolhuis H, Fewer DP, Kopf M, Möke F, Haas F, El-Shehawy R, Hayes P, Bergman B, Sivonen K, Dittmann E, Scanlan DJ, Hagemann M, Stal LJ, Hess WR (2013) Insights into the physiology and ecology of the brackish-water-adapted cyanobacterium *Nodularia spumigena* CCY9414 based on a genome-transcriptome analysis. PLoS ONE 8(3): e60224.

Wang H, Fewer DP, Sivonen K (2011) Genome mining demonstrates the widespread occurrence of gene clusters encoding bacteriocins in cyanobacteria. PLoS ONE 6(7): e22384.

Wang H, Sivonen K, Rouhiainen L, Fewer DP, Lyra C, Rantala-Ylinen A, Vestola J, Jokela J, Rantasärkkä K, Li Z, Liu B (2012) Genome-derived insights into the biology of the hepatotoxic bloom-forming cyanobacterium *Anabaena* sp. strain 90. BMC Genomics 13:613.

Wasylyk JM, Biskupiak JE, Costello CE, Ireland CM (1983) Cyclic peptide structures from the tunicate *Lissoclinum patella* by FAB mass spectrometry. J Org Chem 48:4445-4449.

Welker M, von Döhren H (2006) Cyanobacterial peptides – Nature's own combinatorial synthesis. FEMS Microbial Reviews 30:530–563.

Wesson KJ, Hamann M (1996) Keenamide A, a bioactive cyclic peptide from the marine mollusc *Pleurobranchus forskalii*. J Nat Prod 59:659–631.

Williams DE, Moore RE, Paul VJ (1989) The structure of ulithiacyclamide B. Antitumor evaluation of cyclic peptides and macrolides from *Lissoclinum patella*. J Nat Prod 52:732-739.

Wipf P (1995) Synthetic studies of biologically active marine cyclopeptides. Chem Rev 95:2115–2134.

Zabriskie TM, Foster MP, Stout TJ, Clardy J, Ireland CM (1990) Studies on the solution- and solid-state structure of patellin 2. J Am Chem Soc 112:8080–8084.

Ziemert N, Ishida K, Liaimer A, Hertweck C, Dittmann E (2008a) Ribosomal synthesis of tricyclic depsipeptides in bloom-forming cyanobacteria. Angew Chem Int Ed 47:7756-7759.

Zimert N, Ishida K, Quillardet P, Bouchier C, Hertweck C, de Marsac NT, Dittmann E (2008b) Microcyclamide biosynthesis in two strains of *Microcystis aeruginosa:* from structure to genes and viece versa. Appl Environ Microbiol 74:1791–1797.

Ziemert N, Ishida K, Weiz A, Herweck C, Dittmann E (2010) Exploiting the natural diversity of microviridin gene clusters for discovery of novel tricyclic depsipeptides. Appl Environ Microbiol 76:3568–3574.