

School of Doctoral Studies in Biological Sciences  
University of South Bohemia in České Budějovice  
Faculty of Science  
Department of Ecosystem Biology

**Diversity and Ecological Role of Cyanobacterial Lipopeptides**

Ph.D. Thesis

Mgr. Tomáš Galica

Supervisor:

RNDr. Pavel Hrouzek, Ph.D.

Institute of Microbiology, Czech Academy of Sciences

Co-Supervisor:

RNDr. Jan Mareš, Ph.D.

Institute of Hydrobiology, Biology Centre, Czech Academy of Sciences

České Budějovice

2021

**This thesis should be cited as:**

**Galica T., 2021:** Diversity and ecological role of cyanobacterial lipopeptides. Ph.D. Thesis Series, No. 13 University of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Sciences, České Budějovice, Czech Republic, 2021, pp.

**Annotation**

The presented study focuses on the biosynthesis and distribution of lipopeptides in the context of evolutionary history and life strategies of cyanobacteria. Cyanobacteria are ancient organisms often naturally exposed to situations in which other bacteria, such as *Bacillus* or *Pseudomonas*, were shown to employ lipopeptides to facilitate colony expansion or iron acquisition. The study explores these analogies and maps the distribution and diversity of lipopeptides across cyanobacteria with various lifestyles. Furthermore, the study provides evidence that widespread cyanobacterial lipopeptides containing a particular structural motif with two  $\beta$ -hydroxy aspartate residues are a part of iron stress response and likely facilitate iron acquisition. The gathered information aims to integrate and deepen our knowledge of the ecophysiological roles of cyanobacterial lipopeptides.

**Declaration**

I hereby declare that I have written the presented thesis by myself using the literature provided in the list of references.

.....

Tomáš Galica  
České Budějovice  
19. 6. 2021

## **Financial Support**

The work presented in the study was conducted at Algatech (Třeboň, Institute of Microbiology) and at the Institute of Hydrobiology (České Budějovice). Financial support was provided by the Czech Science Foundation (GAČR; #16-09381S), the Grant Agency of the University of South Bohemia project no. 04-158/2016/P, the Ministry of Education, Youth and Sports of the Czech Republic (project Algamic CZ.1.05/2.1.00/19.0392; National Programme of Sustainability I #LO1416 and RVO 67985939) and by the Ministry of Regional Development of the Czech Republic - Cross-Border cooperation Czech-Bavaria project #41.

## **Acknowledgments**

First of all, I would like to thank my dear supervisors, Pavel Hrouzek and Jan Mareš for their patience and support, not just in the lab, but also as friends outside working hours. In the later stages of my studies, their support was supplemented with sufficient space and much enjoyed academic freedom to develop my own ideas. I would also like to thank Klára Řeháková, who supervised me during my first year and greatly facilitated the start of my studies in České Budějovice.

Further I would like to thank all the people from Opaťák, or Centre Algatech of Institute of Microbiology, for the amazing atmosphere they create daily at this wonderful place to do science. There is enough good humor, fun and inspiration for everyone seeking a break from heavy duty science work. Yet, on the other hand, there is also enough support, advice and consolation, when experiments (scientific or personal) do not go all too well. Similarly, my gratitude goes to the people of the Department of Ecology, for maintaining a friendly and welcoming atmosphere. The annual conferences of the department were always a big and appreciated event, that I was looking forward to.

Many thanks for the chance to work abroad go to David P. Fewer and prof. Kaarina Sivonen. It was an important and pleasant internship, which further facilitated my professional growth. Moreover, in their laboratory I have found fine company that made my stay abroad a very pleasant and smooth experience.

Last but not least, I would like to thank my family and dear friends for the comforting support and company during the long years of my study.

## List of scientific publications and the author's contribution

The presented thesis is based on the following articles:

### Paper I.

Urajová, P., Hájek, J., Wahlsten, M., Jokela, J., **Galica, T.**, Fewer, D., Kust, A., Zapomělová-Kozlíková, E., Delawska, K., Sivonen, K., Kopecký, J. & Hrouzek, P. (2016). A liquid chromatography-mass spectrometric method for the detection of cyclic beta-amino fatty acid lipopeptides. *Journal of Chromatography A*, 1438, 76-83.

<https://doi.org/10.1016/j.chroma.2016.02.013> (IF=4.049)

*TG participated in the mass spectrometry experiments investigating the relationship between the collision energy and fragmentation patterns of different lipopeptides, subsequent data analysis and presentation. (8.333%)*

### Paper II.

**Galica, T.**, Hrouzek, P. & Mareš, J. (2017). Genome mining reveals high incidence of putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP ligase genes in biofilm-forming cyanobacteria. *Journal of Phycology*, 53(5), 985-998. <https://doi.org/10.1111/jpy.12555> (IF=2.328)

*TG participated in the study design, conducted the bioinformatic screen and wrote the corresponding part of the manuscript. (50.0%)*

### Paper III.

Mareš, J., Hájek, J., Urajová, P., Kust, A., Jokela, J., Saurav, K., **Galica, T.**, Čapková, K., Mattila, A., Haapaniemi, E., Permi, P., Mysterud, I., Skulberg, O., Karlsen, J., Fewer, D., Sivonen, K., Tonnesen, H. & Hrouzek, P. (2019). Alternative Biosynthetic Starter Units Enhance the Structural Diversity of Cyanobacterial Lipopeptides. *Applied and Environmental Microbiology*, 85(4), Article ARTN e02675-18. <https://doi.org/10.1128/AEM.02675-18> (IF=3.960)

*TG participated in the analysis of FAAL domain specificity in relationship to the diversity of the hydrophobic residue of produced puwainaphycins and minutissamides. (5.555%)*

#### **Paper IV.**

**Galica T.**, Borbone N., Mareš J., Kust A., Caso A., Esposito G., Saurav K., Hájek J., Řeháková K., Urajová P., Costantino V. & Hrouzek P. (2021) Cyanochelins, an overlooked class of widely distributed cyanobacterial siderophores, discovered by silent gene cluster awakening. *Journal of Applied Environmental Microbiology*, Epub ahead of print. <https://doi.org/10.1128/AEM.03128-20> (IF=3.960)

*TG designed the study, performed the bioinformatic screening, cultivated the strains, isolated the candidate compounds, characterized the compounds by mass spectroscopy, conducted the photolytic experiments and together with JM and PH wrote the article. (50.0%)*

#### **Paper V.**

Mareš J., Chmelík D., **Galica T.**, Urajová P., Hájek J., Řeháková K. & Hrouzek P. (manuscript under revision) A newly discovered NRPS/PKS pathway of the cyanobacterial lipopeptide muscotoxin utilizes 4-methyl proline generated by the nostopeptolide biosynthetic machinery.

*TG participated in the design of the isotopic labeling, feeding experiments and interpretation of the data. (14.285%)*

**The corresponding authors of the listed publications acknowledge the contribution of TG as stated above:**

**RNDr. Pavel Hrouzek, Ph.D.**

**RNDr. Jan Mareš, Ph. D.**

# Table of Contents

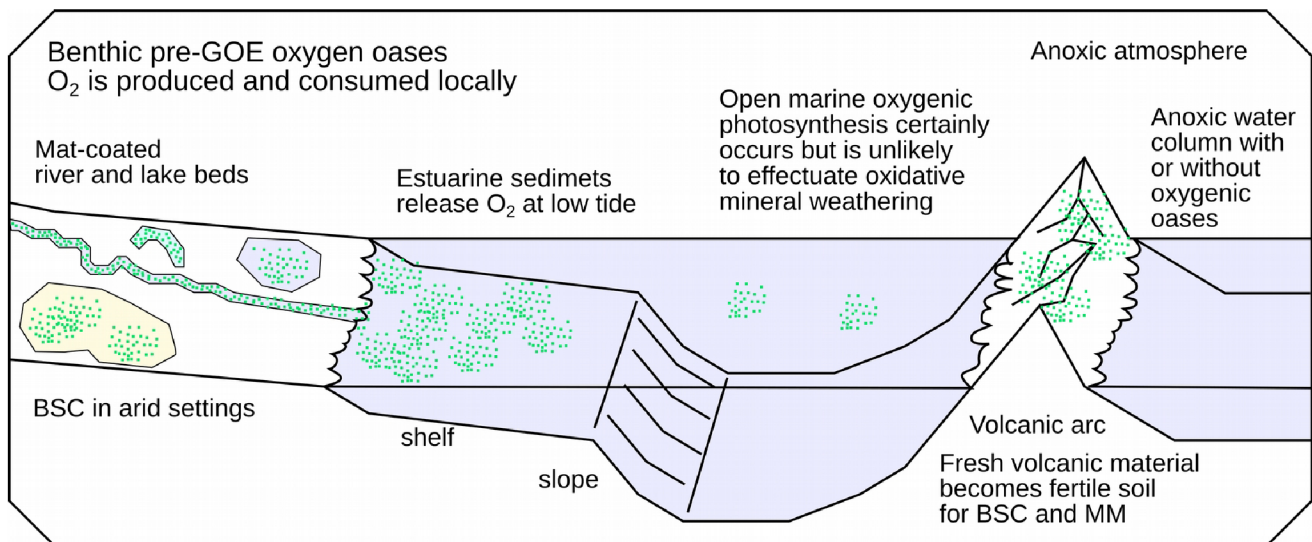
<b>1. Introduction</b>	<b>1</b>
1.1. Global Significance of Cyanobacteria	1
1.2. Cyanobacterial Lifestyles	6
1.3 Taxonomy and Phylogenetic Diversity	8
1.4. Lipopeptides	10
1.4.1. <i>Ecophysiological Roles of Bacterial Lipopeptides</i>	12
Role in Colony Development and Spreading	14
Role in Iron Scavenging	17
1.4.2. <i>Biosynthesis of Lipopeptides</i>	19
1.4.3. <i>Cyanobacterial Lipopeptides</i>	24
Microginins	25
Puwainaphycins and Minutissamides	26
Laxaphycins	27
Anabaenolysins	28
Hassallidins and Baltacidins	29
The Kulolide Superfamily	31
Other Lipopeptides	32
1.5. Summary of the Research Background and Rationale	34
<b>2. Results and Discussion</b>	<b>35</b>
2.1. Outline of the Research	35
2.2. Research Paper I. - Fragmentation of Lipopeptides Under High Collision Energy	36
2.3. Research Paper II. - Search for Novel Lipopeptide Biosynthesis Encoding Gene Clusters in Cyanobacteria and their Distribution	38
2.4. Research paper III. - Investigation of Fatty Acid Loading Modules in Puwainaphycin Gene Clusters.	40
2.5. Research Paper IV. - in press - Discovery of Cyanochelins – Novel and Possibly Widespread Cyanobacterial Siderophores	43
2.6. Research Paper V. - manuscript – Biosynthesis of Muscotoxin in <i>Desmonostoc muscorum</i> and Its Crosstalk with Nostopeptolide Synthesis.	47
2.7. Summary and future perspectives	49
<b>3. References</b>	<b>52</b>
<b>4. Attached Publications</b>	<b>71</b>
4.1. Paper I.	71
4.2. Paper II.	80
4.3. Paper III.	95
4.4. Paper IV.	113
4.5. Paper V. (manuscript)	127
<b>5. Curriculum Vitae</b>	<b>136</b>

# 1. Introduction

## 1.1. Global Significance of Cyanobacteria

Extant cyanobacteria are descendants of an ancient prokaryotic lineage in which the oxygenic photosynthesis evolved (Whitton et al 2012, Sánchez-Baracaldo and Cardona 2020). Their ancestors are considered the driving force behind the accumulation of oxygen in the atmosphere, known as the Great Oxidation Event (GOE), which substantially changed the conditions on Earth around 2.4 billion years ago (bya) and supported further evolution of aerobic life (Whitton et al 2012, Sánchez-Baracaldo and Cardona 2020). Around that time multicellularity evolved within the cyanobacterial phylum and spurred substantial diversification of morphological types that formed the basis of major contemporary lineages (Schirmer et al 2013). Existence of derived multicellular cyanobacteria already at 2 bya is unambiguously supported by the fossil record, thus suggesting their earlier origin (Demoulin et al 2019). In the primordial absence of plants and animals, the coverage of land mass by ancestral cyanobacteria was likely limited by substantially different factors than today and the cyanobacterial mats possibly covered a fairly large area (Finke et al 2019). The observed widespread oxygenic weathering suggests that the oxygen-evolving cyanobacterial mats were spread to the extent that influenced the global geochemical cycles even prior to GOE (Fig. 1; Lalonde and Konhauser 2015). Furthermore, an ancient eukaryote eventually engulfed a cyanobacterium that later became the chloroplast and gave rise to phototrophic algae and modern day plants. The evolutionary invention of oxygenic photosynthesis, the GOE and the genesis of the chloroplast are unique events that shaped the Earth and in which cyanobacteria played an essential role.





**Fig. 1: Model of pre-GOE cyanobacterial biosphere.** Image according to Lalonde and Konhauser (2014). The model emphasizes local production of oxygen by cyanobacterial mats and its immediate consumption by oxidative weathering of mat-covered mineral surface before the accumulation of  $O_2$  in the atmosphere. BSC stands for biological soil crusts; MM for microbial mats.

Planetary processes continue to be influenced by contemporary cyanobacteria, most notably by the marine species that account for a fair share of carbon and nitrogen fixation. The extent of their role is still a subject to scientific debate, yet there is a general agreement on their significance (Moisander et al 2017). Important contributors to marine primary production are small unicellular cyanobacteria of the genera *Prochlorococcus* and *Synechococcus*, which are abundant and distributed over the mixed layer of water throughout vast areas of oligotrophic oceans. Another marine genus, *Trichodesmium*, is found floating in dense accumulations of filaments often forming seasonal blooms in tropical and subtropical waters. Apart from the blooms that influence marine carbon balance, *Trichodesmium* is also a prominent diazotroph and for about 50 years it was considered the main contributor to oceanic nitrogen fixation (Capone et al 2005, Moisander et al 2010, Martínez-Pérez et al 2016). Recently, symbiotic associations of cyanobacteria such as *Richellia intracellularis* or representatives of unicellular cyanobacteria group A (UCYN-A) with diatoms and haptophytes, respectively, were recognized as important marine nitrogen-fixers (Moisander et al 2010, Martínez-Pérez et al 2016). The fixation in such symbioses is performed by the cyanobiont. Additional contributors to marine nitrogen fixation include representatives

of UCYN-B and UCYN-C groups of cyanobacteria (Goebel et al 2010). It remains unclear which groups of diazotrophic cyanobacteria have the greatest impact on marine nitrogen fixation. However, as a whole, cyanobacteria remain important players in marine nitrogen cycling, despite the growing recognition of other non-cyanobacterial marine diazotrophs in the past few years (Moisander et al 2017).

Habitats such as primordial microbial jungle or the vast oceans are large enough so that just their mere size infers their global importance in nutrient cycling. But many cyanobacteria are crucial players in nutrient cycling in habitats of considerably smaller size but scattered across the globe, such as rivers, lakes and ponds, but also arid areas and deserts.

Diverse cyanobacteria such as *Microcystis*, *Aphanizomenon*, *Planktothrix* or *Dolichospermum* can be found alongside algae in the freshwater plankton in various degree of abundance, in accordance with the plankton paradox (Hutchinson 1961). The absence of equilibrium and ever-changing conditions of freshwater pelagic water mass likely prevent competitive exclusion, allowing for a multitude of species to coexist (Borics et al 2021). However, in overly nutritious lakes, cyanobacteria can develop aggressive blooms dominated by one species that are then disturbed by seasonal changes (Paerl and Huisman 2009). Such blooms can have considerable impact on the water body as they can limit the oxygen availability to fish and other aquatic life during nights. Moreover, cyanobacterial blooms can produce toxins and contaminate sources of drinking water to humans and animals (Dittmann et al 2013).

Apart from the pelagic, cyanobacteria of the genera *Anabaena*, *Oscillatoria* or *Phormidium* are successful colonizers of illuminated lake beds, where they grow over various surfaces, including living plants, eventually forming dense mats that can float (Durako et al 1982, Bouma-Gregson et al 2017). Similarly, in marine coastal areas *Moorea* or *Lyngbya* are found growing in the benthos, over corals or other plant surfaces (Engene et al 2012). The epilithic cyanobacteria such as *Schizothrix*, *Nunduva* or *Rivularia* prefer shallow lagoons and disturbed intertidal and splash zones, where they grow over rocks and can eventually form stromatolites (González-Resendiz et al 2018, Reid et al 2000).

Cyanobacteria often thrive as pioneering species in otherwise inhospitable parts of the world, where competitive pressure from eukaryotic phototrophs is less pronounced. Species of *Microcoleus*, *Leptolyngbya* or *Nostoc* cope well with desiccation, nutrient scarcity and freeze-thaw cycles and form biological soil crusts found in deserts, including those in high altitudes (Campbell 1979, Oren et al 2017, Řeháková et al 2007). The psychrotrophic *Oscillatoria*, *Phormidium* or *Nostoc* are prevalent in microbial mats in soils, wetted surfaces and lakes in polar regions of both hemispheres (Vincent et al 1993, Tang et al 1997, de los Rios et al 2015). In both the arid and the polar regions, cyanobacteria are often the major primary producers and a source of fixed nitrogen (Oren et al 2017, Janatková et al 2013). In regions with a less severe climate, terrestrial subaerophytic cyanobacterial mats are confined to rather marginal places not easily accessible by plants, such as rocks, and play a minor supportive role in the primary production (Finke et al 2019, Williams et al 2016, Hauer et al 2015). A peculiar and rather rare habitat with extreme physico-chemical conditions that frequently support development of cyanobacterial mats are the hot springs. A prominent example is the globally distributed hot spring dweller *Mastigocladus laminosus* (formerly *Fischerella*), often found growing around sources of hot water up to temperatures of 57°C (Miller et al 2007), but other cyanobacterial genera such as *Synechococcus* or *Phormidium* also include thermophilic species (Whitton 2012).

Probably the smallest habitat that cyanobacteria inhabit is other organisms: close ecological and physical proximity led to associations that have thrived ever since their establishment. The oldest symbiotic association of once a cyanobacterium with an ancestral eukaryote gave rise to the chloroplast, which still today provides the host cell with energy in the form of reduced carbon. More recently, organisms entered symbiotic relationships with cyanobacteria to benefit from their capability to fix atmospheric nitrogen (Ran et al 2010, Hilton et al 2013). A wide array of organisms have evolved symbioses with cyanobacteria with varying level of cyanobiont-host intimacy, these include mainly: mosses, ferns, cycads, fungi and lichens, diatoms, haptophytes, corals, sponges and molluscs (Whitton 2012). In the tight symbioses, where the cyanobiont is securely transferred to the next generation, it gradually loses its metabolic

capabilities and slowly undergoes genome reduction (*Nostoc azollae*, *Richellia intracellularis*; Ran et al 2010, Hilton et al 2013). In contrast, the cyanobacteria forming facultative symbioses *de novo* in every generation need to preserve higher metabolic skills to survive without the host, but also additional signaling machinery to successfully find and establish a new symbiosis (Hilton et al 2013, Caputo et al 2019).

## 1.2. Cyanobacterial Lifestyles

The habitat in which cyanobacteria originally evolved and spread from remains a subject of ongoing scientific debate. Based on phylogenomic analysis and modern ecology of basal strains of cyanobacteria some authors propose low salinity or fresh-water origin of ancestral cyanobacteria (Blank and Sánchez-Baracaldo 2010, Sánchez-Baracaldo 2015). On the contrary some chose to emphasize the geological record which captured cyanobacteria mostly from estuarine or shallow marine environments (Demoulin et al 2019). However, both the phylogenetic and paleontological approach suggest that the ancestral cyanobacteria have lived either in the benthos or have formed microbial mats (Sánchez-Baracaldo 2015, Demoulin et al 2019). According to current knowledge cyanobacterial mats had grown undisturbed over considerable areas prior to the emergence of eukaryotes; moreover, some suggest this to be the original cyanobacterial lifestyle (Finke et al 2019, Lalonde and Konhauser 2014). The colonization of the open waters seems to have happened later and may have occurred repeatedly after the formation of crown groups cyanobacteria with the prominent pioneers being recruited from different lineages (Sánchez-Baracaldo 2015). As mentioned above small unicellular cyanobacteria such as *Prochlorococcus* and *Synechococcus* dominate the illuminated portions of large areas of oligotrophic oceans. Being small-sized, their surface to volume ratio ensures an easier access to limiting nutrients in comparison to their larger phytoplankton competitors (Raven 1998, Partensky et al 1999). In contrast, the widespread *Trichodesmium* descended from a separate lineage of filamentous cyanobacteria and prefers open water with a higher trophic state and its triumph over marine eukaryotic algae could be attributed to its high temperature optimum (Capone et al 2005, Breitbarth et al 2007).

Although no particular study was dedicated to the origin of cyanobacteria of the freshwater pelagic zone, the colonization therein had likely happened in a similar way. In addition, on a timescale counting in millions of years, cyanobacterial lineages may have switched from pelagic to benthic repeatedly, expanding on occasional floating of cyanobacterial mats. In this sense, the life history of *Microcystis* that inhabits the pelagic

during summer while overwintering in the sediment is particularly interesting (Reynolds et al 1981, Sánchez-Baracaldo 2019).

### 1.3 Taxonomy and Phylogenetic Diversity

Historically, the Cyanobacteria or Cyanophyceae fell under the interest of phycology as a part of botany and hence their taxonomy and especially nomenclature followed botanical conventions (Oren and Ventura 2017). With the advancement of electron-microscopic and molecular methods, the prokaryotic nature of cyanobacteria became widely acknowledged (Oren and Ventura 2017, Willis and Woodhouse 2020).

In the field of prokaryotic taxonomy, the species concept remains an open question. The Evolutionary Species Concept, the Ecotypic Species Concept, the Phylogenetic Species Concept, and the Monophyletic Species Concept are all considered relevant for cyanobacteria (Johansen and Casamatta 2005). Although the concepts differ in where the species boundaries are drawn, they all recognize the importance of shared evolutionary history and genetic information. The presented thesis aims to investigate the production of lipopeptides across the cyanobacterial phylum. For that purpose, the precise delimitation of species is not essential. More important are the phylogenetic relationships among the isolated cyanobacterial strains and pronounced lineages reflecting the evolutionary history, and offering an interesting context for the possible spread of the genes coding lipopeptide biosynthesis along the lineage or by a horizontal gene transfer.

The classification of cyanobacteria is currently undergoing revision as accumulation of genomic data reveals a complicated branching of the cyanobacterial tree of life. Across the phylum there are clearly monophyletic and species-rich derived lineages, interwoven with long branches with a few representatives (Mareš 2018). Moreover, the distinct morphological features – such as the filamentous or colonial growth – were repeatedly acquired and lost during the evolution, hence cyanobacteria with a similar morphology frequently descended from different lineages and, vice versa, a monophyletic group can include filamentous as well as unicellular species (Sánchez-Baracaldo 2015, Schirromeister 2013, Mareš 2018).

As was recognized by Mareš (2018) the previously established orders Gloeobacterales, Spirulinales, Chroococciopsidales and Nostocales appear to withstand the reevaluation by molecular phylogenetics. In

contrast, Pleurocapsales and Chroococcales, despite being mixed, are compact, while Synechococcales and Oscillatoriales are extensively polyphyletic and in desperate need for taxonomic revision (Mareš 2018). Due to the complicated situation in cyanobacterial taxonomy and phylogeny whenever the distribution of lipopeptides in the phylum was investigated during my work it was accompanied by a phylogenetic tree rather than referring to the currently elusive taxonomic ranks.



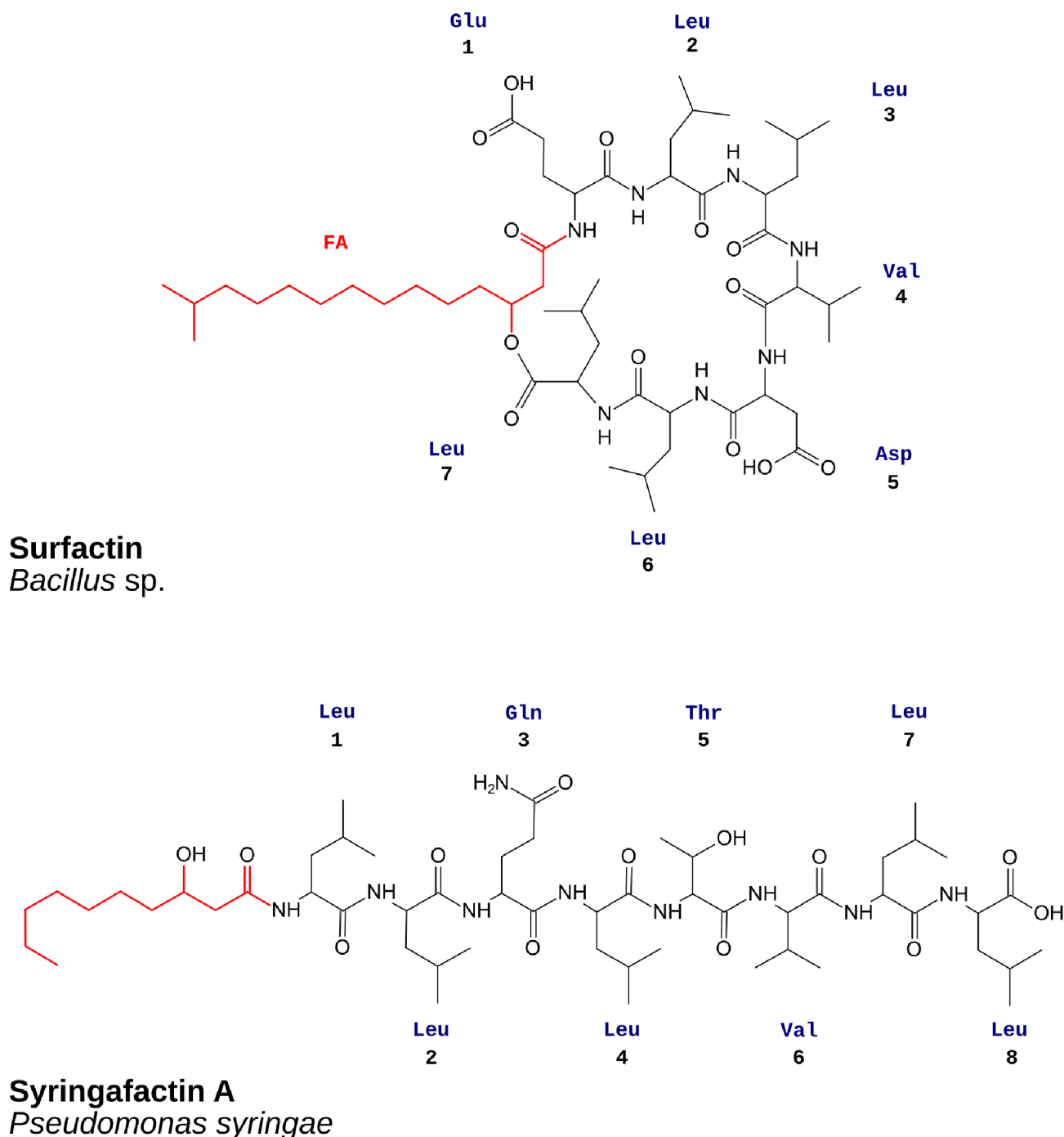
## 1.4. Lipopeptides

Microorganisms harbor the potential to produce complex organic molecules of immense structural variability that could possibly be employed for biotechnological or pharmaceutical purposes. Considerable effort has been dedicated to discovery of such compounds that are often called secondary metabolites or natural products. In attempts to discover potentially valuable natural products, the investigation of their benefits for their actual producers is neglected in favor of their possible usage (Traxler and Kolter 2015). Secondary metabolites are compounds of small molecular weight unnecessary for autonomous viability in laboratory conditions, in which related strains with a similar life strategy and energy metabolism may exhibit equal growth, but substantial difference in their secondary metabolite content. However, the significance of secondary metabolites comes to light when investigated in particular ecological contexts. In specific naturally occurring situations the secondary metabolites are important for interactions among microbes or with the environment to a degree that can decide species success (Chevrette et al 2020, Traxler and Kolter 2015). The term “specialized metabolite” is often used to address those secondary metabolites, for which the benefit they provide to the producer have been described.

Structural variability of secondary metabolites is overwhelming. There are small polyketides, non-ribosomally synthesized peptides and various hybrids of the two, then there are large ribosomally synthesized peptides and many more (Kehr et al 2011, Welker and von Döhren 2006). To focus on a particular group is hence a practical necessity. In the presented study I have focused on lipopeptides from cyanobacteria.

Lipopeptides consist of several amino acid residues connected by peptide bonds and a hydrophobic (lipid) moiety (Fig. 2). The definition, however, is not extremely strict on how many peptide bonds a lipopeptide should contain or how the hydrophobic moiety should look like (jamaicamide vs. hassallidin, Edwards et al 2004, Neuhof et al 2005). Perhaps more important is the resulting functional amphiphilicity of the compounds, a property that demands the presence of hydrophilic and hydrophobic structural elements, with the former often ensured by the extended hydrocarbon chain (Cochrane and Vederas 2016). The amphiphilic nature of

lipopeptides may limit their solubility and diffusivity in aqueous solutions but ensures their potential to reduce surface tension or to form micelles and pores in biological membranes under certain circumstances (Cochrane and Vederas 2016, Kem and Butler 2009). The pronounced hydrocarbon chain of lipopeptides hence ensures an important functionality that the non-acylated bacterial peptides lack.

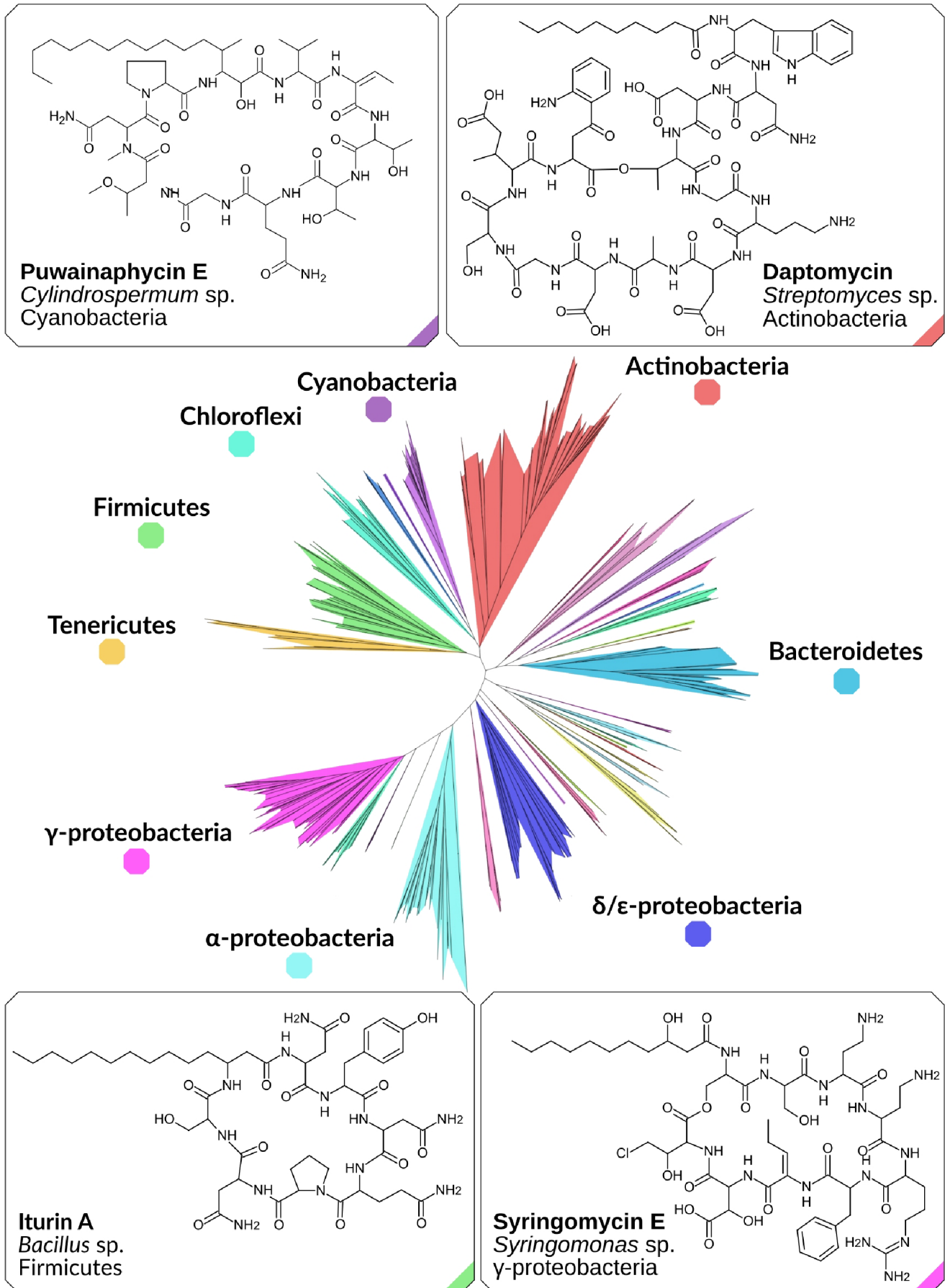


**Fig. 2: Examples of cyclic and linear bacterial lipopeptides.** Fatty acid-derived moiety (FA) is highlighted in red.

### **1.4.1. Ecophysiological Roles of Bacterial Lipopeptides**

Production of secondary metabolites is a widespread phenomenon in the prokaryotic domain of life. The secondary metabolites of my interest, the lipopeptides, have been frequently found in representatives of firmicutes (*Bacillus*, *Paenibacillus* - surfactin, iturin; Cochrane and Vederas 2016), actinobacteria (*Streptomyces* – daptomycin; Miao et al 2005 ) and  $\gamma$ -proteoacteria (*Pseudomonas* – syringafactin, Berti et al 2007; *Marinobacter* – marinobactins, Martinez et al 2000) but they are also representatives from  $\beta$ -proteobacteria (*Cupriavidus* – cupriachelin, Kreutzer et al 2012; *Paraburkholderia* – gramibactin, Hermenau et al 2018),  $\delta$ -proteobacteria (*Stigmatella* – myxochromides, Wenzel et al 2005) and  $\alpha$ -proteobacteria (*Thalassospira* – thalassospiramide, Oh et al 2007; Fig. 3.).

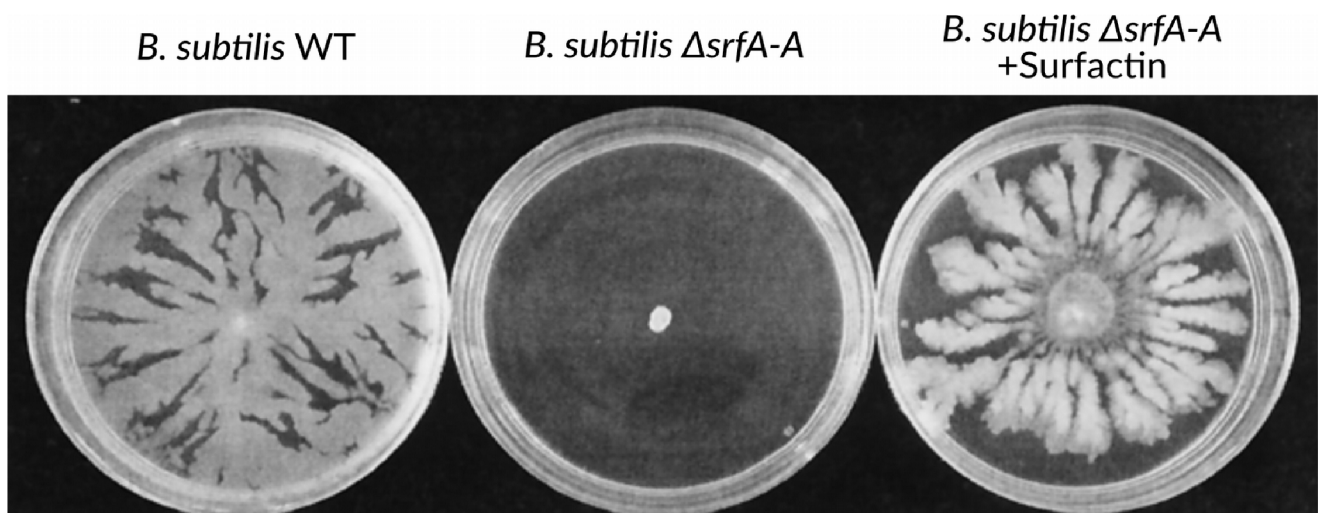
The vast majority of described lipopeptides were not investigated for the benefit they provide to the producing strain yet the ones that were offer valuable insight into the exciting world of microbial interactions.



**Fig. 3: Examples of lipopeptides found across the bacterial kingdom. Phylogenomic tree as in Khaledian et al 2020.**

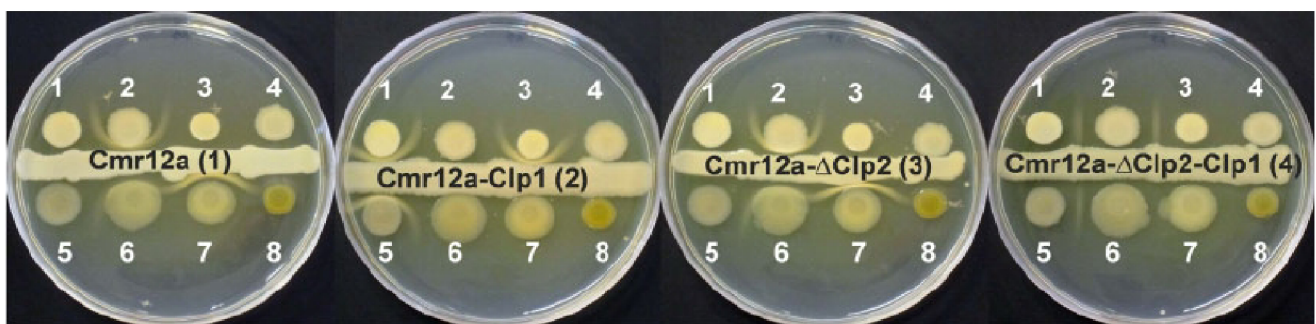
### ***Role in Colony Development and Spreading***

Probably the most notoriously known lipopeptide with an extensively described role in colony development and microbial communication is surfactin (Fig. 4), produced by various representatives of the genus *Bacillus*, a gram-positive endospore-forming representative of firmicutes. Surfactin is a  $\beta$ -hydroxy-fatty acid-containing cyclic heptapeptide that due to its amphiphilic nature is able to reduce surface tension of aqueous solutions and form pores in biological membranes. These key features dispose surfactin to enhance colony growth over new surfaces and to induce potassium ion fluxes in the neighboring cells that through a signaling cascade trigger the production of extracellular matrix and orchestrate the biofilm development (Kinsinger et al 2003, Lopez et al 2009, van Gestel et al 2015). Moreover, it was found that surfactin mediates interactions between its producer and other microbes including other firmicutes, actinomycetes or  $\gamma$ -proteobacteria (Straight et al 2006, Luzzato-Knaan et al 2016). For example, surfactin is thought to facilitate the engulfment of competing colonies of *B. simplex* by *B. subtilis* (Rosenberg et al 2016) or to suppress the development of aerial hyphae of *Streptomyces* (Straight et al 2006).



**Fig. 4: Role of surfactin in colony development of *B. subtilis*.** Surfactin is essential to colony morphology and spreading of *B. subtilis*. Image from Kinsinger et al 2003. On the left plate: growth of wild-type *B. subtilis* (Marburg, 6051); on the middle plate: disruption of *srfA-A* gene in the original strain ( $\Delta$ srfA-A) causes surfactin deficiency and results in impaired colony growth; on the right: the external addition of surfactin rescues the phenotype. For detailed cultivation conditions see the original publication.

Further important insight into biological activity of lipopeptides was provided by the studies of various strains of *Pseudomonas*. This genus of  $\gamma$ -proteobacteria is found to produce a plethora of different lipopeptides among others cichofatins, syringafactins, viscosin, massetolides and more. As was recently reviewed by Götze and Stallforth (Götze and Stallforth 2020) some of these were found to participate in a wide array of motility-related processes, promote swarming motility or, conversely, to suppress it and support the formation of biofilms (see table 1 in Götze and Stallforth 2020). For example, the swarming motility of *Pseudomonas syringae* is facilitated by a linear lipopeptide syringafactin (Berti et al 2007), the motility of *P. cichori* depends on cichafactin (Pauwelyn et al 2013) and the swarming of *P. poae* depends on poeamide (Zachow et al 2015). Often in the absence of lipopeptides facilitating swarming motility, such as via experimental disruption of their biosynthesis, a complementary strategy of biofilm formation is promoted instead (Berti et al 2007, Pauwelyn et al 2013). In contrast, tolaasin, sessilin and massetolide A were found to suppress the swarming motility in their respective strains in favor of biofilm development (D'aes et al 2014). Finally, tolaasin and related sessilin physically interact with the swarming-promoting orfamide family of lipopeptides to form precipitates, demonstrated on agar plates as distinctive white lines, to stop the spread of potentially competing strains (Fig. 5; D'aes et al 2014, Henkels et al 2014).



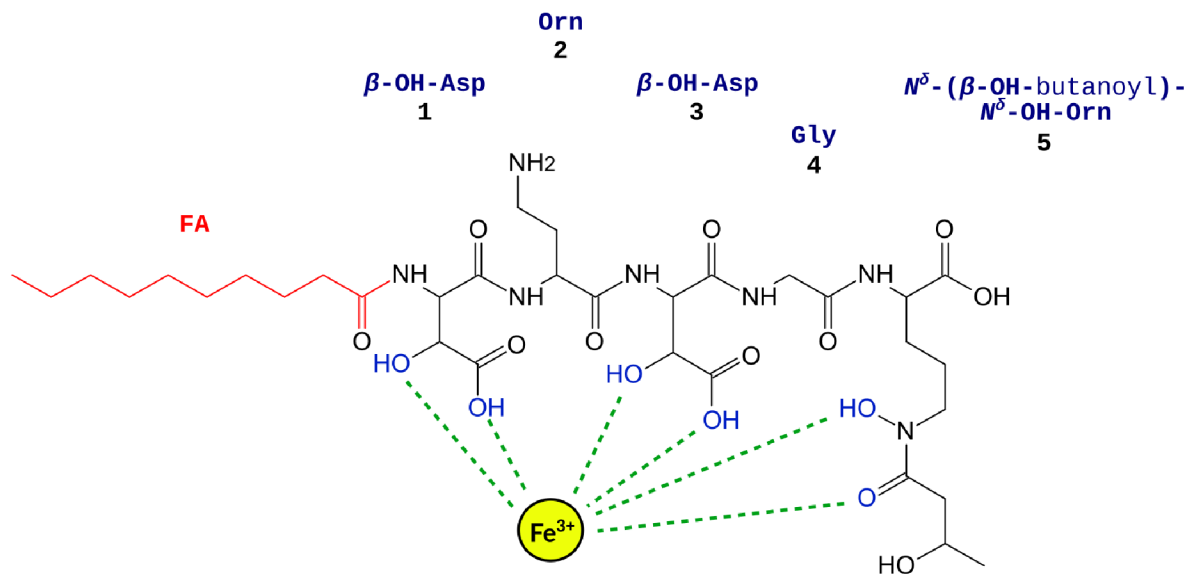
**Fig. 5: Interactions between *Pseudomonas* strains producing different lipopeptides.** Antagonistic lipopeptides can precipitate each other and form distinctive white lines between the producing strains upon cultivation on agar plates. Image taken from D'aes et al 2014. Strains used: (1) *Pseudomonas* sp. CMR12a (orfamide and sessilin); (2) CMR12a-Clp1 (orfamide); (3) CMR12a- $\Delta$ Clp2 (sessilin); (4) CMR12a- $\Delta$ Clp2-Clp1; (5) *P. tolaasii* NCPPB 2192 (tolaasin); (6) *P. 'reactans'* NCPPB 387 (WLIP); (7) *P. protegens* Pf-5 (orfamide); (8) *P. fluorescens* R1SS101 (massetolide). For more information on strain origin and cultivation condition see the original study.

As can be seen from the provided examples, lipopeptides participate in a complex interplay of antagonistic signaling pathways determining the motility patterns of pseudomonads (D'aes et al 2014, Götze and Stallforth 2020). The pseudomonal motility patterns are then important for shaping the relationship between the producing bacterium and the various plants they colonize. The relationship that may be of mutual benefit or for the misery of the plant. This has considerable ecological and agricultural impact beyond the scope of this text.

### ***Role in Iron Scavenging***

Several lipopeptides were observed to be involved in the facilitation of iron uptake (Kem and Butler 2015). Iron is an important micro-nutrient, essential to many components of respiratory chain and photosynthetic machinery (Sandy and Butler 2009). It is the fourth most abundant element in Earth's crust; however, it is often found in its oxidized  $\text{Fe}^{3+}$  form, which forms insoluble precipitates that complicate its uptake by living organisms. Furthermore, sedimentation of these precipitates in the water column causes iron depletion in the pelagic of aquatic environments. To cope with these complications, microorganisms produce compounds, siderophores, which facilitate the uptake of iron and prevent its oxidation and precipitation. Siderophores bind iron atoms via coordination of electron pairs in its free d-orbitals. Such a bond requires several properly positioned donors of electron pairs (Fig. 6); it is reversible and frequently demonstrated as a characteristic UV-Vis absorption spectra of the complexes. To bind, siderophores most frequently employ catecholates,  $\alpha$ -hydroxycarboxylates and hydroxamates (Sandy and Butler 2009). Other residues, such as *N*-nitrosohydroxylamines or a particular arrangement of oxazole and phenolate, are also employed, although less frequently (Hermenau et al 2018, Sritharan 2016). Interestingly, the  $\alpha$ -hydroxycarboxylate residue, found in many siderophores as  $\beta$ -hydroxyaspartate ( $\beta$ -OH-Asp), provides additional functionality. The siderophore-bound ferric iron is reduced upon exposure to UV-light as the siderophore is cleaved close to the  $\beta$ -OH-Asp residue (Kreutzer et al 2012, Hardy and Butler 2019).





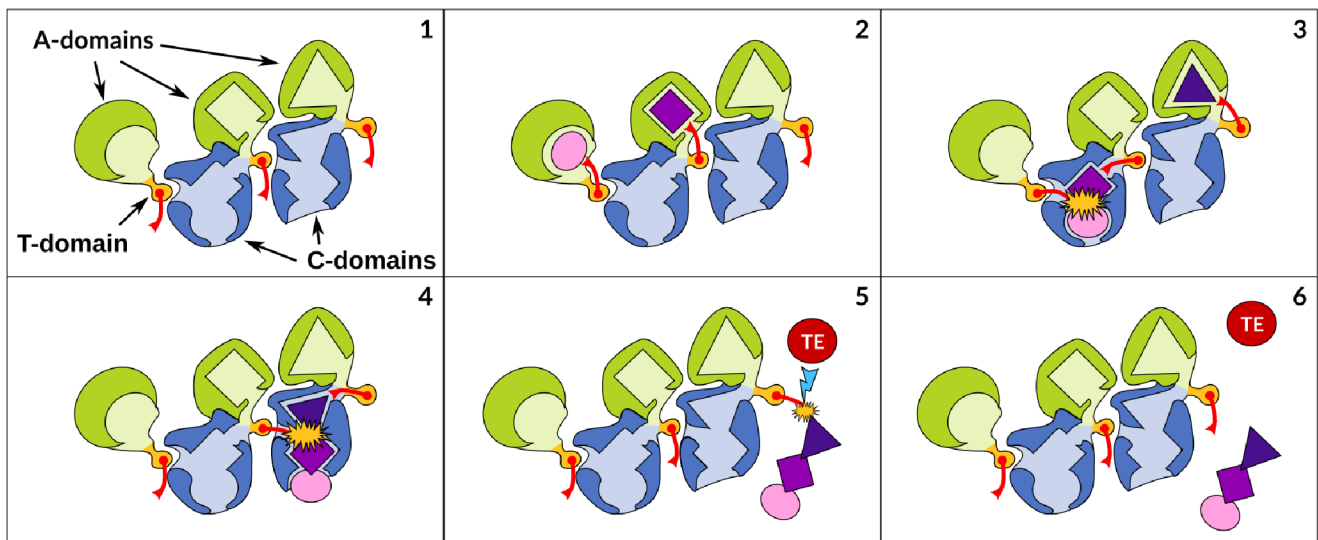
### Cupriachelin *Cupriavidus* sp.

**Fig. 6: Example of lipopeptide siderophore cupriachelin.** Two  $\beta$ -OH-Asp residues ( $\alpha$ -hydroxycarboxylate) and an  $N^{\delta}$ -( $\beta$ -OH-butanoyl)- $N^{\delta}$ -hydroxy-ornithine (hydroxamate) residue are employed for chelation of trivalent iron. Hydroxyl residues involved in chelation are highlighted in blue. Fatty acid-derived moiety is highlighted in red)

Several lipopeptides, e. g. aquachelins and marinobactins (Martinez et al 2000), cupriachelin (Kreutzer et al 2012), mycobactin (Sriharan 2016) and gramibactin (Hermenau et al 2018), contain residues capable of binding iron and were shown to do so. The prolonged hydrocarbon chain of lipopeptides is not essential for the iron chelation. However, it seems to be crucial for adjusting the distribution and diffusivity of the compounds in the natural context, where excessive diffusion can lead to efficient loss of siderophore to the producer in case it diffuses too far (Völker and Wolf-Gladrow 1999). Cleavage of the acyl residue as a possible mechanism for conversion to more diffusive variants has been observed in marinobactins, where in the late stationary phase of an iron-limited culture the bacterium excreted an acylase to cleave off the acyl residue to increase its diffusion rate (Gauglitz et al 2014).

### 1.4.2. Biosynthesis of Lipopeptides

Bioactive peptides found across a wide array of microorganisms attracted the attention of many researchers due to a frequent presence of amino acids that are not recognized by the conventional proteosynthetic machinery. Investigation of this phenomenon led to the discovery of non-ribosomal peptide synthesis (NRPS) and the identification of its core enzymes and mechanism of action (Stachelhaus and Marahiel 1995, Stachelhaus et al 1996, Fig. 7). Essential to NRPS are three domains: the adenylation domain (A-domain), thiolation domain (T-domain) and condensation domain (C-domain).



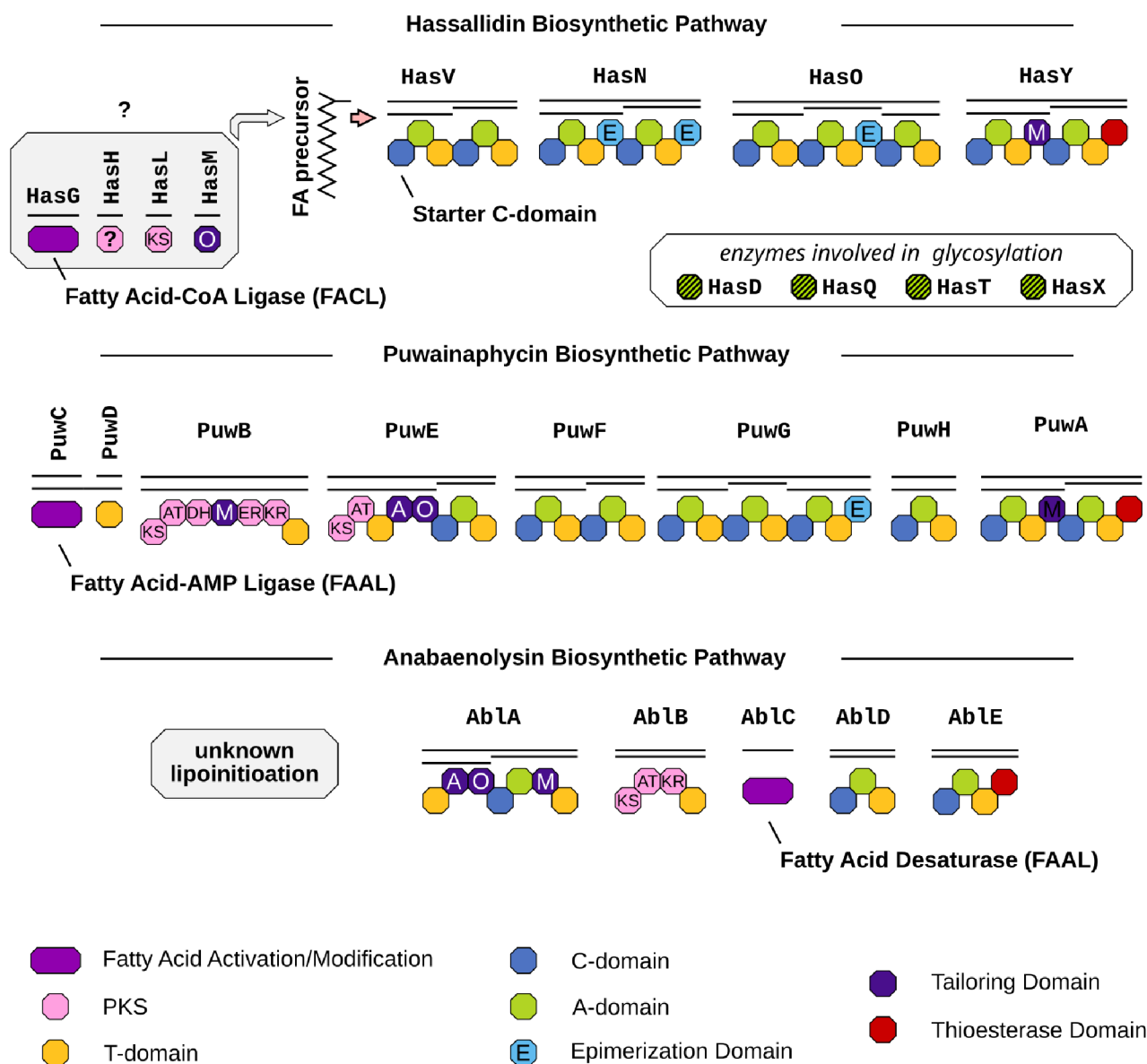
**Fig. 7: Schematic depiction of NRPS synthesis.** 1 - Simplified NRPS assembly line. 2 - Amino acids are activated by corresponding A-domains and bond to 4'-phosphopantetheinyl of T-domains. 3 - The 4'-phosphopantetheinyl-bound amino acids meet at the C-domain where the peptide bond between them is formed, while another amino acid is being activated by the next A-domain. 4 - Nascent peptide is prolonged by the amino acid. 5,6 - The product is eventually released from the assembly line by a thioesterase.

The A-domain's primary function is to activate specific amino acid(s) and catalyze the formation of amino acid-adenylate (Substrate + ATP → Substrate-AMP + PP<sub>i</sub>; Stachelhaus et al 1999). The activated molecule is then passed to the T-domain, where it is bound to the thiol functional group of 4'-phosphopantetheinyl residue, which acts as a swinging arm and is responsible for the transport of the intermediate among domains (Stachelhaus et al 1996, Weber et al 2000). The 4'-phosphopantetheinyl attached to a conserved serine residue of T-domains is one of the essential features shared by NRPS, PKS and fatty acid biosynthetic pathways and

that enables passing of intermediates among these different biosynthetic machineries (Lambalot et al 1996, Stachelhaus et al 1996). The activated PAN-bound intermediates meet deep in the solvent channel of the NRPS C-domain, where the amino group of the downstream intermediate acts as a nucleophile and attacks the thioester bond of the upstream intermediate, thus forming a peptide bond (Keating et al 2002).

The three core domains are organized in modules that are assembled in a larger complex where the nascent peptide passes from module to module. Apart from the core domains, the module may include additional tailoring domains that modify the nascent product or activated monomer. Such modifications include amination, methylation, hydroxylation, dehydration, halogenation, epimerization, heterocyclization and more, with some yet to be discovered (Hermenau et al 2018, Kreutzer et al 2012). Interestingly, the C-domain can reject to fuse the improper or improperly modified residue to the nascent product and hence provide additional layer of proof-reading activity (Kreutzer et al 2012, Reitz et al 2019, Rausch et al 2007).

An essential feature of lipopeptides is the presence of the hydrophobic moiety, which may be biosynthetically achieved in two described ways (Fig. 8). A condensation of an acyl-CoA recognized by a specialized C-starter domain to the following amino acid of the prospective lipopeptide is seen in the biosynthesis of surfactin, lichenysin, fengycin, and arthrofactin (Rausch et al 2007). Second is the activation of fatty acid by a fatty acid-AMP ligase (FAAL) and its subsequent binding to a T-domain, in a way highly similar to that of A-domains, as seen in the biosynthesis of for example mycosubtilin or daptomycin (Duitman et al 1999, Miao et al 2005). In some cases, evidence suggests that neither of the two mechanisms is employed. For example, in the case of anabaenolysins the information available for the biosynthetic pathway does not support employment of neither of the two known mechanisms, yet the peptide is produced with an acyl residue of considerable size. Such cases point to the existence of additional mechanisms yet to be discovered (Shishido et al 2015).

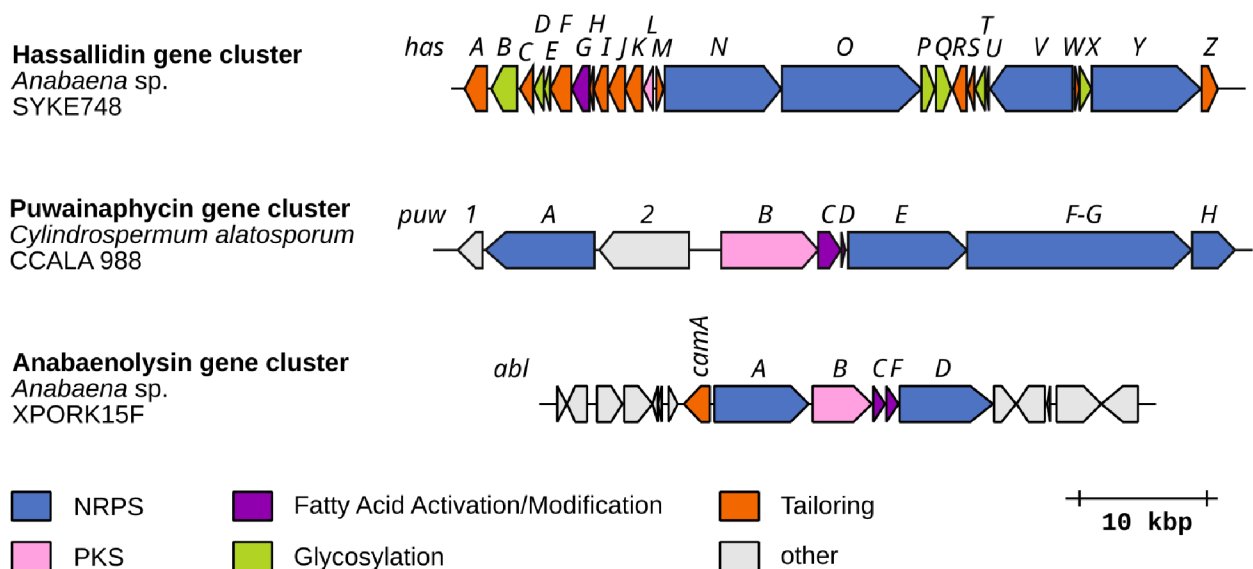


**Fig. 8: Examples of different ways of lipoinitiation in the biosynthesis of cyanobacterial lipopeptides.** For PKS domain the following abbreviations are used: KS - ketosynthase, AT - acyltransferase, DH - dehydrogenase, ER - enoylreductase, KR - ketoreductase. For tailoring domains the following abbreviations are used: A - aminotransferase, O - oxygenase, M - methyltransferase.

Pathways for the biosynthesis of polyketides are organized in a way analogous to that of NRPS. Polyketide synthases (PKS) are modular enzymes that also employ the 4'-phosphopantetheinyl arm for the secured passage of intermediates. However, while being analogous to the T-domain, the domain that binds the 4'-phosphopantetheinyl in the context of PKS modules is designated as ACP (acyl carrier protein) instead. Activation of monomers, frequently a malonyl or a methyl-malonyl residue, is done by acyltransferase and the formation of C-C bond is performed by a

ketosynthase. Elements of pathways for polyketide synthesis (PKS) are frequently combined with NRPS resulting in increased structural diversity of secondary metabolites. In biosynthesis of lipopeptides there are most frequently found to elongate the fatty acyl moiety. Such elongation, however, is not just for the sake of the length of the hydrocarbon chain. The PKS-mediated elongation allows crucial modifications of the acyl chain, such as hydroxylation, amination or eventually cyclization into an aromatic ring. For example, during the biosynthesis of the lipopeptide mycosubtilin, the activated fatty acid is prolonged by PKS modules that introduce an amino residue to the  $\beta$ -carbon of the final N-terminal hydrophobic moiety (Duitman et al 1999). The presence of the  $\beta$ -amino group allows for cyclization of the lipopeptide at the end of its synthesis (Duitman et al 1999, Mareš et al 2014).

Each step in the biosynthesis of lipopeptides is performed by a dedicated protein or a protein domain within a large multi-domain protein typical for NRPS biosynthesis. Nevertheless, each protein domain responsible for each biosynthetic step has to be encoded in the genome. Luckily, in prokaryotes, genes for a biosynthetic pathway are found together in the so-called biosynthetic gene clusters. Organization of such gene clusters does not necessarily reflect the order of biosynthetic steps included, yet it can provide valuable insight into the composition of the encoded product.



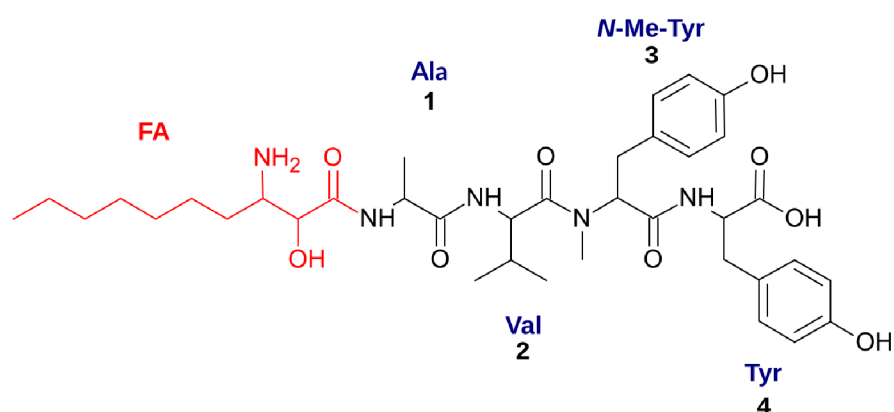
**Fig. 9: Examples of biosynthetic gene clusters.**

In recent years many researchers took advantage of accumulating sequences of microbial genomes to search for such gene clusters. Bioinformatic software dedicated to discovery and analysis of biosynthetic gene clusters, such as AntiSmash (Blin et al 2019) or Prism (Skinnider et al 2020), has been extensively developed and its accuracy is still gradually increasing.

### **1.4.3. *Cyanobacterial Lipopeptides***

There are many cyanobacterial lipopeptides that have been isolated from mixed material collected in the field and described briefly, with elucidated structure and assessment of bioactivity, but without hints on their biosynthesis, mechanism of action or even details of their producer. While this approach certainly brings valuable information on the diversity and pharmaceutical potential, it is less informative on the importance of the compounds for their producers and evolutionary relationships between them. Another complication for the systematic approach to natural compounds is that the authors often favor giving a new name to a compound analogous to another, already described one.

In the following section I will outline several groups of cyanobacterial metabolites that were studied in a considerable detail and for which the biosynthetic pathway was suggested along with a corresponding gene cluster.



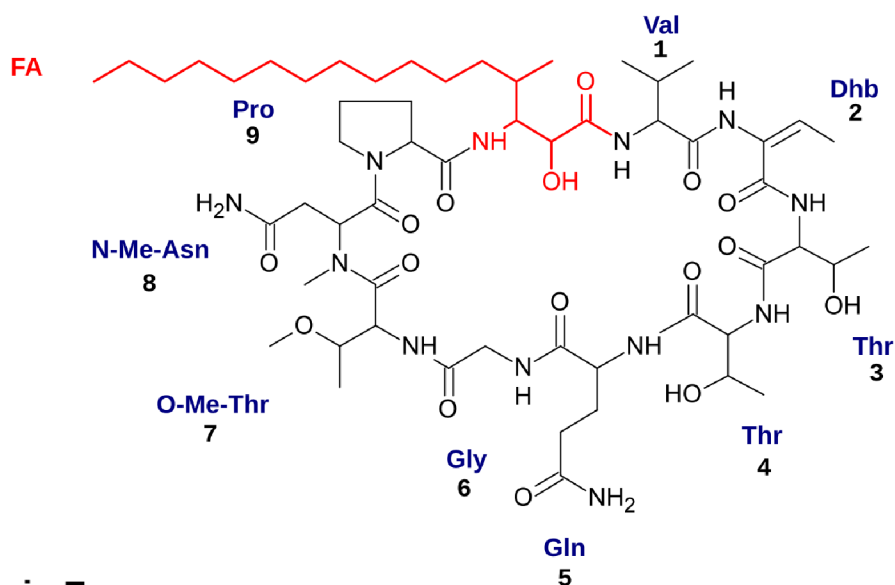
**Microginin**  
*Microcystis aeruginosa*

**Fig. 10: Structure of microginin.** Modified fatty acid residue (FA) highlighted in red.

### **Microginins**

Linear lipopeptides have been repeatedly isolated mostly from freshwater bloom-forming cyanobacteria of genus *Microcystis*, but also from *Nostoc* or *Oscillatoria* (Okino et al 1993, Ishida et al 2000, Strangman and Wright 2016, Stewart et al 2018). Microginins consist of 3-5 amino acids and an N-terminal  $\beta$ -amino,  $\alpha$ -hydroxy C8 or C10 carboxylic acid (Fig. 10, above). The fatty acid may be further modified by mono- or dichlorination on the  $\omega$ -carbon (Ishida et al 2000, Strangman and Wright 2016). Various amino acids have been found in microginins and there is no strict amino acid sequence to distinguish them. It is worth noting that tyrosine is remarkably frequent especially on the C-terminus and that *N*-methylation of amino acids is a regular modification (Ishida et al 2000). The biosynthesis of microginins has been proposed in a patented article by Kramer to be a hybrid PKS/NRPS type (Kramer 2010).



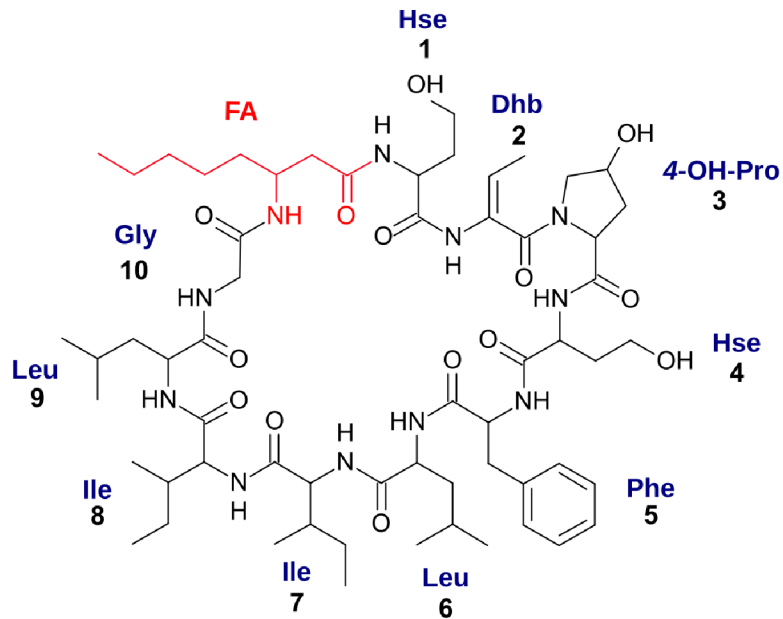


**Puwainaphycin E**  
*Anabaena* sp.

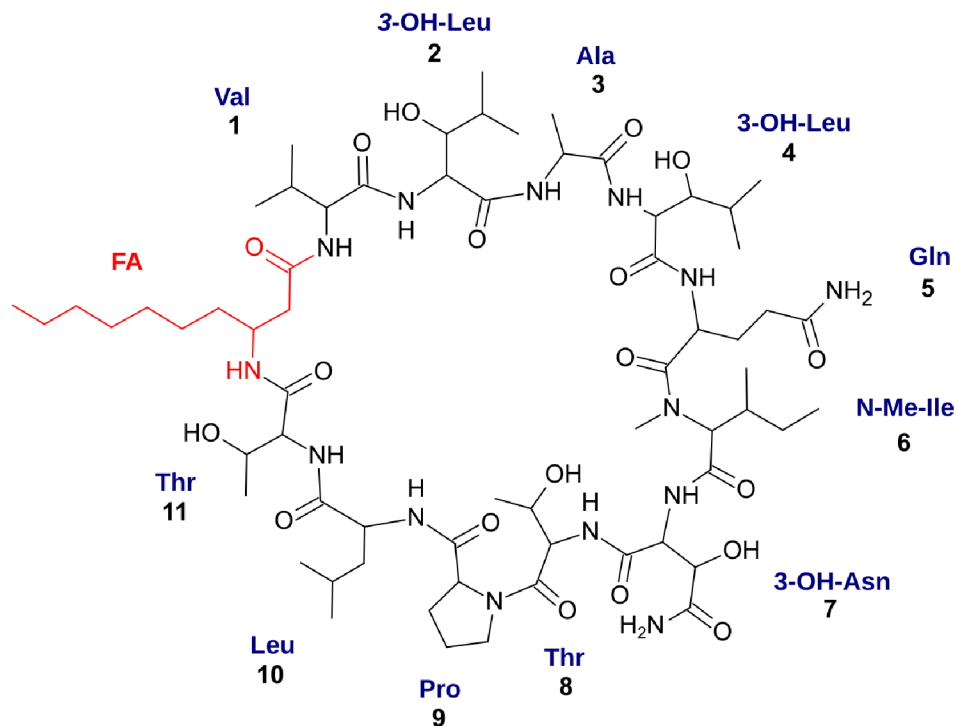
**Fig. 11: Structure of puwainaphycin E.** Modified fatty acid residue (FA) highlighted in red.

***Puwainaphycins and Minutissamides***

Structurally analogous puwainaphycins A-G (PUW) and minutissamides A-L (MIN) are amphipathic lipopeptides circularized via the formation of a peptide bond between the  $\beta$ -amino residue of the fatty acid and the C-terminal carboxyl residue (Fig. 11, Moore et al 1989, Gregson et al 1992, Kang et al 2011, Kang et al 2012, Hrouzek et al 2012, Mareš et al 2014). The 10-membered peptide ring, including the modified fatty acid residue, exhibits a degree of conservation, especially around the place of integration of the fatty acid into to the cycle, since, in the described puwainaphycins, the amino acids 1, 2, 8 and 9 are invariant. In contrast, the fatty acid-derived residue is considerably more variable: its length ranges from C10 to C18 and it can contain further modification such as hydroxylation, oxosubstitution or halogenation (Mareš et al 2014). Puwainaphycins and minutissamides are thought to be synthesized by the same hybrid PKS/NRPS pathway encoded by a ~57 kb long gene cluster (Mareš et al 2014). Fatty acid-AMP ligase (FAAL) is thought to activate several fatty acids with different lengths that are subsequently prolonged and modified by two PKS modules. An important modification performed by the second PKS module is the introduction of the  $\beta$ -amino residue later employed for cyclization. Puwainaphycins and minutissamides were reported from genera *Anabaena* and *Cylindrospermum* (Mareš et al 2014).



**Laxaphycin A**  
*Anabaena laxa*



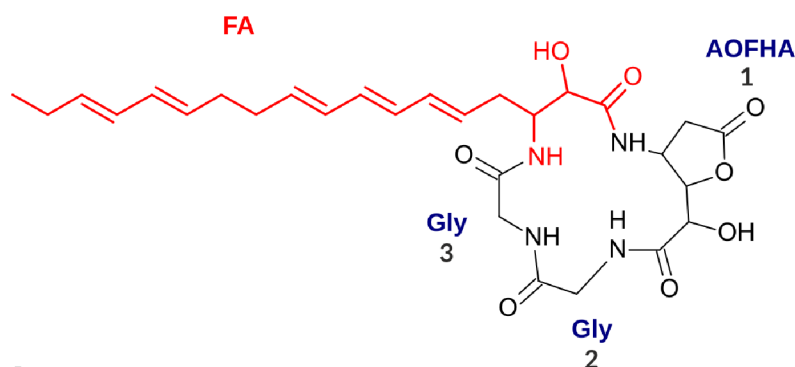
**Laxaphycin B**  
*Anabaena laxa*

**Fig. 12: Structure of A and B type laxaphycins.** Modified fatty acid residue (FA) highlighted in red.

### ***Laxaphycins***

Produced foremostly by *Anabaena laxa*, the laxaphycins (Fig. 12) are cyclic lipopeptides circularized via a peptide bond between C-terminal carboxyl residue and the  $\beta$ -amino residue of the fatty acid (Frankmölle et al. 1992). The laxaphycins show little variability in the hydrophobic moiety

and employ either a  $\beta$ -aminooctanoic or a  $\beta$ -aminodecanoic acid. However, they exhibit a considerable variability in the amino acid composition. Based on the number of amino acids employed, two types of laxaphycins are recognized. The laxaphycin A employs 10 amino acids and together with the N-terminal fatty acid forms a 11-membered ring while the laxaphycin B employs an additional amino acid and forms a 12-membered ring (Frankmölle et al. 1992; Luo et al., 2015). The two peptide cores differ considerably in the amino acid sequence, yet both types occur simultaneously and are thought to be synthesized by a shared unusually branched hybrid PKS/NRPS biosynthetic pathway that employs FAAL for lipoinitiation and is encoded by a common 96-kb gene cluster (Heinilä et al 2020). Moreover, the observation that the antifungal and antiproliferative effects of laxaphycins are strengthened if both types are provided, suggests a functional co-evolution of laxaphycin A and B (Cai et al 2018). Shall the classification of natural compounds be based strictly on structural features, laxaphycin A and B type should be considered separately; however, their shared biosynthetic pathway, occurrence and likely a functional co-evolution, could justify their shared classification from a biological perspective.



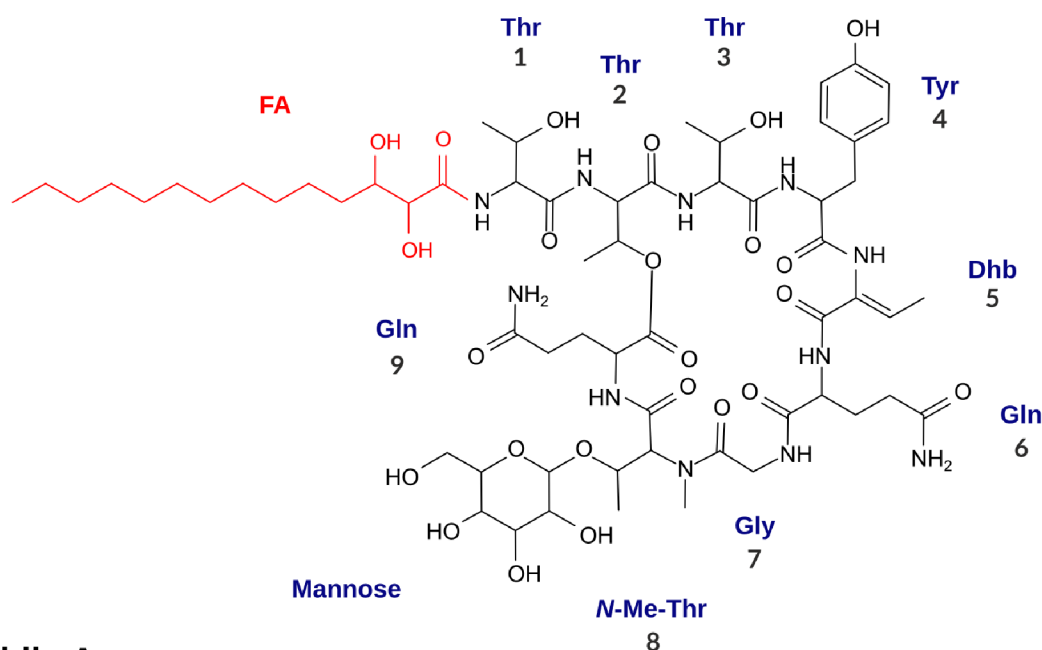
**Anabaenolysin A**  
*Anabaena* sp.

**Fig. 13: Structure of anabaenolysin A.** Modified fatty acid residue (FA) highlighted in red.

### ***Anabaenolysins***

Multiple strains of *Anabaena* isolated from benthos of the Baltic Sea were found to produce anabaenolysins, rather small lipopeptides with a distinctive hydrophobic residue (Fig. 13). Anabaenolysins A-G feature rare unsaturated  $\beta$ -amino fatty acid with a conjugated triene system that is incorporated into a 4-membered peptide ring containing another unusual

structural residue, the 2-(3-amino-5-oxytetrahydrofuran-2-yl)-2-hydroxyacetic acid moiety and two glycines, alternatively with the last glycine substituted by a glutamine (Jokela et al 2012, Shishido et al 2015). The variants differ in the length of the fatty acid (C13-C19) and the position of the double bonds on the extended hydrocarbon chain. A compact 23-kb PKS/NRPS gene cluster is considered responsible for anabaenolysin biosynthesis. The cluster encodes enzymatic machinery for elaboration of the fatty acid, including a PKS module for the introduction of the  $\beta$ -amino residue for cyclization but notably also the fatty acid desaturase held responsible for the formation of double bonds on the hydrocarbon chain (Shishido et al 2015). However, the cluster does not include any genes that could explain the lipoinitiation, hence the first step of anabaenolysin synthesis remains an open question.



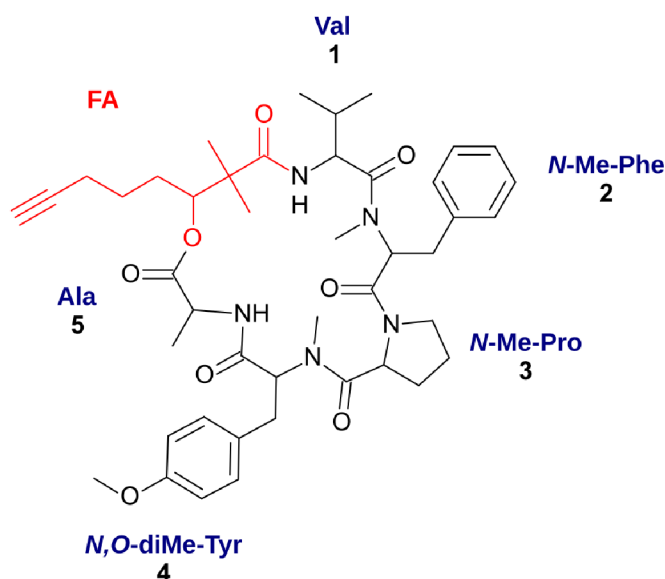
**Hassallidin A**  
*Hassallia* sp.

**Fig. 14: Structure of hassallidin A.** Modified fatty acid residue (FA) highlighted in red.

### ***Hassallidins and Baltacidins***

The most structurally complex lipopeptides isolated from cyanobacteria so far are the hassallidins that in addition to the peptide core and an exocyclic hydrophobic moiety are also glycosylated (Fig. 14, Neuhof et al 2005). Baltacidins A-D share most of the structural features including the majority of the peptide core and could be efficiently considered to belong to the same family (Bui et al 2014, Vestola et al 2014). The peptide core of

hassallidins is largely conserved with notable variation of the fourth amino acid, which could be tyrosine or hydroxy-tyrosine, and the last amino acid (9th) that could be either glutamine, lysine or tyrosine. The hydrophobic moiety consists of a C14 or C16 hydrocarbon chain with two hydroxyl groups and in the case of balticidins A and B, it is also chlorinated. The hydroxyl group of the conserved *N*-methyl-threonine (8th amino acid) and on the  $\beta$ -carbon of the fatty acid offer spots for glycosylation that is the main source of hassallidin structural variability. The  $\beta$ -hydroxyl of the fatty acid can be glycosylated by either rhamnose or arabinose, with arabinose possibly further linked to *N*-acetyl glucose or galacturonic acid. The *N*-Me-Thr at position 8 can be linked to either mannose, diacetylmannose or hexose. The 48-59 kb hassallidin gene cluster encodes a biosynthetic pathway with 9 NRPS modules, several glycosyl transferases responsible for incorporation of sugars and a cassette for synthesis of a modified fatty acid precursor. A specific starter C-domain ensures the recognition of the modified fatty acid and its fusion to the first amino acid of the nascent product (Vestola et al 2014, Pancrace et al 2017). The hassallidins are found in various genera, i.e. *Hassallia*, *Tolypothrix*, *Aphanizomenon*, *Cylindrospermopsis*, *Nostoc*, *Planktothrix* and *Anabaena*.



**Dudawalamide B**  
*Moorea producens*

**Fig. 15: Structure of dudawalamide B, a representative of kulolide superfamily of lipopeptides.** Modified fatty acid residue (FA) highlighted in red.

***The Kulolide Superfamily***

Another interesting, rather broad group of compounds isolated mostly from field samples of marine cyanobacteria are the lipopeptides of the kulolide superfamily. The family contains kulolides, dudawalamides (Fig. 15), pitipeptolides, viequeamides, wewakpeptin A (Almaliti et al 2017, Boudreau et al 2012) and eventually could include hantupeptins, palmyramides, trungapeptins, hectochlorine and some lyngbyabellins (A-C, G; Tripathi et al 2010, Taniguchi et al 2010, Bunyajetpong et al 2006, Han et al 2005, Nagarajan et al 2012). The members of the kulolide superfamily have a loosely similar scaffold, but often share important features of the hydrophobic moiety. Frequently the hydrophobic chain is 8 carbons long, methylated on the  $\alpha$ -carbon, hydroxylated on the  $\beta$ -carbon and has a triple bond or is chlorinated at  $\omega-1$ . The  $\beta$ -hydroxyl is also the spot of cyclization. Interestingly, in the kulolide superfamily the residue next to the modified fatty acid may be an amino acid but also frequently an  $\alpha$ -hydroxy carboxylic acid, hence employing an ester instead of a peptide bond. A biosynthetic pathway was proposed for hectochlorine and although it is not clear whether to consider hectochlorine a member of the kulolide family, it definitely features fatty acid modifications typical for the kulolide

family. Nevertheless, FAAL is employed to fuse the recognized fatty acid to a T-domain and the fatty acid is passed to a PKS module. The fatty acid is prolonged, methylated and the keto residue on the  $\beta$ -carbon is reduced to hydroxyl to allow cyclization of the product. The rest of the pathway includes several NRPS modules that would vary for the individual members of the kulolide family (Ramaswamy et al 2007). The compound from the kulolide superfamily were isolated from marine cyanobacteria, mostly *Lyngbya* and *Moorea*, from various locations in the Pacific, Indian and Atlantic oceans and are likely considerably widespread (Boudreau et al 2012).

### ***Other Lipopeptides***

The above provided list consists of prominent classes of cyanobacterial lipopeptides for which there are also data on their biosynthesis. There are additional cyanobacterial lipopeptides fitting the most consensual definition of a lipopeptide, such as microcolins, almiramides, dragonamides and muscotoxins (Yu et al 2019, Sanchez et al 2010, Tomek et al 2015). However, little or no information on their biosynthesis has been published so far.

It is worth noting that a relevant, rather broader understanding, considers peptides with aromatic branched or heavily modified hydrocarbon moieties as lipopeptides. Such definition would also include the microcystins, the infamous freshwater toxins with an N-terminal hydrophobic moiety that includes an aromatic ring on the distant end of the hydrocarbon chain (Dittmann et al 2013). Similarly, nodularins, nostophycins and eventually several aeruginosins and spumigins would fit in the definition (Welker and von Döhren 2006, Fewer et al 2011). Furthermore, there are compounds with a branched hydrophobic residue such as the pahayokolides (An et al 2007).

Another edge of the definition is the minimal length of a hydrocarbon chain that would classify the final structure as a lipopeptide. There are peptides with modified carboxy acid at the N-terminus that is considerably short. A good example is notable cyanobacterial peptide, the nostopeptolide, where the butyric acid is the N-terminal residue (Golakoti et al 2000, Liaimer et al 2015). Similarly, the cyanopeptolins contain the N-terminal hexanoic acid and the malevamide A contain  $\beta$ -amino-A-methyl

hexanoic acid (Welker and von Döhren 2006, Horgen et al 2000). One of the key features of the lipopeptides is their amphipathic nature and short-chain fatty acid are not more hydrophobic than amino acid such as leucine. Employment of such short fatty acids in peptides is undoubtedly interesting, yet the compounds are not considered as lipopeptides.

Finally, the peptides are recognized by the presence of a distinct bond. There are compounds with extended polyketide moieties that dominate the overall structure, however, they are synthesized through intermediates with obvious peptide bonds that are subsequently modified in such a way that the bonds lose their peptide nature. Cyanobacteria produce compounds such as barbamide, curacin A, jamaicamide, lynbyatoxin or somocystinamide that have few peptide bonds and would fit the broad definition of lipopeptide if the presence of a hydrophobic moiety and a peptide bond were the only criteria (Edwards and Gerwick 2004, Edwards et al 2004, Wrasidlo et al 2008, Nogle and Gerwick 2002). However, such compounds are considered a mixed-polyketide peptides rather than lipopeptides. Regardless of their final structure, the biosynthesis of these compounds shares key enzymatic activities with the synthesis of lipopeptides and provides insight into the exciting tailoring potential of hybrid biosynthetic pathways (Chang et al 2002, Chang et al 2004, Edwards and Gerwick 2004, Edwards et al 2004, Ramaswamy et al 2007). In the present study, the peptides with aromatic or branched hydrophobic moiety will not be reviewed, neither will be the peptides with incorporated short fatty acid (C6 and shorter) or the mixed-polyketide peptides.



## **1.5. Summary of the Research Background and Rationale**

Cyanobacteria are ancient prokaryotic organisms with a considerable global impact and important roles in various habitats. The phylum comprehends various species with a remarkable degree of morphological and metabolic plasticity. Cyanobacteria can live in a close association with other organisms in microbial mats or in symbioses with algae or plants and even nowadays remain an influential group of phototrophs.

Lipopeptides are structurally diverse secondary metabolites frequently found to be produced by various cyanobacteria. The essential features of lipopeptides, especially their amphiphilicity that gives the ability to reduce surface tension of aqueous solutions or form pores in biological membranes, have made them important factors in the colonization of novel surfaces and biofilm formation in other known bacteria. Moreover, the lipopeptides were found to mediate interactions within bacterial colonies, microbial communities or eventually between bacteria and plants. Many lipopeptides were isolated from cyanobacteria, but the relationship between lipopeptides and the sessile vs. buoyant lifestyle was not investigated yet, neither was their role in plant-cyanobacteria symbioses.

Finally, in the recent decade, a considerable amount of cyanobacterial genomes has been analyzed and the sequences are publicly available. Frequently, the data do not come from studies dedicated to secondary metabolism yet with the available bioinformatic tools they offer a great opportunity to discover novel cyanobacterial metabolites. Perhaps even more importantly, it is an opportunity to gain insight into lipopeptide distribution across the phylum and their mode of action and natural roles in cyanobacteria.

## **2. Results and Discussion**

### **2.1. Outline of the Research**

The thesis revolves around different aspects of cyanobacterial lipopeptides. The first paper investigates the fragmentation of lipopeptides upon collision induced dissociation (CID) during mass spectroscopic analysis to facilitate the detection of lipopeptides in natural isolates or the identification of novel lipopeptides from cyanobacterial cultures. The second study takes the advantage of the publicly available genomes and the available bioinformatics software to explore the distribution of biosynthetic gene clusters possibly encoding lipopeptide biosynthesis across the cyanobacterial phylum. The third paper investigates in detail the different organization of biosynthetic gene clusters encoding the synthesis of puwainaphycins. The fourth study elaborates on the observation from the second study, where a particular pattern of NRPS and tailoring genes was frequently found in the analyzed genomes of phylogenetically distant cyanobacteria. We decided to investigate what may be the structural feature synthesized by this recurring pattern and what functional implication it may have for the produced lipopeptide. The fifth and final study is in the form of a manuscript and studies the origin of  $\gamma$ -methyl proline (4-Me-Pro) in muscotoxin B. The results suggest an intermediate leak from a nostopeptolide biosynthetic machinery that is utilized by the muscotoxin biosynthetic pathway to produce an additional variant.

## 2.2. Research Paper I. - Fragmentation of Lipopeptides Under High Collision Energy

An important structural feature of lipopeptides, responsible for their amphiphilic nature, is the more or less modified acyl residue. The study investigates the possibilities offered by high resolution mass spectrometry (HRMS), especially the collision induced dissociation (CID), to identify lipopeptides with beta-amino acyl residues. Two high-pressure liquid chromatography – high resolution mass spectrometry (HPLC-HRMS) methods differing in the collision energy, one with 50 eV and the other with 100eV, were proposed for the efficient analysis of the peptide sequence and the fatty acid identity. The methods were tested on iturin A, with the 50 eV CID method offering spectra suitable for the analysis of amino acid composition while the 100 eV method revealed prominent peaks of the immonium ions of the acyl chain. Subsequently we applied the two complementary methods to analyze a complex material – an extract of biomass of the puwainaphycin-producing strain *Cylindrospermum alatosporum* CCALA 998 – and along with a successful identification of known puwainaphycins we discovered additional novel variants differing mostly in the acyl chain. Moreover, the retention time of lipopeptide variants was linearly dependent on the length of the fatty acid over the range of C10 – C17 in iturins and C8 – C16 in puwainaphycins, which further facilitate the identification of the fatty acid length. Application of the method to the anabaenolysin family of lipopeptides, which are considerably smaller in size ( $m/z$  559.2758 for anabaenolysin A in comparison to  $m/z$  >1100 for the majority of puwainaphycins), required a considerable adjustment of the collision energy to be successful. Such an adjustment poses a considerable limitation for the application of the method for molecules with  $m/z$  <600 Da. Nevertheless, lowering the energy to 35eV we were able to detect six anabaenolysins in the crude extract of *Anabaena* sp. XPORK 15F. Finally, we applied the dual method to identify a  $\beta$ -amino-fatty acid-containing lipopeptides in cyanobacterial strains with unknown metabolic profiles. The collected data revealed the presence of 10 candidate lipopeptides in 8 of 240 screened strains; however we did not pursue their isolation and detailed structure characterization.

The development of the method and the investigation of the lipopeptide properties under different collision energies provided an essential background for the later work where we searched for novel lipopeptides and pursued their characterization.

### **2.3. Research Paper II. - Search for Novel Lipopeptide Biosynthesis Encoding Gene Clusters in Cyanobacteria and their Distribution**

The study takes an advantage of the notion that majority of the acylation of the known NRPS/PKS secondary metabolites of cyanobacteria is performed by fatty acid-AMP ligase (FAAL). The presence of a gene with an encoded FAAL domain in a NRPS (NRPS/PKS) gene cluster could thus be considered as a sign that the gene cluster encodes the biosynthesis of a lipopeptide. A position-specific scoring matrix (PSSM) was generated from a MAFFT alignment of the described cyanobacterial and bacterial FAAL domains known to be involved in the biosynthesis of lipopeptides or related molecules – namely: from PuwC (AIW82280, puwainaphycins), DptE (AAX31555, daptomycin), microginin acyl synthase (WP\_026798330), HctA (AAY42393, hectochlorin), CoIA (AKQ09578, columbamides), JamA (AAS98774, jamaicamides), Tar4 (AHH53502, taromycin), MycA (AAF08795, mycosubtilin), ItuA (BAB69698, iturin A), BmyA (ABS74181, bacillomycin), and PaIA (AHD05679, paenilarvins). In total the search yielded 79 putative lipopeptide biosynthetic gene clusters from 59 strains of cyanobacteria. The data was manually curated and novel gene clusters were categorized into 3 groups based on the topology of PKS/NRPS elements. The group without PKS consisted of 10 clusters, the group in which the FAAL was followed by PKS counted 11 clusters and finally the clusters in which FAAL was fused to NRPS and only then to PKS accounted for 31 clusters. Another 18 clusters were considered incomplete, since they were interrupted by a boundary of sequencing scaffold and hence were not classified. Similarly, the gene clusters with known secondary metabolites were not considered for this classification. However, all the found clusters were considered for a phylogenomic analysis and for an analysis of lipopeptide-to-lifestyle relationship. The putative lipopeptide gene clusters were found in numerous independent lineages across the cyanobacterial phylum with no evidence of a common origin. The observation is congruent with current views on the evolution of secondary metabolite gene clusters, which considers the evolution of gene clusters as a complex phenomenon with multiple modes of recombination within and among the clusters, horizontal gene transfer and various evolutionary forces in either divergent or convergent direction (Chevrette et al 2020). Although the

lipopeptide gene clusters were found throughout the phylogenomic tree of cyanobacteria, they were more frequent in derived lineages with larger cells, complex morphology and diversified thalli. This finding also follows the observed trend that NRPS-PKS biosynthetic gene clusters are more frequent in cyanobacteria with complex morphology and larger genome size (Shih et al 2013, Wang et al 2016). To find a possible relationship between the lipopeptide production and the strain ecology, we evaluated the origin of 376 strains with genome information available at the time via NCBI GenBank. Positive for the presence of a putative lipopeptide gene clusters were 59 strains (16%). The strains were categorized as planktic, substrate-associated or symbiotic. Of the 92 substrate-associated strains, 44 (48%) were found to harbor candidate gene clusters. In contrast, only 11 of the 250 (4%) planktic strains, and 1 of the 16 (6%) symbiotic cyanobacteria contained a gene clusters with matching criteria.

The observation that putative lipopeptide gene clusters are more frequently found in substrate-associated cyanobacteria is in line with the information obtained from firmicutes or delta-proteobacteria, where lipopeptides play a crucial role in colony development and surface colonization. Similarly, the observed analogies among the gene clusters suggests that they are shaped by similar evolutionary mechanisms as the secondary metabolite gene clusters in other microorganisms. However, the cyanobacterial gene clusters were more frequently unidirectional with genes for tailoring enzymes scattered among PKS/NRPS genes, contrasting the cassettes of tailoring enzymes found in other microorganisms (Hardy and Butler 2019, vs Paper IV).

Several clusters were observed to share a similar topology involving a particular C-domain, that was later shown to be crucial for  $\beta$ -OH-Asp incorporation, and a dioxygenase that was previously shown to catalyze the conversion of a Asp to a  $\beta$ -OH-Asp (Reitz et al 2019, Kreutzer et al 2012). This observation was further successfully elaborated in paper IV.

## 2.4. Research paper III. - Investigation of Fatty Acid Loading Modules in Puwainaphycin Gene Clusters.

Puwainaphycins (PUW) and minutissamides (MIN) are structurally analogous lipopeptides produced by terrestrial cyanobacteria around the world (Moore et al 1989, Hrouzek et al 2012, Mareš et al 2014). The study provides an in-depth insight into the variability of PUWs and MINs and the underlying differences in the composition of their biosynthetic gene clusters. Six producing strains, five of which were newly sequenced, were investigated and provided five known and more than 50 novel variants of PUW/MIN. The pool of the variants was sufficient to analyze the PUW lipopeptide class for conserved positions in the peptide ring. The gathered evidence suggests conservation of the peptide core, with hydrophobic amino acid in close proximity to the acyl chain, suggesting their functional importance, possibly in strengthening the interaction of the acyl chain with biological membranes.

The obtained data showed large variability of the acyl chain of PUWs/MINs. Interestingly we identified two types of FAAL presumably initiating the biosynthesis of PUWs/MINs. Type I. FAAL was found in three strains of *Cylindrospermum*, type II. was found in *Symplocastrum muelleri* and the two strains of *Anabaena* contained both types. Each type appeared to activate a distinct set of fatty acids resulting in particular lipopeptide profile, with strains harboring both types producing representatives of each. The observed metabolic pattern suggests that in case both are present the different initiation modules act as alternating starters of the biosynthesis. Analogous pattern was observed in anabaenopeptins where two NRPS modules alternate in the initiation of the biosynthesis (Rouhianen et al 2010). Inverse situation was recently reported for the laxaphycins, where the biosynthesis is forked after the PKS module into two NRPS pathways (Heinilä et al 2020; Fig. 8).

Comparison of the gene clusters and the metabolic profile between the individual strains pointed to several novel candidate tailoring enzymes encoded therein. Presence of *puwJ*, encoding a putative cytochrome P450-like oxidase, just downstream of the type II. starter module is correlated with the presence of PUW/MIN variants hydroxylated on the  $\omega-2$  carbon of the acyl residue. A putative halogenase-encoding *puwK*, also associated

with the type II. starter is likely responsible for  $\omega$ -chlorination of MIN B. And finally the *puwL* encoding a putative O-acetyltransferase, found in *S. muelleri* catalyzes acetylation of the hydroxy residue on the far end of the fatty acid and generates five additional PUW variants. Although co-occurrence of the tailoring genes and the corresponding variants is not a direct proof of their activity, such as mutational analysis, it could still be considered as a very strong indication especially in cases where the investigated organisms are not routinely accessible by genetic engineering. PUWs/MINs were previously shown to be bio-active, in particular toxic to human cells (Hrouzek et al 2012). Considering the antimicrobial effect of other lipopeptides such as iturin or daptomycin, we decided to investigate the activity of PUW F and MIN A, C and D against a panel of selected Gram-positive and Gram-negative bacterial strains (13). The tested compounds did not exhibit antimicrobial effect on the bacterial strains. However the PUW F variant has impaired the growth of *Candida albicans* and *Saccharomyces cerevisiae* at minimal inhibitory concentration 5.5  $\mu$ M. As for the other tested compounds, MIN A exhibited hints of inhibitory zone in the disk diffusion assays against the yeasts, while MIN C and D did not exhibit any considerable antagonistic effect. Minor structural differences are likely responsible for the behavior. In contrast to PUW F, MIN A features an acyl residue that is 2 carbons shorter, while MIN C and D feature a hydroxy or oxo group respectively on C16 that likely alter the membrane interactions required for toxicity. Interestingly, based on our but also previously published data the PUWs/MINs seem to be active only against eukaryotic cells (mostly tested on human cell lines; Kang et al 2012, Hrouzek et al 2012) while leaving the prokaryotes unaffected. This may have considerable implications as PUWs/MINs are frequently produced by soil inhabiting cyanobacteria such as *Anabaena* and *Cylindrospermum*, but more importantly also by *Symplocastrum muelleri* that commonly forms thick biofilms in boglands of the subarctics, where also mysterious intoxication of sheep, known as alveld, occurs (Mysterud et al 2016). At the time the possible involvement of cytotoxic cyanobacterial lipopeptides in genesis of the disease was not deeply investigated and could not be unambiguously excluded.



The study provides information on structural variation of a prominent class of cyanobacterial lipopeptides, PUWs and MINs, along with corresponding genetic basis for their production. Intriguing are also the hints pointing to selective bioactivity of particular variants of PUWs/MINs against eukaryotes. On multiple occasions bacteria were shown to produce lipopeptides that suppress the growth of other bacteria in the vicinity of the producer (D'aes et al 2014, Henkels et al 2014, Bais et al 2004, Raajimakers et al 2010). So far cyanobacterial lipopeptides appear to preferably target eukaryotes, but whether this is true, is an open question worth investigating. Similarly, worth of further attention is the possible toxicity of PUWs/MINs to animals via skin exposure and ingestion.

## 2.5. Research Paper IV. - in press - Discovery of Cyanochelins – Novel and Possibly Widespread Cyanobacterial Siderophores

Building up on the previous experience with the detection of lipopeptides and their respective gene clusters as well as with the biosynthetic potential of FAAL we decided to investigate a group of clusters noted during the work on paper II. One of the strains identified therein was obtained and a candidate lipopeptide was detected with the method developed in paper I. Similar lipopeptides were detected in other two candidate strains with considerably similar gene clusters.

A prominent feature of the detected lipopeptides was the presence of  $\beta$ -OH-Asp residue. A literature review on  $\beta$ -OH-Asp showed that it is a common component of the iron chelating lipopeptides (Kem and Butler 2015). The literature further suggested that it comes from an Asp activated by an Asp-specific A-domain and subsequently hydroxylated on its  $\beta$ -carbon by a dioxygenase (Kreutzer et al 2012). The process is further controlled by a specific C-domain that allows only modified Asp – the  $\beta$ -OH-Asp – to be incorporated into a nascent NRPS peptide. The particular organization of the genes encoding the enzymes for the incorporation of  $\beta$ -OH-Asp was then retrospectively identified as the pattern in the gene clusters that originally caught our attention.

In the paper we provide bioinformatic description of the candidate lipopeptide siderophores from *Rivularia* sp. PCC 7116, *Leptolyngbya* sp. NIES 3755 and *Rubidibacter lacunae* KORDI 51-2. The clusters encode biosynthesis of lipopeptide 6-7 AA long with two  $\beta$ -OH-Asp incorporating NRPS modules. Furthermore, in vicinity of clusters from *Rivularia* and *Leptolyngbya*, there are siderophore transporters and DNA-regulatory elements in the promotor region of the first biosynthetic gene that are responsive to iron stress (FurA binding site). These findings supported our suspicion that the encoded lipopeptides would be novel siderophores, of a structural type previously not found in cyanobacteria.

Interestingly, despite the general expectation that cyanobacteria should produce siderophores, only three classes of siderophores, in particular schizokinen, synechobactin and anachelin and their minor variants, were identified in the phylum (Årstøl and Hohmann-Marriott 2019). Moreover, cyanobacteria contain a variety of gene clusters that are silent in the

soothing environment of culture collections, with all nutrients available in plenty and in the absence of a malevolent competitor or parasite (Shih et al 2013). The benefits of secondary metabolite production are, however, fully manifested in a more complex environmental matrices, often in the presence of nutritional stress or other organisms, e. g. as was described in the section natural role of lipopeptides. If proper triggers are provided during the cultivation, they may lead to identification of novel compounds, including siderophores (Gross 2007).

Upon iron starvation the strains of interest accumulated compounds with the MS/MS fragmentation spectra corresponding with the bioinformatically predicted peptides. The analysis revealed the presence of a compound with  $m/z$  781.36 in *Rivularia* sp. PCC 7116 including partial amino acid sequence [Gly]-[Gly]-[ $\beta$ -OH-Asp]-[Gly]. Candidate compounds detected in *Leptolyngbya* ( $m/z$  1,026.51) and *Rubidibacter* ( $m/z$  1,045.51) contained [Phe]-[Gly]-[ $\beta$ -OH-Asp]-[Ser]-[ $\beta$ -OH-Asp] and [Gly]-[Ala]-[ $\beta$ -OH-Asp]-[Ala]-[ $\beta$ -OH-Asp], respectively. The stimulation of siderophore production after iron starvation is not surprising and was observed repeatedly in bacteria. What is a bit surprising, is that it has not been applied more often in order to identify the products of the orphan NRPS gene clusters found in cyanobacteria.

The candidate siderophores were isolated by HPLC in small amounts and tested by a mass spectrometric experiment for the ability to form complexes with  $Fe^{3+}$ . The measurements confirmed the formation of iron-lipopeptide complexes, as we observed a shift of the molecular ions of lipopeptides by 52.91, corresponding to the molecular weight of iron bound instead of three hydrogens. Moreover, the HPLC/MS analysis of crude extracts included ions corresponding to iron-lipopeptide complexes, further supporting the selectivity of the compounds to iron. We concluded that the candidate compounds are indeed lipopeptide siderophores.

A typical feature of the  $\beta$ -OH-Asp siderophores is their ability to reduce  $Fe^{3+}$  to  $Fe^{2+}$  (Kreutzer et al 2012, Hardy et al 2019). We decided to investigate if our candidate compounds are capable of that too. Aliquots of the isolated compounds were exposed to sunlight in the presence of  $Fe^{3+}$  and bathophenanthroline disulfonic acid (BPDS, 0.1mM), which develops colorful complexes when bound to  $Fe^{2+}$ . The sunlight-exposed

aliquots of all three candidate compounds developed strong pink coloration indicative of Fe<sup>2+</sup> presence while the light-protected samples did not. Furthermore, expected photolytic fragments of the lipopeptides (cleaved at N-terminal side of the  $\beta$ -OH-Asp) were detected by a subsequent HPLC/MS analysis.

Compound 781.36 was isolated in a sufficient amount (2.2 mg) to undergo NMR structural analysis. The MS/MS data along with NMR measurements yielded the structure of a compound named cyanochelin A, the first representative of a potential class of cyanobacterial siderophores. Unfortunately, the other two cyanochelins were not acquired in sufficient amounts for their structures to be evaluated in time. However, their structure will likely be evaluated in near future as a part of a follow-up study.

Finally, we looked at the distribution of the clusters encoding the biosynthesis of analogous compounds, the cyanochelin-like gene clusters, throughout the cyanobacterial phylogenetic tree. We found additional 27 clusters in 29 strains, which included FAAL and two  $\beta$ -OH-Asp specific NRPS modules. Of these 25 further included genes for siderophore transport machinery in their vicinity. The detected clusters, including the three analyzed in detail, were found throughout the entire phylum, within both basal and derived lineages of cyanobacteria. The analysis has shown that cyanochelins or cyanochelin-like  $\beta$ -OH-Asp siderophores are considerably frequent and widespread.

The work reported in paper IV. provides important insights into the ecophysiology of cyanobacteria. Most importantly, it reveals the immense potential of cyanobacteria to produce  $\beta$ -OH-Asp siderophores that is not limited to any particular phylogenetic lineage nor habitat type or cell morphology and that has been overlooked so far. The study contributes several new compounds to the limited arsenal of known cyanobacterial siderophores and provides a lead to the discovery of additional new ones. Moreover, the reported information further develops mapping of cyanobacterial gene clusters and provides significant clues to the possible products of the many orphan biosynthetic gene clusters. Most intriguing, however, was the photoreactive aspect of the discovered compounds, since cyanobacteria are phototrophic organisms dependent on light that

can severely shorten the half-life of produced siderophores. It would be exciting to investigate what additional benefit could the photoreactive nature of cyanochelins bring to the phototrophic producer, or eventually to the whole microbial community, and in what habitats this is of significant importance. A study by Barbeau (Barbeau et al 2001) reports possible exchange of siderophore-reduced iron for exudates in the open-ocean consortium of bacteria and phototrophic algae, but how this would apply to a situation where the phototroph is also the siderophore producer is an intriguing question. For now, the natural function and regulation of cyanobacterial lipopeptides remains an under-investigated field to which the work presented in paper IV. aims to contribute.

## **2.6. Research Paper V. - manuscript – Biosynthesis of Muscotoxin in *Desmonostoc muscorum* and Its Crosstalk with Nostopeptolide Synthesis.**

The study investigates the diversity and biosynthesis of muscotoxins produced by a strain of *Desmonostoc muscorum*. In contrast to PUWs/MINs, the detected muscotoxins show invariant incorporation of 2,5-dihydroxy-3-amino-dodecanoic acid (5-OH-Ahdoa). The muscotoxin variants differed at position 4, 6, and 11. At position 4 and 11 either proline or  $\gamma$ -methyl proline (4-Me-Pro) were found while position 6 is occupied either by Val or Ile.

The genome of *D. muscorum* was assembled and the 53 kbp long putative muscotoxin gene cluster was identified. The cluster included genes *mcxA* encoding a PKS, *mcxB* coding for a hybrid PKS/NRPS, *mcxC* encoding a NRPS gene and *mcxD* likely responsible for synthesis of a transporter system. The biosynthetic genes exhibited considerable homology to the biosynthetic gene cluster of the cyanobacterial  $\beta$ -amino cyclic heptapeptide, the nostophycin. The biosynthesis of nostophycins starts with an adenylation domain of NpnA; however, only a rudiment of such a starter A-domain is found in McxA, the first enzyme in the proposed biosynthetic pathway. And while it is highly probable that structurally similar muscotoxins and nostophycins shared an ancestral biosynthetic pathway, the starter A-domain is very likely not active in the muscotoxin pathway and an alternative mechanism must be responsible for the recognition and activation of the fatty acid. Perhaps the ketoreductase domain found in the McxA, could ensure the recognition of the correct substrate. A lack of fatty acid-activating enzymes within the biosynthetic gene cluster of lipopeptides is known from other cases. For example, the anabaenolysin gene cluster also does not include FA-activating enzymes and neither does the surfactin biosynthetic gene cluster. In such cases the FA is provided in-trans, probably from intermediates of primary metabolism.

Interesting is also the incorporation of 4-Me-Pro into muscotoxins as the biosynthetic gene cluster did not include any genes for the biosynthesis of 4-Me-Pro. Instead, a 4-Me-Pro synthesizing cassette was found in a biosynthetic gene cluster of another cyanobacterial peptide, the nostopeptolide, also present elsewhere in the genome. Analogously as was

observed in the biosynthesis of spumigins and aeruginosins (Liu et al 2015) we hypothesized that NosE and NosF proteins of the nostopeptolide biosynthetic pathway cyclize Leu into 4-Me-Pro, which is subsequently incorporated into muscotoxins. The origin of 4-Me-Pro was investigated by employing isotopically-labeled Leu. The experiments showed that the 4-Me-Pro incorporated into muscotoxins does not come from Pro, but from Leu that is converted to 4-Me-Pro likely by the accessory enzymes of the nostopeptolide pathway. The proportion of labeled muscotoxin (19.5%) is roughly equal to that of nostopeptolide (20.1%) and suggests that after its synthesis by NosE and NosF, the 4-Me-Pro is likely made generally available rather than being streamed exclusively to the nostopeptolide pathway. Hence the observed ratio between the muscotoxin A with Pro and muscotoxin B with 4-Me-Pro is likely not dependent on the amount of 4-Me-Pro leaked, but rather the result of the specificity and preferences of the respective A-domain.

The study was submitted to ChemBioChem journal, whose reviewers requested additional data on specificity of the A-domain. Although the request was not strictly mandatory, we decided to investigate the issue in more detail as we find it interesting. As a consequence, the publication of the paper has been postponed, requiring me to attach it to the thesis in the form of an advanced manuscript.

In general, however, the paper investigates particular issues in the field of secondary metabolites of cyanobacteria, such as the less prevalent modes of lipoinitiation, the cross-talk between the biosynthetic pathways but, in a sense, also the objective limitations of the bioinformatic prediction of lipopeptides produced by a particular biosynthetic gene cluster in a wider genomic context. The study thus provides valuable details to the knowledge of cyanobacterial lipopeptides and their biosynthesis.

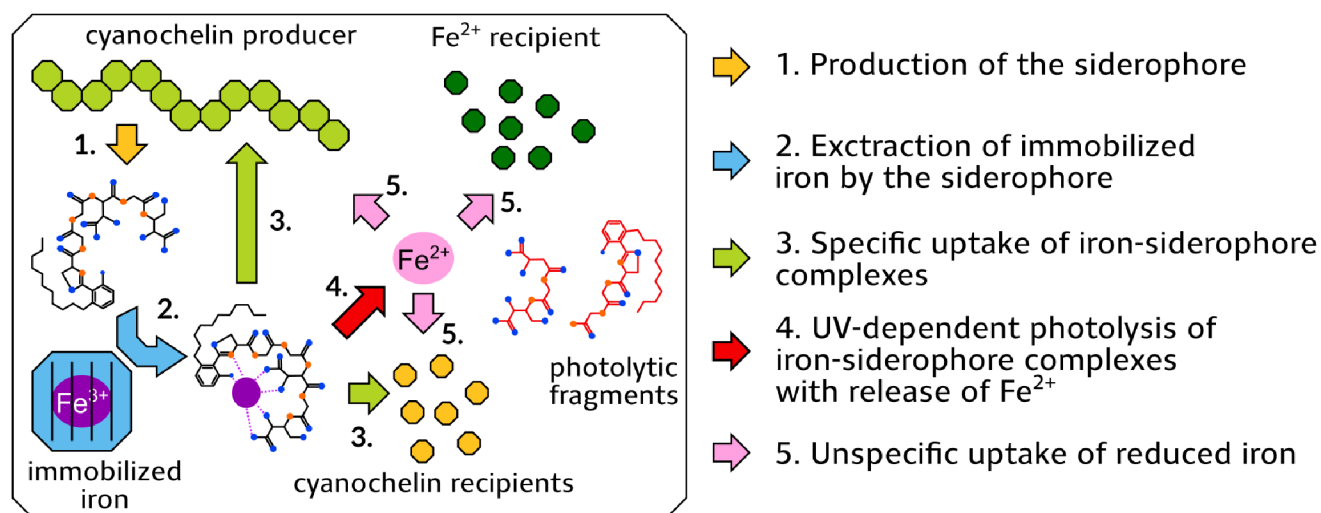
## 2.7. Summary and future perspectives

In its initial stage the project aspired to explore the diversity and potential eco-physiological roles of cyanobacterial lipopeptides. The original idea was to investigate the relationship between the lipopeptide production and the cyanobacterial lifestyle. We assumed, analogously to what is known from *Bacillus* or *Pseudomonas*, that lipopeptides would be required for a sessile vs. planktonic growth of the producing cyanobacterium. The bioinformatic screening revealed that lipopeptide biosynthetic gene clusters are more frequent in the genomes of substrate-associated cyanobacteria; however, that may not necessarily imply the involvement of the lipopeptides in the colony growth (paper II.). Several strains harboring the candidate gene clusters were inspected for the production of lipopeptides; however, only in one case we were able to identify the product. In an attempt to upscale the production to obtain sufficient amount of the candidate compound for further characterization, the compound production ceased. A revision of the structural features of the candidate compound indicated that it could be a siderophore, produced only under iron stress. The original aim to investigate the possible ecophysiological role of cyanobacterial lipopeptides thus gained a fresh momentum. Subsequent work (paper IV.) led to the identification of cyanochelins, the cyanobacterial lipopeptide siderophores that employ  $\beta$ -OH-Asp for the chelation of iron and are possibly widespread across cyanobacteria. The study significantly expands the arsenal of known cyanobacterial siderophores on one hand, and on the other it provides valuable additions with a possible ecophysiological function to the repertoire of cyanobacterial lipopeptides, eventually showing that cyanochelins and similar molecules, may actually be very frequent in cyanobacteria. This is what I personally consider the most important contribution of the presented work to the field of cyanobacterial secondary metabolites.

Interestingly the discovered siderophores can reduce ferric iron to ferrous iron if exposed to UV-light. This has been reported previously (Barbeau et al 2001, Hardy et al 2019) but in the novel context of photoautotrophic cyanobacteria, it raises exciting questions on the direction of the fluxes of iron scavenged and reduced by photoreactive siderophores (Fig. 16). Are



the producing cyanobacteria the only recipients of the siderophore-acquired iron? What other organisms may benefit from the production of cyanochelins and similar siderophores by cyanobacteria? What benefit do the siderophore/reduced iron recipients offer in exchange? How big is the proportion of iron leaked from the photolyzed siderophore to the non-producing organisms? The literature suggest that some microorganisms are dependent on the provision of siderophores by other organisms (D'Onofrio et al 2010), so how will this project into epibiont recruitment by cyanobacteria, or the composition of microbial community in cyanobacteria dominated (micro-)habitats?



**Fig. 16: Proposed routes of photolytic siderophore cycling in a microbial community.**

During my studies I was also kindly offered the opportunity to join my colleagues in the other projects that I value highly. The first one (paper I.) provided a solid foundation for my future work with compound isolation, identification and fragmentation analysis using HPLC-HRMS/MS. Puwainaphycins and minutissamides are extensively studied in the lab that I am a part of. The study I participate in (paper III.) illustrates how the structural diversity within a compound class could be achieved. The study provides a valuable insight into the biosynthesis of cyanobacterial lipopeptides and considerable count of diverse puwainaphycin variants, which most importantly point out the conserved and most likely functionally important residues within the compound scaffold. I also had the chance to participate in the so far unpublished study of muscotoxin (paper V.), where I think the most significant part is the evaluation of the crosstalk between the two biosynthetic pathways. The lifespan of the

secondary metabolites as well as the rate of their synthesis has been addressed only occasionally (Orr and Jones 1998) and is naturally highly dependent on the growth rate and conditions.

Overall, I believe that the presented work has met its original aim to investigate the diversity and possible ecophysiological role of cyanobacterial lipopeptides and provided valuable contribution to the field of study. I think that the production of photoreactive siderophores by cyanobacteria deserves further exploration, especially in the context of other co-habiting organisms. And as for the role of lipopeptides in colony development, I think that the advances in the genetic manipulation of cyanobacteria will soon open novel possibilities. Yet, on the other hand, I think that the cultivation techniques would need important development, so that we would know the conditions that trigger motility in the model cyanobacterial strains with same degree of certainty as we know it for *Bacillus* or *Pseudomonas*.

### 3. References

- Almaliti, J., Malloy, K. L., Glukhov, E., Spadafora, C., Gutierrez, M., & Gerwick, W. H. (2017). Dudawalamides A-D, Antiparasitic Cyclic Depsipeptides from the Marine Cyanobacterium *Moorea producens*. *Journal of Natural Products*, *80*(6), 1827-1836.  
<https://doi.org/10.1021/acs.jnatprod.7b00034>
- An, T., Kumar, T. K., Wang, M., Liu, L., Lay, J. O., Liyanage, R., Berry, J., Gantar, M., Marks, V., Gawley, R. E., & Rein, K. S. (2007). Structures of pahayokolides A and B, cyclic peptides from a *Lyngbya* sp. *Journal of Natural Products*, *70*(5), 730-735. <https://doi.org/10.1021/np060389p>
- Årstøl, E., & Hohmann-Marriott, M. F. (2019). Cyanobacterial Siderophores- Physiology, Structure, Biosynthesis, and Applications. *Marine Drugs*, *17*(5). <https://doi.org/10.3390/md17050281>
- Bais, H. P., Fall, R., & Vivanco, J. M. (2004). Biocontrol of *Bacillus subtilis* against infection of *Arabidopsis* roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiology*, *134*(1), 307-319. <https://doi.org/10.1104/pp.103.028712>
- Barbeau, K., Rue, E. L., Bruland, K. W., & Butler, A. (2001). Photochemical cycling of iron in the surface ocean mediated by microbial iron(iii)-binding ligands. *Nature*, *413*(6854), 409-413. <https://doi.org/10.1038/35096545>
- Berti, A. D., Greve, N. J., Christensen, Q. H., & Thomas, M. G. (2007). Identification of a biosynthetic gene cluster and the six associated lipopeptides involved in swarming motility of *Pseudomonas syringae* pv. tomato DC3000. *Journal of Bacteriology*, *189*(17), 6312-6323.  
<https://doi.org/10.1128/JB.00725-07>
- Blank, C. E., & Sánchez-Baracaldo, P. (2010). Timing of morphological and ecological innovations in the cyanobacteria - a key to understanding the rise in atmospheric oxygen. *Geobiology*, *8*(1), 1-23.  
<https://doi.org/10.1111/j.1472-4669.2009.00220.x>
- Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S. Y., Medema, M. H., & Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, *47*(W1), W81-W87. <https://doi.org/10.1093/nar/gkz310>

- Borics, G., Abonyi, A., Salmaso, N., & Ptacnik, R. (2021). Freshwater phytoplankton diversity: models, drivers and implications for ecosystem properties. *Hydrobiologia*, *848*(1), 53-75.  
<https://doi.org/10.1007/s10750-020-04332-9>
- Boudreau, P. D., Byrum, T., Liu, W. T., Dorrestein, P. C., & Gerwick, W. H. (2012). Viequeamide A, a Cytotoxic Member of the Kulolide Superfamily of Cyclic Depsipeptides from a Marine Button Cyanobacterium. *Journal of Natural Products*, *75*(9), 1560-1570. <https://doi.org/10.1021/np300321b>
- Bouma-Gregson, K., Power, M. E., & Bormans, M. (2017). Rise and fall of toxic benthic freshwater cyanobacteria (*Anabaena* spp.) in the Eel river: Buoyancy and dispersal. *Harmful Algae*, *66*, 79-87.  
<https://doi.org/10.1016/j.hal.2017.05.007>
- Breitbarth, E., Oschlies, A., & LaRoche, J. (2007). Physiological constraints on the global distribution of Trichodesmium - effect of temperature on diazotrophy. *Biogeosciences*, *4*(1), 53-61. <https://doi.org/10.5194/bg-4-53-2007>
- Bui, T. H., Wray, V., Nimtz, M., Fossen, T., Preisitsch, M., Schröder, G., Wende, K., Heiden, S. E., & Mundt, S. (2014). Balticidins A-D, antifungal hassallidin-like lipopeptides from the Baltic Sea cyanobacterium *Anabaena cylindrica* Bio33. *Journal of Natural Products*, *77*(6), 1287-1296.  
<https://doi.org/10.1021/np401020a>
- Bunyajetpong, S., Yoshida, W. Y., Sitachitta, N., & Kaya, K. (2006). Trungapeptins A-C, cyclodepsipeptides from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, *69*(11), 1539-1542.  
<https://doi.org/10.1021/np050485a>
- Cai, W., Matthew, S., Chen, Q.-Y., Paul, V. J., & Luesch, H. (2018). Discovery of new A- and B-type laxaphycins with synergistic anticancer activity. *Bioorganic & Medicinal Chemistry*, *26*(9), 2310-2319.  
<https://doi.org/10.1016/j.bmc.2018.03.022>
- Campbell, S. E. (1979). Soil Stabilization by a Prokaryotic Desert Crust - Implications for Precambrian Land Biota. *Origins of Life and Evolution of the Biosphere*, *9*(4), 335-348. <https://doi.org/10.1007/BF00926826>
- Capone, D. G., Burns, J. A., Montoya, J. P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A. F., & Carpenter, E. J. (2005). Nitrogen fixation by

*Trichodesmium* spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean. *Global Biogeochemical Cycles*, 19(2), Article ARTN GB2024. <https://doi.org/10.1029/2004GB002331>

Caputo, A., Nylander, J. A. A., & Foster, R. A. (2019). The genetic diversity and evolution of diatom-diazotroph associations highlights traits favoring symbiont integration. *FEMS Microbiology Letters*, 366(2), Article ARTN fny297. <https://doi.org/10.1093/femsle/fny297>

Chang, Z., Sitachitta, N., Rossi, J. V., Roberts, M. A., Flatt, P. M., Jia, J., Sherman, D. H., & Gerwick, W. H. (2004). Biosynthetic pathway and gene cluster analysis of curacin A, an antitubulin natural product from the tropical marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 67(8), 1356-1367. <https://doi.org/10.1021/np0499261>

Chang, Z. X., Flatt, P., Gerwick, W. H., Nguyen, V. A., Willis, C. L., & Sherman, D. H. (2002). The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene*, 296(1-2), 235-247, Article Pii s0378-1119(02)00860-0. [https://doi.org/10.1016/s0378-1119\(02\)00860-0](https://doi.org/10.1016/s0378-1119(02)00860-0)

Chevrette, M. G., Gutiérrez-García, K., Selem-Mojica, N., Aguilar-Martínez, C., Yañez-Olvera, A., Ramos-Aboites, H. E., Hoskisson, P. A., & Barona-Gómez, F. (2020). Evolutionary dynamics of natural product biosynthesis in bacteria. *Natural Product Reports*, 37(4), 566-599. <https://doi.org/10.1039/c9np00048h>

Cochrane, S. A., & Vederas, J. C. (2016). Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A Gold Mine of Antibiotic Candidates. *Medicinal Research Reviews*, 36(1), 4-31. <https://doi.org/10.1002/med.21321>

D'aes, J., Kieu, N. P., Léclère, V., Tokarski, C., Olorunleke, F. E., De Maeyer, K., Jacques, P., Höfte, M., & Ongena, M. (2014). To settle or to move? The interplay between two classes of cyclic lipopeptides in the biocontrol strain *Pseudomonas* CMR12a. *Environmental Microbiology*, 16(7), 2282-2300. <https://doi.org/10.1111/1462-2920.12462>

D'Onofrio, A., Crawford, J. M., Stewart, E. J., Witt, K., Gavrish, E., Epstein, S., Clardy, J., & Lewis, K. (2010). Siderophores from neighboring organisms

promote the growth of uncultured bacteria. *Chemistry & Biology*, 17(3), 254-264. <https://doi.org/10.1016/j.chembiol.2010.02.010>

de los Rios, A., Ascaso, C., Wierzchos, J., Vincent, W. F., & Quesada, A. (2015). Microstructure and cyanobacterial composition of microbial mats from the High Arctic. *Biodiversity and Conservation*, 24(4), 841-863. <https://doi.org/10.1007/s10531-015-0907-7>

Demoulin, C. F., Lara, Y. J., Cornet, L., François, C., Baurain, D., Wilmotte, A., & Javaux, E. J. (2019). Cyanobacteria evolution: Insight from the fossil record. *Free Radical Biology and Medicine*, 140, 206-223. <https://doi.org/10.1016/j.freeradbiomed.2019.05.007>

Dittmann, E., Fewer, D. P., & Neilan, B. A. (2013). Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiology Reviews*, 37(1), 23-43. <https://doi.org/10.1111/j.1574-6976.2012.12000.x>

Duitman, E. H., Hamoen, L. W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F., Reinhardt, R., Schmidt, M., Ullrich, C., Stein, T., Leenders, F., & Vater, J. (1999). The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23), 13294-13299. <https://doi.org/10.1073/pnas.96.23.13294>

Durako, M. J., Medlyn, R. A., & Moffler, M. D. (1982). Particulate Matter Resuspension via Metabolically Produced Gas-Bubbles from Benthic Estuarine Microalgae Communities. *Limnology and Oceanography*, 27(4), 752-756. <https://doi.org/10.4319/lo.1982.27.4.0752>

Edwards, D. J., & Gerwick, W. H. (2004). Lyngbyatoxin biosynthesis: sequence of biosynthetic gene cluster and identification of a novel aromatic prenyltransferase. *Journal of American Chemical Society*, 126(37), 11432-11433. <https://doi.org/10.1021/ja047876g>

Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A., & Gerwick, W. H. (2004). Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chemistry & Biology*, 11(6), 817-833. <https://doi.org/10.1016/j.chembiol.2004.03.030>

- Engene, N., Rottacker, E. C., Kastovsky, J., Byrum, T., Choi, H., Ellisman, M. H., Komarek, J., & Gerwick, W. H. (2012). *Moorea producens* gen. nov., sp nov and *Moorea bouillonii* comb. nov., tropical marine cyanobacteria rich in bioactive secondary metabolites. *International Journal of Systematic and Evolutionary Microbiology*, *62*, 1171-1178.  
<https://doi.org/10.1099/ijs.0.033761-0>
- Fewer, D. P., Österholm, J., Rouhiainen, L., Jokela, J., Wahlsten, M., & Sivonen, K. (2011). Nostophycin biosynthesis is directed by a hybrid polyketide synthase-nonribosomal peptide synthetase in the toxic cyanobacterium *Nostoc* sp. strain 152. *Applied and Environmental Microbiology*, *77*(22), 8034-8040. <https://doi.org/10.1128/AEM.05993-11>
- Finke, N., Simister, R. M., O'Neil, A. H., Nomosatryo, S., Henny, C., MacLean, L. C., Canfield, D. E., Konhauser, K. O., Lalonde, S. V., Fowle, D. A., & Crowe, S. A. (2019). Mesophilic microorganisms build terrestrial mats analogous to Precambrian microbial jungles. *Nature Communications*, *10*, Article ARTN 4323. <https://doi.org/10.1038/s41467-019-11541-x>
- Frankmölle, W. P., Knübel, G., Moore, R. E., & Patterson, G. M. L. (1992). Antifungal Cyclic-Peptides from the Terrestrial Blue-Green-Alga *Anabaena laxa*. 2. Structures of Laxaphycin-A Laxaphycin-B, Laxaphycin-D and Laxaphycin-E. *Journal of Antibiotics*, *45*(9), 1458-1466.  
<https://doi.org/10.7164/antibiotics.45.1458>
- Gauglitz, J. M., Inishi, A., Ito, Y., & Butler, A. (2014). Microbial Tailoring of Acyl Peptidic Siderophores. *Biochemistry*, *53*(16), 2624-2631.  
<https://doi.org/10.1021/bi500266x>
- Goebel, N. L., Turk, K. A., Achilles, K. M., Paerl, R., Hewson, I., Morrison, A. E., Montoya, J. P., Edwards, C. A., & Zehr, J. P. (2010). Abundance and distribution of major groups of diazotrophic cyanobacteria and their potential contribution to N<sub>2</sub> fixation in the tropical Atlantic Ocean. *Environmental Microbiol*, *12*(12), 3272-3289.  
<https://doi.org/10.1111/j.1462-2920.2010.02303.x>
- Golakoti, T., Yoshida, W. Y., Chaganty, S., & Moore, R. E. (2000). Isolation and structures of nostopeptolides A1, A2 and A3 from the cyanobacterium *Nostoc* sp GSV224. *Tetrahedron*, *56*(46), 9093-9102.  
[https://doi.org/10.1016/S0040-4020\(00\)00764-X](https://doi.org/10.1016/S0040-4020(00)00764-X)

- González-Resendiz, L., Johansen, J. R., Alba-Lois, L., Segal-Kischinevzky, C., Escobar-Sánchez, V., Jiménez-García, L. F., Hauer, T., & León-Tejera, H. (2018). Nunduva, a new marine genus of Rivulariaceae (Nostocales, Cyanobacteria from marine rocky shores [Article]. *Fottea*, *18*(1), 86-105. <https://doi.org/10.5507/fot.2017.018>
- Götze, S., & Stallforth, P. (2020). Structure, properties, and biological functions of nonribosomal lipopeptides from pseudomonads. *Natural Product Reports*, *37*(1), 29-54. <https://doi.org/10.1039/c9np00022d>
- Gregson, J. M., Chen, J.-L., Patterson, G. M. L., & Moore, R. E. (1992). Structure of Puwainaphycins A-E. *Tetrahedron*, *48*(18), 3727-3734. [https://doi.org/10.1016/S0040-4020\(01\)92264-1](https://doi.org/10.1016/S0040-4020(01)92264-1)
- Gross, H. (2007). Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Applied Microbiology and Biotechnology*, *75*(2), 267-277. <https://doi.org/10.1007/s00253-007-0900-5>
- Han, B. N., McPhail, K. L., Gross, H., Goeger, D. E., Mooberry, S. L., & Gerwick, W. H. (2005). Isolation and structure of five lyngbyabellin derivatives from a Papua New Guinea collection of the marine cyanobacterium *Lyngbya majuscula*. *Tetrahedron*, *61*(49), 11723-11729. <https://doi.org/10.1016/j.tet.2005.09.036>
- Hardy, C. D., & Butler, A. (2019). Ambiguity of NRPS Structure Predictions: Four Bidentate Chelating Groups in the Siderophore Pacifibactin. *Journal of Natural Products*, *82*(4), 990-997. <https://doi.org/10.1021/acs.jnatprod.8b01073>
- Hauer, T., Mühlsteinová, R., Bohunická, M., Kaštovský, J., & Mareš, J. (2015). Diversity of cyanobacteria on rock surfaces. *Biodiversity and Conservation*, *24*(4), 759-779. <https://doi.org/10.1007/s10531-015-0890-z>
- Heinilä, L. M. P., Fewer, D. P., Jokela, J. K., Wahlsten, M., Jortikka, A., & Sivonen, K. (2020). Shared PKS Module in Biosynthesis of Synergistic Laxaphycins. *Frontiers in Microbiology*, *11*, Article ARTN 578878. <https://doi.org/10.3389/fmicb.2020.578878>
- Henkels, M. D., Kidarsa, T. A., Shaffer, B.T., Goebel, N. C., Burlinson, P., Mavrodi, D. V., Bentley, M. A., Rangel, L. I., Davis II, E. W., Thomashow, L. S., Zabriskie, T. M., Preston, G. M., & Loper, J. E. (2014). *Pseudomonas*



*protegens* Pf-5 Causes Discoloration and Pitting of Mushroom Caps Due to the Production of Antifungal Metabolites. *Molecular Plant-Microbe Interactions*, 27(7), 733-746. <https://doi.org/10.1094/MPMI-10-13-0311-R>

Hermenau, R., Ishida, K., Gama, S., Hoffmann, B., Pfeifer-Leeg, M., Plass, W., Mohr, J. F., Wichard, T., Saluz, H.-P., & Hertweck, C. (2018). Gramibactin is a bacterial siderophore with a diazeniumdiolate ligand system. *Nature Chemical Biology*, 14(9), 841-843. <https://doi.org/10.1038/s41589-018-0101-9>

Hilton, J. A., Foster, R. A., Tripp, H. J., Carter, B. J., Zehr, J. P., & Villareal, T. A. (2013). Genomic deletions disrupt nitrogen metabolism pathways of a cyanobacterial diatom symbiont. *Nature Communications*, 4, Article ARTN 1767. <https://doi.org/10.1038/ncomms2748>

Horgen, F. D., Yoshida, W. Y., & Scheuer, P. J. (2000). Malevamides A-C, new depsipeptides from the marine cyanobacterium *Symploca laete-viridis*. *Journal of Natural Products*, 63(4), 461-467. <https://doi.org/10.1021/np990449+>

Hrouzek, P., Kuzma, M., Černý, J., Novák, P., Fišer, R., Šimek, P., Lukešová, A., & Kopecký, J. (2012). The Cyanobacterial Cyclic Lipopeptides Puwainaphycins F/G Are Inducing Necrosis via Cell Membrane Permeabilization and Subsequent Unusual Actin Relocalization. *Chemical Research in Toxicology*, 25(6), 1203-1211. <https://doi.org/10.1021/tx300044t>

Hutchinson, G. E. (1961). The Paradox of the Plankton. *American Naturalist*, 95(882), 137-145. <https://doi.org/10.1086/282171>

Ishida, K., Kato, T., Murakami, M., Watanabe, M., & Watanabe, M. F. (2000). Microginins, zinc metalloproteases inhibitors from the cyanobacterium *Microcystis aeruginosa*. *Tetrahedron*, 56(44), 8643-8656. [https://doi.org/10.1016/S0040-4020\(00\)00770-5](https://doi.org/10.1016/S0040-4020(00)00770-5)

Janatková, K., Řeháková, K., Doležal, J., Šimek, M., Chlumská, Z., Dvorský, M., & Kopecký, M. (2013). Community structure of soil phototrophs along environmental gradients in arid Himalaya. *Environmental Microbiology*, 15(9), 2505-2516. <https://doi.org/10.1111/1462-2920.12132>

Johansen, J. R., & Casamatta, D. A. (2005). Recognizing cyanobacterial diversity through adoption of a new species paradigm. *Algological*

- Studies/Archiv für Hydrobiologie, Supplement Volumes, 117, 71-93.*  
<https://doi.org/10.1127/1864-1318/2005/0117-0071>
- Jokela, J., Oftedal, L., Herfindal, L., Permi, P., Wahlsten, M., Doskeland, S. O., & Sivonen, K. (2012). Anabaenolysins, Novel Cytolytic Lipopeptides from Benthic *Anabaena* Cyanobacteria [Article]. *Plos ONE*, 7(7), Article ARTN e41222. <https://doi.org/10.1371/journal.pone.0041222>
- Kang, H.-S., Kronic, A., Shen, Q., Swanson, S. M., & Orjala, J. (2011). Minutissamides A-D, Antiproliferative Cyclic Decapeptides from the Cultured Cyanobacterium *Anabaena minutissima*. *Journal of Natural Products*, 74(7), 1597-1605. <https://doi.org/10.1021/np2002226>
- Kang, H.-S., Sturdy, M., Kronic, A., Kim, H., Shen, Q., Swanson, S. M., & Orjala, J. (2012). Minutissamides E-L, antiproliferative cyclic lipodecapeptides from the cultured freshwater cyanobacterium cf. *Anabaena* sp. *Bioorganic & Medicinal Chemistry*, 20(20), 6134-6143. <https://doi.org/10.1016/j.bmc.2012.08.017>
- Keating, T. A., Marshall, C. G., Walsh, C. T., & Keating, A. E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nature Structural Biology*, 9(7), 522-526. <https://doi.org/10.1038/nsb810>
- Kehr, J., Picchi, D. G., & Dittmann, E. (2011). Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. *Beilstein Journal of Organic Chemistry*, 7, 1622-1635. <https://doi.org/10.3762/bjoc.7.191>
- Kem, M. P., & Butler, A. (2015). Acyl peptidic siderophores: structures, biosyntheses and post-assembly modifications. *BioMetals*, 28(3), 445-459. <https://doi.org/10.1007/s10534-015-9827-y>
- Khaledian, E., Brayton, K. A., & Broschat, S. L. (2020). A Systematic Approach to Bacterial Phylogeny Using Order Level Sampling and Identification of HGT Using Network Science. *Microorganisms*, 8(2), Article ARTN 312. <https://doi.org/10.3390/microorganisms8020312>
- Kinsinger, R. F., Shirk, M. C., & Fall, R. (2003). Rapid surface motility in *Bacillus subtilis* is dependent on extracellular surfactin and potassium ion. *Journal of Bacteriology*, 185(18), 5627-5631. <https://doi.org/10.1128/jb.185.18.5627-5631.2003>

Kramer, D. (2010). Microginin producing proteins and nucleic acids encoding a microginin gene cluster as well as methods for creating novel microginins. *European patent office*, patent #: EP1792981A1

Kreutzer, M. F., Kage, H., & Nett, M. (2012). Structure and Biosynthetic Assembly of Cupriachelin, a Photoreactive Siderophore from the Bioplastic Producer *Cupriavidus necator* H16. *Journal of the American Chemical Society*, *134*(11), 5415-5422. <https://doi.org/10.1021/ja300620z>

Lalonde, S. V., & Konhauser, K. O. (2015). Benthic perspective on Earth's oldest evidence for oxygenic photosynthesis. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(4), 995-1000. <https://doi.org/10.1073/pnas.1415718112>

Lambalot, R. H., Gehring, A. M., Flugel, R. S., Zuber, P., LaCelle, M., Marahiel, M. A., Reid, R., Khosla, C., & Walsh, C. T. (1996). A new enzyme superfamily - The phosphopantetheinyl transferases. *Chemistry & Biology*, *3*(11), 923-936. [https://doi.org/10.1016/S1074-5521\(96\)90181-7](https://doi.org/10.1016/S1074-5521(96)90181-7)

Liaimer, A., Helfrich, E. J. N., Hinrichs, K., Guljamow, A., Ishida, K., Hertweck, C., & Dittmann, E. (2015). Nostopeptolide plays a governing role during cellular differentiation of the symbiotic cyanobacterium *Nostoc punctiforme*. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(6), 1862-1867. <https://doi.org/10.1073/pnas.1419543112>

Liu, L., Budnjo, A., Jokela, J., Haug, B. E., Fewer, D. P., Wahlsten, M., Rouhiainen, L., Permi, P., Fossen, T., & Sivonen, K. (2015). Pseudoaeruginosins, nonribosomal peptides in *Nodularia spumigena*. *ACS Chemical Biology*, *10*(3), 725-733. <https://doi.org/10.1021/cb5004306>

Lopez, D., Fischbach, M. A., Chu, F., Losick, R., & Kolter, R. (2009). Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(1), 280-285. <https://doi.org/10.1073/pnas.0810940106>

Luo, S., Kang, H.-S., Kronic, A., Chen, W.-L., Yang, J., Woodard, J. L., Fuchs, J. R., Cho, S. H., Franzblau, S. G., Swanson, S. M., & Orjala, J. (2015). Trichormamides C and D, antiproliferative cyclic lipopeptides from the cultured freshwater cyanobacterium cf. *Oscillatoria* sp. UIC 10045.

*Bioorganic & Medicinal Chemistry*, 23(13), 3153-3162.

<https://doi.org/10.1016/j.bmc.2015.04.073>

Luzzatto-Knaan, T., Melnik, A. V., & Dorrestein, P. C. (2019). Mass Spectrometry Uncovers the Role of Surfactin as an Interspecies Recruitment Factor. *ACS Chemical Biology*, 14(3), 459-467.

<https://doi.org/10.1021/acscchembio.8b01120>

Mareš, J. (2018). Multilocus and SSU rRNA gene phylogenetic analyses of available cyanobacterial genomes, and their relation to the current taxonomic system. *Hydrobiologia*, 811(1), 19-34. <https://doi.org/10.1007/s10750-017-3373-2>

Mareš, J., Hájek, J., Urajová, P., Kopecký, J., & Hrouzek, P. (2014). A hybrid non-ribosomal peptide/polyketide synthetase containing fatty-acyl ligase (FAAL) synthesizes the  $\beta$ -amino fatty acid lipopeptides puwainaphycins in the Cyanobacterium *Cylindrospermum alatosporum*. *PLoS ONE*, 9(11), e111904. <https://doi.org/10.1371/journal.pone.0111904>

Martinez, J. S., Zhang, G. P., Holt, P. D., Jung, H.-T., Carrano, C. J., Haygood, M. G., & Butler, A. (2000). Self-assembling amphiphilic siderophores from marine bacteria. *Science*, 287(5456), 1245-1247. <https://doi.org/10.1126/science.287.5456.1245>

Martínez-Pérez, C., Mohr, W., Löscher, C. R., Dekaezemacker, J., Littmann, S., Yilmaz, P., Lehnen, N., Fuchs, B. M., Lavik, G., Schmitz, R. A., LaRoche, J., & Kuypers, M. M. M. (2016). The small unicellular diazotrophic symbiont, UCYN-A, is a key player in the marine nitrogen cycle. *Nature Microbiology*, 1(11), Article ARTN 16163.

<https://doi.org/10.1038/NMICROBIOL.2016.163>

Miao, V., Coeffet-LeGal, M. F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S., Ford, R., Parr, I., Bouchard, M., Silva, C. J., Wrigley, S. K., & Baltz, R. H. (2005). Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology-Sgm*, 151, 1507-1523.

<https://doi.org/10.1099/mic.0.27757-0>

Miller, S. R., Castenholz, R. W., & Pedersen, D. (2007). Phylogeography of the thermophilic cyanobacterium *Mastigocladus laminosus*. *Applied and*

*Environmental Microbiology*, 73(15), 4751-4759. <https://doi.org/10.1128/AEM.02945-06>

Moisander, P. H., Beinart, R. A., Hewson, I., White, A. E., Johnson, K. S., Carlson, C. A., Montoya, J. P., & Zehr, J. P. (2010). Unicellular Cyanobacterial Distributions Broaden the Oceanic N<sub>2</sub> Fixation Domain. *Science*, 327(5972), 1512-1514. <https://doi.org/10.1126/science.1185468>

Moisander, P. H., Benavides, M., Bonnet, S., Berman-Frank, I., White, A. E., & Riemann, L. (2017). Chasing after Non-cyanobacterial Nitrogen Fixation in Marine Pelagic Environments. *Frontiers in Microbiology*, 8, Article ARTN 1736. <https://doi.org/10.3389/fmicb.2017.01736>

Moore, R. E., Bornemann, V., Niemczura, W. P., Gregson, J. M., Chen, J. L., Norton, T. R., Patterson, G. M. L., & Helms, G. L. (1989). Puwainaphycin-C, a Cardioactive Cyclic Peptide from the Blue-Green-Alga *Anabaena* Bq-16-1 - Use of Two-Dimensional C-13-C-13 And C-13-N-15 Correlation Spectroscopy in Sequencing the Amino-Acid Units. *Journal of the American Chemical Society*, 111(16), 6128-6132. <https://doi.org/10.1021/ja00198a021>

Mysterud, I., Koller, G., Hoiland, K., Carlsen, T., & Sletten, A. (2016). The lamb disease alveld: Search for fungi and bacteria on *Narthecium ossifragum* foliage and roots. *Small Ruminant Research*, 136, 179-186.

Nagarajan, M., Maruthanayagam, V., & Sundararaman, M. (2012). A review of pharmacological and toxicological potentials of marine cyanobacterial metabolites. *Journal of Applied Toxicology*, 32(3), 153-185. <https://doi.org/10.1002/jat.1717>

Neuhof, T., Schmieder, P., Preussel, K., Dieckmann, R., Pham, H., Bartl, F., & von Döhren, H. (2005). Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium *Hassallia* sp. *Journal of Natural Products*, 68(5), 695-700. <https://doi.org/10.1021/np049671r>

Nogle, L. M., & Gerwick, W. H. (2002). Somocystinamide A, a novel cytotoxic disulfide dimer from a Fijian marine cyanobacterial mixed assemblage. *Organic Letters*, 4(7), 1095-1098. <https://doi.org/10.1021/ol017275j>

Oh, D.-C., Strangman, W. K., Kauffman, C. A., Jensen, P. R., & Fenical, W. (2007). Thalassospiramides A and B, immunosuppressive peptides from

the marine bacterium *Thalassospira* sp.. *Organic Letters*, 9(8), 1525-1528.  
<https://doi.org/10.1021/ol070294u>

Okino, T., Matsuda, H., Murakami, M., & Yamaguchi, K. (1993). Microginin, an Angiotensin-Converting Enzyme-Inhibitor from The Blue-Green-Alga *Microcystis aeruginosa*. *Tetrahedron Letters*, 34(3), 501-504.  
[https://doi.org/10.1016/0040-4039\(93\)85112-a](https://doi.org/10.1016/0040-4039(93)85112-a)

Oren, A., & Ventura, S. (2017). The current status of cyanobacterial nomenclature under the "prokaryotic" and the "botanical" code. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, 110(10), 1257-1269. <https://doi.org/10.1007/s10482-017-0848-0>

Oren, N., Raanan, H., Murik, O., Keren, N., & Kaplan, A. (2017). Dawn illumination prepares desert cyanobacteria for dehydration. *Current Biology*, 27(19), R1056-R1057. <https://doi.org/10.1016/j.cub.2017.08.027>

Orr, P. T., & Jones, G. J. (1998). Relationship between microcystin production and cell division rates in nitrogen-limited *Microcystis aeruginosa* cultures. *Limnology and Oceanography*, 43(7), 11.  
[https://doi.org/DOI: 10.4319/lo.1998.43.7.1604](https://doi.org/DOI:10.4319/lo.1998.43.7.1604)

Paerl, H. W., & Huisman, J. (2009). Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environmental Microbiology Reports*, 1(1), 27-37. <https://doi.org/10.1111/j.1758-2229.2008.00004.x>

Panrace, C., Jokela, J., Sassoon, N., Ganneau, C., Desnos-Ollivier, M., Wahlsten, M., Hurnisto, A., Calteau, A., Bay, S., Fewer, D. P., Sivonen, K., & Gugger, M. (2017). Rearranged Biosynthetic Gene Cluster and Synthesis of Hassallidin E in *Planktothrix serita* PCC 8927 [Article]. *ACS Chemical Biology*, 12(7), 1796-1804. <https://doi.org/10.1021/acscchembio.7b00093>

Partensky, F., Hess, W. R., & Vaulot, D. (1999). Prochlorococcus, a marine photosynthetic prokaryote of global significance. *Microbiology and Molecular Biology Reviews*, 63(1), 106-127.  
<https://doi.org/10.1128/MMBR.63.1.106-127.1999>

Pauwelyn, E., Huang, C.-J., Ongena, M., Léclère, V., Jacques, P., Bleyaert, P., Budzikiewicz, H., Schäfer, M., & Höfte, M. (2013). New Linear Lipopeptides Produced by *Pseudomonas cichorii* SF1-54 Are Involved in Virulence,

Swarming Motility, and Biofilm Formation. *Molecular Plant-Microbe Interactions*, 26(5), 585-598. <https://doi.org/10.1094/MPMI-11-12-0258-R>

Raaijmakers, J. M., De Bruijn, I., Nybroe, O., & Ongena, M. (2010). Natural functions of lipopeptides from *Bacillus* and *Pseudomonas*: more than surfactants and antibiotics. *FEMS Microbiology Reviews*, 34(6), 1037-1062. <https://doi.org/10.1111/j.1574-6976.2010.00221.x>

Ramaswamy, A. V., Sorrels, C. M., & Gerwick, W. H. (2007). Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 70(12), 1977-1986. <https://doi.org/10.1021/np0704250>

Ran, L., Larsson, J., Vigil-Stenman, T., Nylander, J. A. A., Ininbergs, K., Zheng, W., Lapidus, A., Lowry, S., Haselkorn, R., & Bergman, B. (2010). Genome Erosion in a Nitrogen-Fixing Vertically Transmitted Endosymbiotic Multicellular Cyanobacterium. *PLoS ONE*, 5(7), Article ARTN e11486. <https://doi.org/10.1371/journal.pone.0011486>

Rausch, C., Hoof, I., Weber, T., Wohlleben, W., & Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evolutionary Biology*, 7, 78. <https://doi.org/10.1186/1471-2148-7-78>

Raven, J. A. (1998). The twelfth Tansley Lecture. Small is beautiful: the picophytoplankton. *Functional Ecology*, 12(4), 503-513. <https://doi.org/10.1046/j.1365-2435.1998.00233.x>

Řeháková, K., Johansen, J. R., Casamatta, D. A., Xuesong, L., & Vincent, J. (2007). Morphological and molecular characterization of selected desert soil cyanobacteria: Three species new to science including *Mojavia pulchra* gen. et sp nov.. *Phycologia*, 46(5), 481-502.

Reid, R. P., Visscher, P. T., Decho, A. W., Stolz, J. F., Bebout, B. M., Dupraz, C., Macintyre I. G., L., Paerl, H. W., Pinckney, J. L., Prufert-Bebout, L., Steppe, T. F., & DesMarais, D. J. (2000). The role of microbes in accretion, lamination and early lithification of modern marine stromatolites. *Nature*, 406(6799), 989-992.

Reitz, Z. L., Hardy, C. D., Suk, J., Bouvet, J., & Butler, A. (2019). Genomic analysis of siderophore beta-hydroxylases reveals divergent stereocontrol and expands the condensation domain family. *Proceedings of the National*

*Academy of Sciences of the United States of America*, 116(40), 19805-19814. <https://doi.org/10.1073/pnas.1903161116>

Reynolds, C. S., Jaworski, G. H. M., Cmiech, H. A., & Leedale, G. F. (1981). On the Annual Cycle of The Blue-Green-Alga *Microcystis aeruginosa* Kutz Emend Elenkin. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 293(1068), 419-477. <https://doi.org/10.1098/rstb.1981.0081>

Rosenberg, G., Steinberg, N., Oppenheimer-Shaanan, Y., Olender, T., Doron, S., Ben-Ari, J., Sirota-Madi, A., Bloom-Ackermann, Z., & Kolodkin-Gal, I. (2016). Not so simple, not so subtle: the interspecies competition between *Bacillus simplex* and *Bacillus subtilis* and its impact on the evolution of biofilms. *npj Biofilms and Microbiomes*, 2, Article ARTN 15027. <https://doi.org/10.1038/npjbiofilms.2015.27>

Rouhiainen, L., Jokela, J., Fewer, D. P., Urmann, M., & Sivonen, K. (2010). Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). *Chemistry & Biology*, 17(3), 265-273. <https://doi.org/10.1016/j.chembiol.2010.01.017>

Sanchez, L. M., Lopez, D., Vesely, B. A., Della Togna, G., Gerwick, W. H., Kyle, D. E., & Lington, R. G. (2010). Almiramides A-C: Discovery and Development of a New Class of Leishmaniasis Lead Compounds. *Journal of Medicinal Chemistry*, 53(10), 4187-4197. <https://doi.org/10.1021/jm100265s>

Sánchez-Baracaldo, P. (2015). Origin of marine planktonic cyanobacteria. *Scientific Reports*, 5, Article ARTN 17418. <https://doi.org/10.1038/srep17418>

Sánchez-Baracaldo, P., Bianchini, G., Di Cesare, A., Callieri, C., & Christmas, N. A. M. (2019). Insights Into the Evolution of Picocyanobacteria and Phycoerythrin Genes (mpeBA and cpeBA). *Frontiers in Microbiology*, 10, Article ARTN 45. <https://doi.org/10.3389/fmicb.2019.00045>

Sánchez-Baracaldo, P., & Cardona, T. (2020). On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytologist*, 225(4), 1440-1446. <https://doi.org/10.1111/nph.16249>



- Sandy, M., & Butler, A. (2009). Microbial iron acquisition: marine and terrestrial siderophores. *Chemical Reviews*, *109*(10), 4580-4595. <https://doi.org/10.1021/cr9002787>
- Schirrmeister, B. E., de Vos, J. M., Antonelli, A., & Bagheri, H. C. (2013). Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(5), 1791-1796. <https://doi.org/10.1073/pnas.1209927110>
- Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A., Cai, F., Tandeau de Marsac, N., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K. W., Han, C. S., Rubin, E. M., Eisen, J. A., Woyke, T., Gugger, M., & Kerfeld, C. A. (2013). Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 1053-1058. <https://doi.org/10.1073/pnas.1217107110>
- Shishido, T. K., Jokela, J., Kolehmainen, C.-T., Fewer, D. P., Wahlsten, M., Wang, H., Rouhiainen, L., Rizzi, E., De Bellis, G., Permi, P., & Sivonen, K. (2015). Antifungal activity improved by coproduction of cyclodextrins and anabaenolysins in Cyanobacteria. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(44), 13669-13674. <https://doi.org/10.1073/pnas.1510432112>
- Skinnider, M. A., Johnston, C. W., Gunabalasingam, M., Merwin, N. J., Kieliszek, A. M., MacLellan, R. J., Li, H., Ranieri, M. R. M., Webster, A. L. H., Cao, M. P. T., Pfeifle, A., Spencer, N., To, Q. H., Wallace, D. P., Dejong, C. A., & Magarvey, N. A. (2020). Comprehensive prediction of secondary metabolite structure and biological activity from microbial genome sequences. *Nature Communications*, *11*(1), 6058. <https://doi.org/10.1038/s41467-020-19986-1>
- Sritharan, M. (2016). Iron Homeostasis in *Mycobacterium tuberculosis*: Mechanistic Insights into Siderophore-Mediated Iron Uptake. *Journal of Bacteriology*, *198*(18), 2399-2409. <https://doi.org/10.1128/JB.00359-16>
- Stachelhaus, T., Hüser, A., & Marahiel, M. A. (1996). Biochemical characterization of peptidyl carrier protein (PCP), the thiolation domain of

multifunctional peptide synthetases. *Chemistry & Biology*, 3(11), 913-921. [https://doi.org/10.1016/S1074-5521\(96\)90180-5](https://doi.org/10.1016/S1074-5521(96)90180-5)

Stachelhaus, T., & Marahiel, M. A. (1995). Modular structure of genes encoding multifunctional peptide synthetases required for non-ribosomal peptide synthesis. *FEMS Microbiology Letters*, 125(1), 3-14. <https://doi.org/10.1111/j.1574-6968.1995.tb07328.x>

Stachelhaus, T., Mootz, H. D., & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology*, 6(8), 493-505. [https://doi.org/10.1016/S1074-5521\(99\)80082-9](https://doi.org/10.1016/S1074-5521(99)80082-9)

Stewart, A. K., Ravindra, R., Van Wagoner, R. M., & Wright, J. L. C. (2018). Metabolomics-Guided Discovery of Microginin Peptides from Cultures of the Cyanobacterium *Microcystis aeruginosa*. *Journal of Natural Products*, 81(2), 349-355. <https://doi.org/10.1021/acs.jnatprod.7b00829>

Straight, P. D., Willey, J. M., & Kolter, R. (2006). Interactions between *Streptomyces coelicolor* and *Bacillus subtilis*: Role of surfactants in raising aerial structures. *Journal of Bacteriology*, 188(13), 4918-4925. <https://doi.org/10.1128/JB.00162-06>

Strangman, W. K., & Wright, J. L. C. (2016). Microginins 680, 646, and 612—new chlorinated AhoA-containing peptides from a strain of cultured *Microcystis aeruginosa*. *Tetrahedron Letters*, 57(16), 1801-1803. <https://doi.org/10.1016/j.tetlet.2016.03.039>

Tang, E. P. Y., Tremblay, R., & Vincent, W. F. (1997). Cyanobacterial dominance of polar freshwater ecosystems: Are high-latitude mat-formers adapted to low temperature?. *Journal of Phycology*, 33(2), 171-181. <https://doi.org/10.1111/j.0022-3646.1997.00171.x>

Taniguchi, M., Nunnery, J. K., Engene, N., Esquenazi, E., Byrum, T., Dorrestein, P. C., & Gerwick, W. H. (2010). Palmyramide A, a Cyclic Depsipeptide from a Palmyra Atoll Collection of the Marine Cyanobacterium *Lyngbya majuscula*. *Journal of Natural Products*, 73(3), 393-398. <https://doi.org/10.1021/np900428h>

Tomek, P., Hrouzek, P., Kuzma, M., Sýkora, J., Fišer, R., Černý, J., Novák, P., Bártová, S., Šimek, P., Hof, M., Kavan, D., & Kopecký, J. (2015). Cytotoxic Lipopeptide Muscotoxin A, Isolated from Soil Cyanobacterium

- Desmonostoc muscorum*, Permeabilizes Phospholipid Membranes by Reducing Their Fluidity. *Chemical Research in Toxicology*, 28(2), 216-224. <https://doi.org/10.1021/tx500382b>
- Traxler, M. F., & Kolter, R. (2015). Natural products in soil microbe interactions and evolution. *Natural Product Reports*, 32(7), 956-970. <https://doi.org/10.1039/c5np00013k>
- Tripathi, A., Puddick, J., Prinsep, M. R., Lee, P. P. F., & Tan, L. T. (2010). Hantupeptins B and C, cytotoxic cyclodepsipeptides from the marine cyanobacterium *Lyngbya majuscula*. *Phytochemistry*, 71(2-3), 307-311. <https://doi.org/10.1016/j.phytochem.2009.10.006>
- van Gestel, J., Vlamakis, H., & Kolter, R. (2015). From Cell Differentiation to Cell Collectives: *Bacillus subtilis* Uses Division of Labor to Migrate. *Plos Biology*, 13(4), 29, Article e1002141. <https://doi.org/10.1371/journal.pbio.1002141>
- Vestola, J., Shishido, T. K., Jokela, J., Fewer, D. P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L., & Sivonen, K. (2014). Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. *Proceedings of the National Academy of Sciences of the United States of America*, 111(18), E1909-1917. <https://doi.org/10.1073/pnas.1320913111>
- Vincent, W. F., Downes, M. T., Castenholz, R. W., & Howard-Williams, C. (1993). Community Structure and Pigment Organization of Cyanobacteria-Dominated Microbial Mats in Antarctica. *European Journal of Phycology*, 28(4), 213-221. <https://doi.org/10.1080/09670269300650321>
- Völker, C., & Wolf-Gladrow, D. A. (1999). Physical limits on iron uptake mediated by siderophores or surface reductases. *Marine Chemistry*, 65(3-4), 227-244. [https://doi.org/10.1016/S0304-4203\(99\)00004-3](https://doi.org/10.1016/S0304-4203(99)00004-3)
- Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., & Sivonen, K. (2014). Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(25), 9259-9264. <https://doi.org/10.1073/pnas.1401734111>
- Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., & Holak, T. A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier

- domains of modular peptide synthetases. *Structure*, 8(4), 407-418.  
[https://doi.org/10.1016/S0969-2126\(00\)00120-9](https://doi.org/10.1016/S0969-2126(00)00120-9)
- Welker, M., & von Döhren, H. (2006). Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiology Reviews*, 30(4), 530-563. <https://doi.org/10.1111/j.1574-6976.2006.00022.x>
- Wenzel, S. C., Kunze, B., Höfle, G., Silakowski, B., Scharfe, M., Blöcker, H., & Müller, R. (2005). Structure and biosynthesis of myxochromides S1-3 in *Stigmatella aurantiaca*: Evidence for an iterative bacterial type I polyketide synthase and for module skipping in nonribosomal peptide biosynthesis. *Chembiochem*, 6(2), 375-385. <https://doi.org/10.1002/cbic.200400282>
- Whitton, B. A. (2012). *Ecology of Cyanobacteria II : Their Diversity in Space and Time*. Springer Netherlands. <https://doi.org/10.1007/978-94-007-3855-3>
- Williams, L., Loewen-Schneider, K., Maier, S., & Büdel, B. (2016). Cyanobacterial diversity of western European biological soil crusts along a latitudinal gradient. *FEMS Microbiology Ecology*, 92(10), Article ARTN fiw157. <https://doi.org/10.1093/femsec/fiw157>
- Willis, A., & Woodhouse, J. N. (2020). Defining Cyanobacterial Species: Diversity and Description Through Genomics. *Critical Reviews in Plant Sciences*, 39(2), 101-124.  
<https://doi.org/10.1080/07352689.2020.1763541>
- Wrasidlo, W., Mielgo, A., Torres, V. A., Barbero, S., Stoletov, K., Suyama, T. L., Klemke, R. L., Gerwick, W. H., Carson, D. A., & Stupack, D. G. (2008). The marine lipopeptide somocystinamide A triggers apoptosis via caspase 8. *Proceedings of the National Academy of Sciences of the United States of America*, 105(7), 2313-2318. <https://doi.org/10.1073/pnas.0712198105>
- Yu, H.-B., Glukhov, E., Li, Y., Iwasaki, A., Gerwick, L., Dorrestein, P. C., Jiao, B.-H., & Gerwick, W. H. (2019). Cytotoxic Microcolin Lipopeptides from the Marine Cyanobacterium *Moorea producens*. *Journal of Natural Products*, 82(9), 2608-2619. <https://doi.org/10.1021/acs.jnatprod.9b00549>

Zachow, C., Jahanshah, G., de Bruijn, I., Song, C., Ianni, F., Pataj, Z., Gerhardt, H., Pianet, I., Lammerhofer, M., Berg, G., Gross, H., & Raaijmakers, J. M. (2015). The Novel Lipopeptide Poaeamide of the Endophyte *Pseudomonas poae* RE star 1-1-14 Is Involved in Pathogen Suppression and Root Colonization. *Molecular Plant-Microbe Interactions*, 28(7), 800-810. <https://doi.org/10.1094/MPMI-12-14-0406-R>

## **4. Attached Publications**

### **4.1. Paper I.**



## A liquid chromatography–mass spectrometric method for the detection of cyclic $\beta$ -amino fatty acid lipopeptides



Petra Urajová<sup>a</sup>, Jan Hájek<sup>a,b,c</sup>, Matti Wahlsten<sup>d</sup>, Jouni Jokela<sup>d</sup>, Tomáš Galica<sup>a,b</sup>, David P. Fewer<sup>d</sup>, Andreja Kust<sup>a,b,c</sup>, Eliška Zapomělová-Kozlíková<sup>c</sup>, Kateřina Delawská<sup>a,b</sup>, Kaarina Sivonen<sup>d</sup>, Jiří Kopecký<sup>a,b</sup>, Pavel Hrouzek<sup>a,b,\*</sup>

<sup>a</sup> Centre Algatech, Institute of Microbiology, The Czech Academy of Sciences (CAS), Opatovický mlýn, 379 81, Třeboň, Czech Republic

<sup>b</sup> University of South Bohemia, Faculty of Science, Branšovská 1760, České Budějovice, Czech Republic

<sup>c</sup> Biology Centre of CAS, v.v.i., Institute of Hydrobiology, Na Sádkách 7, 370 05 České Budějovice, Czech Republic

<sup>d</sup> Department of Food and Environmental Sciences, Viikki Biocenter 1, University of Helsinki, FI-00014 Helsinki, Finland

### ARTICLE INFO

#### Article history:

Received 29 October 2015

Received in revised form 12 January 2016

Accepted 2 February 2016

Available online 15 February 2016

#### Keywords:

Lipopeptides

Bacteria

Cyanobacteria

LC-HRMS

Fatty acid

Peptide

### ABSTRACT

Bacterial lipopeptides, which contain  $\beta$ -amino fatty acids, are an abundant group of bacterial secondary metabolites exhibiting antifungal and/or cytotoxic properties. Here we have developed an LC-HRMS/MS method for the selective detection of  $\beta$ -amino fatty acid containing cyclic lipopeptides. The method was optimized using the lipopeptides iturin A and puwainaphycin F, which contain fatty acids of similar length but differ in the amino acid composition of the peptide cycle. Fragmentation energies of 10–55 eV were used to obtain the amino acid composition of the peptide macrocycle. However, fragmentation energies of 90–130 eV were used to obtain an intense fragment specific for the  $\beta$ -amino fatty acid ( $C_nH_{2n+2}N^+$ ). The method allowed the number of carbons and consequently the length of the  $\beta$ -amino fatty acid to be estimated. We identified 21 puwainaphycin variants differing in fatty acid chain in the crude extract of cyanobacterium *Cylindrospermum alatosporum* using this method. Analogously 11 iturin A variants were detected. The retention time of the lipopeptide variants showed a near perfect linear dependence ( $R^2 = 0.9995$ ) on the length of the fatty acid chain in linear separation gradient which simplified the detection of minor variants. We used the method to screen 240 cyanobacterial strains and identified lipopeptides from 8 strains. The HPLC-HRMS/MS method developed here provides a rapid and easy way to detecting novel variants of cyclic lipopeptides.

© 2016 Elsevier B.V. All rights reserved.

### 1. Introduction

Bacteria are a well-known source of low-molecular weight secondary metabolites [1]. These compounds represent an important source of novel chemical structures for pharmacy and biotechnology [1,2]. Intensive pharmacological screenings for novel chemical structures have identified immense number of novel microbial secondary metabolites [1,2]. Many microbial secondary metabolites are peptides or have a peptidic portion and are produced through the action of non-ribosomal peptide synthetases (NRPS) [3]. Based on the genome mining studies it is assumed that the NRPS biosynthetic clusters of known chemical structures represent only a small portion of the total cluster number found [4]. Taken together the

known metabolites represent only a small portion of the real microbial metabolite chemical diversity.

Cyclic lipopeptides are an important class of bacterial secondary metabolites and include many antibiotics and antifungal agents [1,5–9]. Their structures contain peptide macrocycle with a fatty acyl residue side chain. There are several basic structural types of bacterial cyclic lipopeptides. The carboxyl group of the fatty acid can form a peptide bond with an amino-acid residue situated on the side chain of the lipopeptide molecule as found in bacterial antibiotics fengycin [6] and daptomycin [7] and the anti-fungal lipopeptides hassallidins [8] and balticidins [9]. Alternatively, the fatty acid can be incorporated into the peptide backbone *via* cyclization (most frequently on  $C_2$  or  $C_3$  carbon) by amine- or hydroxyl-groups [10–12]. The 3-OH substituted fatty acid of surfactin is bound by a peptide bond between the carboxyl and *via* ester bond by the hydroxyl group [13]. If an amino group is introduced during biosynthesis the fatty acid residue allows the formation of two

\* Corresponding author at: Centre Algatech, Institute of Microbiology, The Czech Academy of Sciences (CAS), Opatovický mlýn, 379 81, Třeboň, Czech Republic.  
E-mail address: [hrouzekp@gmail.com](mailto:hrouzekp@gmail.com) (P. Hrouzek).

<http://dx.doi.org/10.1016/j.chroma.2016.02.013>

0021-9673/© 2016 Elsevier B.V. All rights reserved.

peptide bonds (or an isopeptide bond in case of  $\beta$ -NH<sub>2</sub>) with C- and N-terminus of the peptide forming a cycle [10,12].

The structure of iturin family of lipopeptide antibiotics was described already 40 years ago [5]. However, novel structural types of lipopeptides are still reported highlighting their rich chemical diversity in the nature and especially in cyanobacteria [13–15]. Cyanobacteria produce a large diversity of  $\beta$ -amino fatty acid lipopeptides varying in the type and number of amino acids in the cycle as well as the length and functionalization of the fatty acid moiety [16–25]. Usually C<sub>8</sub>–C<sub>18</sub> long fatty acid can be found in cyanobacterial lipopeptides sometimes methylated [18,20], hydroxylated [18,19], halogenated [12,18] or unsaturated [23]. Cyanobacterial cyclic  $\beta$ -amino fatty acid lipopeptides display a broad range of bioactivities including cytotoxicity [16–22], cytolytic activity [26], as well as antifungal and/or antibacterial activity [24,25,27]. Some cyanobacterial lipopeptides permeabilize human plasma membranes and are potentially toxic to humans [20,21].

To date no selective screening method allows the targeted screening of novel lipopeptides from bacteria. Here we report the development of a rapid and selective liquid chromatographic mass spectrometric method for  $\beta$ -amino fatty acid lipopeptides analysis based on the detection of a characteristic fatty acid immonium ion. This method allows rapid screening for novel lipopeptides from various biological samples.

## 2. Experimental

### 2.1. Materials

Methanol for extraction (p.a.) was obtained from Analytika (Czech Republic), acetonitrile and water for HPLC-HRMS/MS analyzes were obtained either from Sigma–Aldrich (Germany) or Merck (New Jersey, USA) and were of LC–MS grade purity. The iturin A standard was obtained from *Bacillus subtilis* ( $\geq 95\%$ ) and contained mainly C<sub>14</sub>–C<sub>15</sub> fatty acids (with traces of C<sub>13</sub>–C<sub>17</sub>) according to manufacturer instructions (Sigma, Germany). 4-Methyl-Ahdoapuwainaphycin F and anabaenolysin A were isolated from crude cyanobacterial biomass by HPLC from *Cylindrospermum alatosporum* CCALA 988 and *Anabaena* sp. XPORK 15F, respectively. Hexakis(1H,1H,2H-perfluoroethoxy) phosphazene (LockMass Tuning Mix ES-TOF 622 Da, 97%) was purchased from ABCR GmbH & Co., KG, Germany).

### 2.2. Methods and procedure

#### 2.2.1. Cultivation of cyanobacterial biomass

*C. alatosporum* CCALA 988 and strains for lipopeptide screening were cultivated in 350 mL glass tubes or Erlenmeyer flasks on liquid Allen Arnon medium [28] and bubbled with 1.5% CO<sub>2</sub>-enriched air at constant temperature of 26 °C, under continuous illumination of 50  $\mu\text{mol m}^{-2} \text{s}^{-1}$ . Strains used in lipopeptide screening originate from culture collections of the Institute of Hydrobiology, Biological Centre CAS and Culture collection of phototrophic microorganisms (CCALA) and the University of Helsinki (HAMBI Culture collection). *Anabaena* sp. XPORK 15F was grown in Z8 medium without a nitrogen source [29] under continuous light (8–14  $\mu\text{mol m}^{-2} \text{s}^{-1}$ ) at 20–25 °C for 20–40 days. The cultures were harvested by centrifugation (1250  $\times g$ , 15 min) after 5–7 days of cultivation, stored at –70 °C and freeze-dried (Leybold–Heraeos freeze-drier Lyovac GT 3).

#### 2.2.2. Extraction

200 mg of dried biomass of *C. alatosporum* CCALA 988 and *Anabaena* sp. XPORK 15F was extracted with 10 mL of methanol/water (70/30, v/v) for 1 h and centrifuged for 10 min.

at 1920  $\times g$ . The supernatant was evaporated until dryness on rotary evaporator, while the temperature did not exceed 40 °C. The residue was redissolved in 1 mL of extraction solvent. 5–30 mg of lyophilized biomass from the 240 cyanobacterial strains was homogenized with sand in Eppendorf tube and 100  $\mu\text{l}$  of 70% MeOH. The samples were centrifuged (12,000  $\times g$ , 7 min.) in a Minispin Eppendorf centrifuge and transferred into vials with inserts for LC-HRMS analysis.

#### 2.2.3. HRMS/MS analysis by direct infusion

Methanolic solutions (acidified by 0.1% HCOOH) of iturin A, puwainaphycin F and anabaenolysin A were analyzed by direct infusion HRMS/MS using collision energy varying from 0 to 200 eV. The experiments were performed in triplicate.

#### 2.2.4. HPLC-HRMS/MS analysis and MS/MS experiments

A standard solution of iturin A (0.1 mg/mL), *C. alatosporum* CCALA 988 and *Anabaena* sp. XPORK 15F 70% methanol extracts were analyzed on a Dionex UltiMate 3000 UHPLC+ (Thermo Scientific, Sunnyvale, CA, USA) equipped with a diode-array detector. Separation of compounds was performed on reversed phase C<sub>18</sub> column (Phenomenex Kinetex, 150  $\times$  4.6 mm, 2.6  $\mu\text{m}$ , Torrance, CA, USA) using H<sub>2</sub>O (A)/acetonitrile (B) both containing 0.1% HCOOH as a mobile phase with the flow rate of 0.5 mL  $\text{min}^{-1}$ . The gradient was as follows: A/B 85/15 (0 min), 85/15 (in 1 min), 0/100 (in 39 min), 0/100 (in 44 min) and 85/15 (in 49 min). The HPLC was connected to an Impact HD high resolution mass spectrometer (Bruker, Billerica, Massachusetts, USA) with electrospray ionization. The following settings were used: drying temperature 200 °C; drying gas flow 12 l  $\text{min}^{-1}$ ; nebulizer gas pressure 3 bar; capillary voltage 3.8 kV; endplate offset 500 V. The spectra were collected in the range  $m/z$  20–2000 with spectra rate of 3 Hz. Collision energy alternated from 35 eV or 50 eV to 100 eV. Mass spectrometer was calibrated with sodium formate clusters and LockMass Tuning Mix ES-TOF (622 Da) was used as the lockmass calibrant. The formulas of obtained molecular peaks and fragments were calculated using Smart Formula in Bruker Compass DataAnalysis software (version 4.2) and MassLynx software (version V4.1). Only fragments with maximal  $m/z$  difference 1 mDa between theoretical and measured value were taken into consideration. This corresponds to ppm values below 10 ppm in MS<sub>2</sub> for all detected ions. However, the majority of the detected ions had values were below 5 ppm (Tables 1–3).

240 strains of cyanobacteria were analyzed by the HPLC-HRMS/MS method as described above. The resulting data were analyzed for characteristic  $\beta$ -amino fatty acid immonium ion from C<sub>8</sub> to C<sub>16</sub> both saturated (C<sub>x</sub>H<sub>(2x+2)</sub>N<sup>+</sup>) and with one double bond (C<sub>x</sub>H<sub>2x</sub>N<sup>+</sup>).

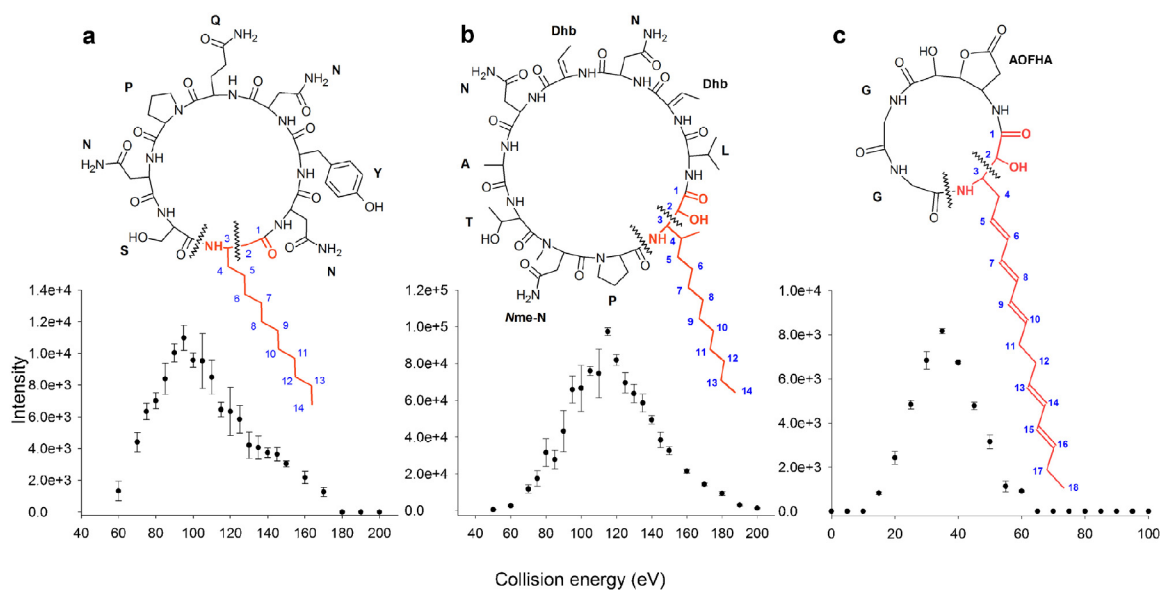
## 3. Results

### 3.1. HRMS/MS analysis of selected lipopeptides

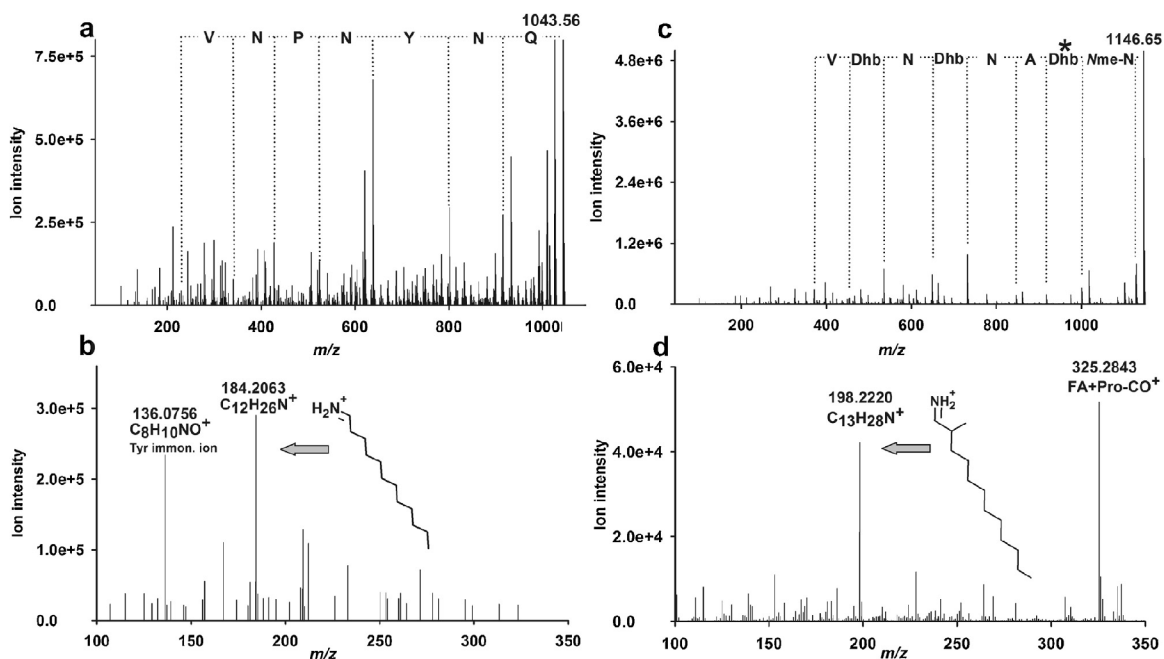
The HRMS/MS method was optimized by obtaining diagnostic fragment ion from iturin A prepartate ( $m/z$  1043.5520 [M+H]<sup>+</sup>), puwainaphycin F ( $m/z$  1146.6517 [M+H]<sup>+</sup>) and anabaenolysin A ( $m/z$  559.2762 [M+H]<sup>+</sup>) isolated from cyanobacterial crude extracts in direct infusion mode using collision energies from 0 to 200 eV. All of the selected lipopeptides possess at least C<sub>14</sub> unit long  $\beta$ -amino acid chain (Fig. 1), but differ in the composition of their peptide core structure and the level of the fatty acid chain saturation.

Iturin A and puwainaphycin F exhibit similar fragmentation patterns. Lower collision energies ( $\sim 50$  eV) led to fragments indicating loss of amino acids, while higher collision energies resulted in formation of diagnostic  $\beta$ -amino fatty acid immonium ions for both tested compounds (Fig. 2). The diagnostic  $\beta$ -amino fatty acid ion for





**Fig. 1.** The structure of the lipopeptides included in this study and the intensity of their diagnostic FA immonium ion  $C_nH_{2n}N^+$  at varying collision energies. a) Ahteia-Itu A provided a diagnostic fragment at  $m/z$  184.2064 with a maximal intensity at 95 eV. b) 4-Methyl-Ahteia-Puw F provided a diagnostic fragment at  $m/z$  198.2220 with a maximal intensity at 115 eV and c) in Ahopa-Abl A the diagnostic fragment at  $m/z$  230.1900 was found at a collision energy of 35 eV. The modified  $\beta$ -amino fatty acids are highlighted in red and the diagnostic immonium FA fragments lacking the  $C_1$  and  $C_2$  carbons are separated by wavy line. Itu stands for iturin, Puw stands for puwainaphycin, Abl stands for anaabanolysin.



**Fig. 2.** Fragmentation of Ahteia-Itu A (a, b) and 4-methyl-Ahteia-Puw F (c, d) at different collision energies. a, c) Fragmentation at 50 eV showing the peptide sequence of the macrocycle. b, d) Fragmentation at 100 eV showing the presence of modified  $\beta$ -amino fatty acid diagnostic immonium ion. The interpretation of the amino acid sequence of 4-methyl-Ahteia-Puw F was derived from  $[M - H_2O + H]^+$  and thus the loss of additional Dhb instead of Thr takes place (marked by asterisk). Itu stands for iturin and Puw stands for puwainaphycin.

iturin A has a  $m/z$  184.2063 with the formula  $C_{12}H_{26}N^+$ . It reached maximal absolute intensity at 95 eV and in the range 85–135 eV it was the base peak of MS/MS spectra (Fig. 1). Although the diagnostic ion consisted of 12 carbon atoms, the iturin A fatty acid chain itself is 14 carbon atoms. We obtained similar results for puwainaphycin

F. The diagnostic  $\beta$ -amino fatty acid ion was at  $m/z$  198.2216 with formula  $C_{13}H_{28}N^+$ . The maximal intensity was reached at 120 eV (Fig. 1) and it was the most intense ion in the range 110–160 eV. Despite having the same length of the fatty acid chain ( $C_{14}$ ) the diagnostic fragments of iturin A and puwainaphycin F differ by a

–CH<sub>2</sub>, which is due to the methylation of the puwainaphycin F on the carbon C4. Both diagnostic fragments lacked the first two carbons of the fatty acid chain (Fig. 1). This made it possible to recognize the presence of β-amino fatty acid chain in the molecule, but it was impossible to determine the exact fatty acid chain length due to branching occurring in some lipopeptides. Although amino acids adjoining the β-amino fatty acid chain differ in iturin A and puwainaphycin F, it is possible to observe fatty acid diagnostic ions by using collision energy of 100 eV for both lipopeptides.

Anabaenolysin A exhibited a different fragmentation behavior. The immonium ion of the fatty acid chain (*m/z* 230.1903) with formula C<sub>16</sub>H<sub>24</sub>N<sup>+</sup> appeared at a collision energy of 20 eV, reached its maximum abundance at 35 eV (relative intensity 67%) and was no longer detectable at 60 eV, but was not the base peak in any MS/MS spectra (Fig. 1).

### 3.2. HPLC-HRMS/MS analysis of iturin A variants

The HPLC-HRMS/MS analysis of the commercial iturin A standard, was composed mainly of variants with a fatty acid chain length C<sub>14</sub>–C<sub>15</sub> (traces of C<sub>13</sub>–C<sub>17</sub>) according to information provided by manufacturer, revealed the presence of eleven protonated molecules in the range *m/z* 987–1086 (Table 1). MS/MS spectra of these ions obtained using low collision energy (50 eV) confirmed the amino-acid sequence of iturin family members (Fig. 2a). Eight variants had common structure FA-Asn-Tyr-Asn-Gln-Pro-Asn-Ser as described for iturin A and in three variants the sequence FA-Gln-Tyr-Asn-Gln-Pro-Asn-Ser were found, showing an exchange of Asn to Gln at the position adjacent to fatty acid chain.

Diagnostic β-amino fatty acid fragment ions of the variants obtained using collision energy 100 eV identified the fatty acid moiety of the lipopeptide (Fig. 2b). The general formula was C<sub>x</sub>H<sub>(2x+2)</sub>N<sup>+</sup>, where *x* represents the length of fatty acid chain reduced of the first and the second carbon. In all cases, the diagnostic β-amino fatty acid fragment ions represent the dominant fragment ions in the MS/MS spectra (Fig. S1a) and are well recognizable by the extracted ion chromatogram. We were able to detect minor iturin A variants with C<sub>10</sub>–C<sub>17</sub> fatty acid chains in the sample analyzed using this approach (Table 1).

All iturin A variants eluted in region between 12 and 23 min, corresponding to linear increase of organic solvent in the mobile phase. When analyzed by linear regression the retention times of lipopeptides showed perfectly linear relationship with the length of amino acid chain. The retention time *y* of eight iturin A variants increased with equation  $y = 1.475x - 2.45$  (*x* = number of fatty acid carbons) and  $R^2 = 0.9941$  and only slightly worse fit was found in Asn/Gln substituted variants ( $y = 1.300x + 0.1$ ,  $R^2 = 0.9826$ ). While the change in presence of one –CH<sub>2</sub> group in fatty acid chain caused marked shift in retention time, the Asn to Gln substitution in the cycle had almost no effect.

### 3.3. HPLC-HRMS/MS analysis of *Cylindrospermum alatosporum* CCALA 988 extract

We have previously shown that cyanobacterium *C. alatosporum* CALA 988 produces the cytotoxic lipopeptides puwainaphycins [12,20]. Now we further analyzed the crude extract of *C. alatosporum* CALA 988 in order to test the method for lipopeptide screening in a complex matrix. The HPLC-HRMS/MS analyzes of the extract allowed us to detect 21 molecular ions in the region *m/z* 1090–1200 (Table 2). Each of these ions could be attributed to puwainaphycin F/G variants based on the amino-acid sequence obtained by fragmentation in lower collision energy at 50 eV (Fig. 2c). Fragmentation spectra obtained at a collision energy of 100 eV revealed diagnostic fragment ions C<sub>x</sub>H<sub>(2x+2)</sub>N<sup>+</sup> as in the case of iturin A. Based on the presence of the diagnostic immonium ion C<sub>x</sub>H<sub>(2x+2)</sub>N<sup>+</sup> we

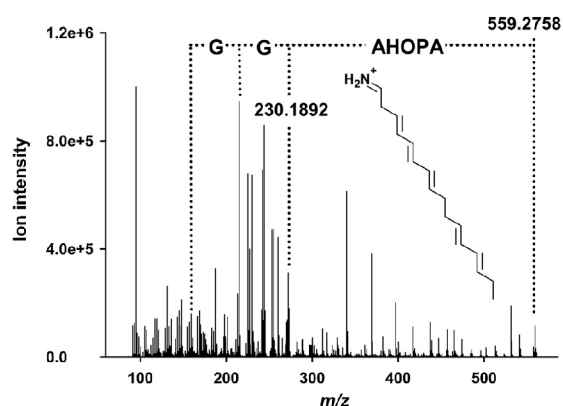


Fig. 3. MS/MS spectrum of anabaenolysin A at collision energy of 35 eV. The presence of modified β-amino fatty acid diagnostic immonium ion C<sub>16</sub>H<sub>24</sub>N<sup>+</sup> and fragments corresponding to amino acids loses are highlighted.

were able to prove the presence of puwainaphycin F/G variants with fatty acid chain lengths ranging from C<sub>10</sub> to C<sub>17</sub> (Fig. S1b). Some puwainaphycin variants are hydroxylated or chlorinated on the fatty acyl chain [12]. All four detected fatty acid substituted puwainaphycin variants (chloro/hydroxy-Ahdoa puwainaphycin F and chloro/hydroxy-Ahtea puwainaphycin F) have provided common diagnostic fragment C<sub>x</sub>H<sub>2x</sub>N<sup>+</sup> when fragmentation energies 100–150 eV were used (Fig. S2). These sum formula of these fragments perfectly matches with presence of a double bond created after dehydration/dechlorination of the fatty acid chain during fragmentation (Table 2).

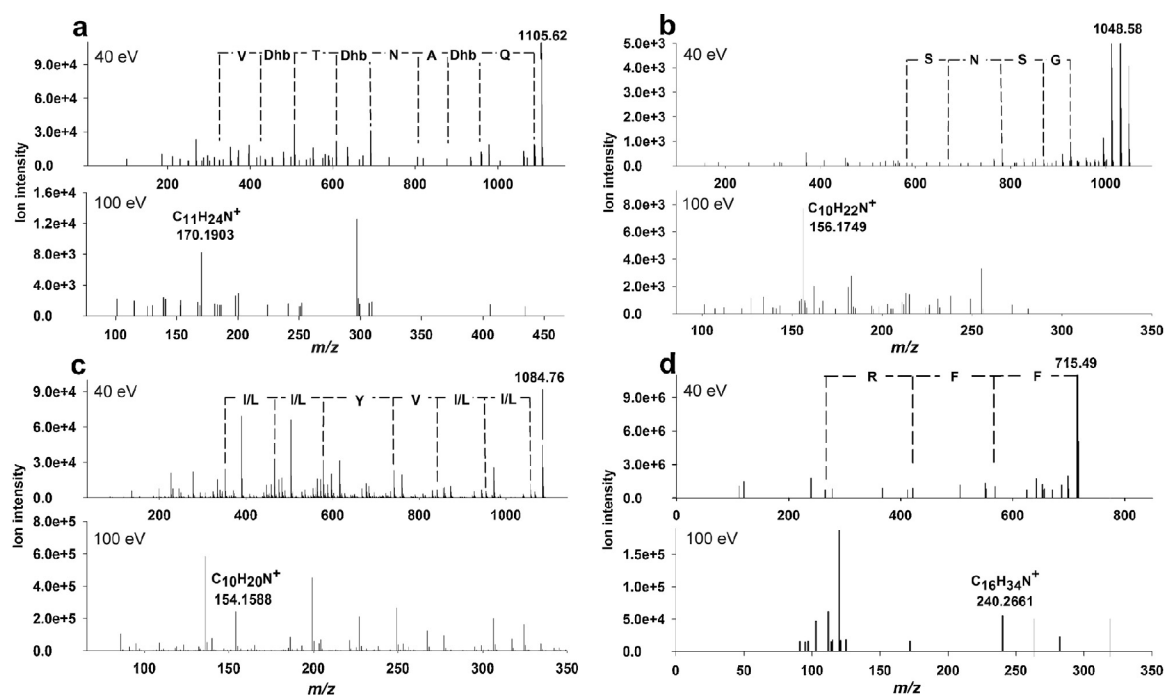
The retention time of puwainaphycin F/G variants on C<sub>18</sub> reversed phase column is only partially dependent on the amino acid composition but strongly dependent on the length of the fatty acid chain as found for iturin A variants. While the shift in retention time of 4-methyl-Ahdoa-puwainaphycin F and 4-methyl-Ahdoa-puwainaphycin G (differing in Asn–Gln substitution) is almost negligible (max 0.1 min), the retention time distinctly shifts with the length of fatty acid chain. A similar shift in retention time with increasing fatty acid chain length was also observed for chlorinated and hydroxylated variants. The relationship between retention time and the length of the fatty acid chain within the gradient used is almost perfectly linear with only a slight deviation from the linear trend in the 4-methyl-Ahhea-puwainaphycin F variant. The increasing chain length allows rotation of the fatty acid chain resulting in intramolecular interactions affecting the retention time. The regression curve for this relationship was determined excluding 4-methyl-Ahhea-puwainaphycin-F  $y = 1.516x + 0.434$  with  $R^2$  value 0.9999 ( $y = 1.440x + 1.323$  with  $R^2$  value 0.9959 including 4-methyl-Ahhea-puwainaphycin F).

### 3.4. HPLC-HRMS/MS analysis of *Anabaena* sp. XPORK 15F extract

In order to test the method also on lipopeptides containing polyunsaturated β-amino fatty acids we analyzed 70% methanolic extract of cyanobacterium *Anabaena* sp. XPORK 15F previously shown to produce anabaenolysin variants [23]. The separation was performed by the same HPLC-HRMS/MS method as for the puwainaphycin and iturin variants. Since anabaenolysin A exhibited different fragmentation behavior in direct infusion experiments in comparison to puwainaphycin and iturin (Fig. 3) we applied a fragmentation energy of 35 eV.

Six anabaenolysin lipopeptides in the range of *m/z* 550–600 were found in the XPORK 15F extract using HPLC-HRMS





**Fig. 4.** Examples of novel putative lipopeptides detected in cyanobacterial strains with unknown metabolite profiles. The partial amino acid sequence was obtained by fragmentation at 40 eV. (a) Novel Puw variant ( $m/z$  1105.62) detected in strain *Anabaena* sp. SMIX. The interpretation of the amino acid sequence was derived from  $[M - H_2O + H]^+$  and thus the loss of additional Dhb instead of Thr takes place (marked by asterisk). (b) Novel lipopeptide ( $m/z$  1048.58) detected in strain *Cylindrospermum* sp. CCALA 993. (c) Novel lipopeptide ( $m/z$  1084.76) detected in strain *Cylindrospermum* sp. CCALA 1000. (d) Novel lipopeptide ( $m/z$  715.49) detected in strain *Anabaena* sp. A-Pastvisko11. Puw stands for puwainaphycin.

### 3.5. Screening for lipopeptides in cyanobacteria with unknown metabolic profiles

We identified new lipopeptides from 240 cyanobacterial strains with an unknown metabolite content using the new method leading to the detection of 2 known and 3 unknown puwainaphycin variants and/or 9 putative lipopeptide structures in 8 cyanobacterial strains.

In *Cylindrospermum* spp. strains CCALA 993 and 994 we detected 4-methyl-Ahdoa-puwainaphycin F by presence of characteristic  $C_{12}$   $\beta$ -amino fatty acid immonium ion ( $m/z$  170.9005). In the third puwainaphycin producing strain (*Anabaena* sp. SMIX) we detected 4-methyl-Ahdoa-puwainaphycin F, 4-methyl-Ahtea-puwainaphycin F and three novel puwainaphycin variants ( $m/z$  1105.3, 1177.4 and 1190.4). In variant 1105.3 characteristic fatty acid immonium ion fragment of elemental composition  $C_{16}H_{34}N^+$  was detected at 100 eV (Fig. 4a) and peptide sequence of Val-Dhb-Thr-Dhb-Asn-Ala-Dhb-Glu was obtained at 40 eV (Fig. 4a). In variants 1177.4 and 1190.4 the fatty acid immonium ion fragment  $C_{15}H_{30}N^+$  was detected and their peptide sequences were identical to 1105 and puwainaphycin F, respectively.

We detected 10 putative novel lipopeptide structures in the 240 strains analyzed (Table S2). These compounds contained the characteristic  $\beta$ -amino fatty acid immonium ion in fragmentation spectra obtained at 100 eV and at least 3 amino-acid residues were detected in fragmentation at 40 eV (see Table S2). Examples of fragmentation spectra of three structures are given in Fig. 4b–d. In *Cylindrospermum* spp. strains CCALA 993 and 994 we detected a protonated molecule at  $m/z$  1048.6 which formed  $C_{10}H_{22}N^+$  fatty acid immonium ion at 100 eV and partial peptide sequence Gly-Ser-Asn-Ser at 40 eV (Fig. 4b). A protonated molecule  $m/z$  1084.8 detected from *Cylindrospermum* sp. CCALA

1000 formed a unsaturated fatty acid immonium ion  $C_{10}H_{20}N^+$  (Fig. 4c) and the 40 eV fragmentation provided the peptide sequence Ile/Leu,Ile/Leu,Val,Tyr,Ile/Leu, Ile/Leu (Fig. 4c). Finally in the strain *Anabaena* sp. A-Pastvisko11 we detected the putative lipopeptide with characteristic fragment  $C_{16}H_{34}N^+$  and short peptide sequence Phe-Phe-Arg (Fig. 4d).

These results demonstrates that the method developed here is capable of detecting not only already described lipopeptides but also variants of known lipopeptides and lipopeptides with novel chemical structures.

## 4. Discussion

Bacteria are a prolific source of  $\beta$ -amino fatty acid lipopeptides typically in molecular range from 700–1400 Da [5–7,16–25]. The majority of known lipopeptide chemical variants possess fully saturated non-substituted  $\beta$ -amino fatty acid chain of fatty acid length  $C_8$ – $C_{12}$  [16–25]. Some of the minor variants can be functionalized by hydroxyl-, keto-, chlorine or bromine group, but substituted variants can be also produced [21]. Complex branching and functionalization of the fatty acid chain have also been reported [19].

Characteristic fragments are routinely used in selective detection of many microbial metabolites as for example hepatotoxic microcystins [30,31]. However, no selective method for the detection of cyclic  $\beta$ -amino fatty acid lipopeptides is currently available. The peptidic nature of the compound is seen from neutral losses of amino acid residues and their immonium ions (the mass range of proteinogenic amino acid even number immonium ions is  $m/z$  30–159) in the product ion spectra produced with collision energies typical for peptides. With higher collision energies intensities of higher mass immonium ions start to dominate which are diagnostic for  $\beta$ -amino fatty acids ( $C_xH_{2x+2}N^+$  for non-functionalized and

$C_xH_{2x}N^+$  for those bearing one functionalization). Fully saturated non-derivatized  $\beta$ -amino  $C_8$ – $C_{20}$  fatty acid form even numbered immonium ions which mass range is  $m/z$  100–270. An even numbered ion mass directly shows the presence of one nitrogen in the fragment. Taking together, by coupling of the information from two separate HPLC–HRMS/MS analyzes and plotting the extracted ion chromatograms for the immonium ions of  $\beta$ -amino fatty acid chains ( $C_8$ – $C_{12}$ , both saturated and once unsaturated) we can obtain information about the presence of  $\beta$ -amino fatty acid lipopeptides in the sample analyzed (for example Fig. S3).

Different derivatives of the fatty acid chain, how it is bound to the peptide part and the structure of the peptide part change the immonium ion formation dramatically in collision induced dissociation. This was demonstrated in case of anabaenolysin A. The immonium ion formed from the fatty acid of anabaenolysin A did not dominate similarly as all the immonium ions formed from iturin and puwainaphycin variants and was obtained at lower collision energies. Identification of the peptidic nature of the compound could be problematic in peptides with 3–5 amino acid residues especially if non-proteogenic amino acids are present since there are no easy assignment tools or databases available for identification of the rare amino acid in the mass spectra analysis software. In anabaenolysin the cyclic peptide part is containing only two proteogenic amino acids and one residue specific for anabaenolysin [23,27]. In anabaenolysin A the two amino acids are both glycines so only the loss of Gly residue from the molecular ion or the presence of Gly immonium ion in the product ion spectrum would straightforward reveal the compounds peptidic structure. In practice the probability of both these events to happening is negligible low. So in screening studies without existing knowledge it could be challenging to recognize lipopeptides with more modified  $\beta$ -amino fatty acids and/or peptide part containing mostly nonproteogenic amino acids possibly together with polyketidic structures. It is also important to mention that the occurrence of multiple modifications on the fatty acid side chain including more complex branching, desaturation or multiple substitutions by –OH and –Cl groups will make the data interpretation more difficult and thus such type of compound will be hardly detected by the method presented.

The majority of characterized bacterial  $\beta$ -amino fatty acid lipopeptides contain mostly proteinogenic amino acids and have saturated fatty acid chain [5,16–25]. Therefore the method developed here has broad applications in bioprospection, pharmaceutical and environmental screening. We previously reported that *C. alatosporum* CCLA 988 is capable of synthesizing puwainaphycin variants with  $C_{10}$  and  $C_{14}$  fatty acid both chlorinated and hydroxylated [12]. However, the method described here and knowledge about the linear dependence of retention time and fatty acid length made it possible to detect puwainaphycin variants from  $C_{10}$  to  $C_{17}$  including majority of their chlorinated and hydroxylated variants.

## 5. Conclusions

Here we report a method for the analysis of crude bacterial extracts by a HPLC–HRMS/MS technique using varying collision energies as a good tool for targeted cyclic  $\beta$ -amino-lipopeptides screening. We detected novel structural variants of puwainaphycin F and G with different  $\beta$ -amino fatty acid in *C. alatosporum* CCLA 988 extract using this method. Selected high collision energy (100 eV) is suitable for both puwainaphycin and iturin A types of lipopeptides to generate diagnostic fragments. On the contrary, in anabaenolysin lower collision energy (35 eV) is needed to obtain the diagnostic  $\beta$ -amino fatty acid immonium ion. The proposed HPLC–HRMS/MS approach here provides rapid and easy method

for detecting novel variants of bioactive cyclic lipopeptides as was demonstrated in studied cyanobacterial strains.

## Acknowledgements

The work was supported by Ministry of Education of the Czech Republic—National Programme of Sustainability I, ID: LO1416 and the Czech Science Foundation (Project Nos. CSF 14-18067S and 16-09381S).

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.chroma.2016.02.013>.

## References

- [1] J. Berdy, Bioactive microbial metabolites—a personal view, *J. Antibiot.* 58 (2005) 1–26.
- [2] G.E. Chilipala, S.Y. Mo, J. Orjala, Chemodiversity in freshwater and terrestrial cyanobacteria—a source for drug discovery, *Curr. Drug Targets* 12 (2011) 1654–1673.
- [3] M. Welker, H. von Dohren, Cyanobacterial peptides—nature's own combinatorial biosynthesis, *Fems Microbiol. Rev.* 30 (2006) 530–563.
- [4] H. Wang, D.P. Fewer, L. Holm, L. Rouhiainen, K. Sivonen, Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) 9259–9264.
- [5] L. Delcambe, F. Peypoux, F. Besson, M. Guinand, G. Michel, Structure of iturin and iturin/like substances, *Biochem. Soc. Trans.* 5 (1977) 1122–1124.
- [6] N. Vanittanakom, W. Loeffler, U. Koch, G. Jung, Fengycin—a novel antifungal lipopeptide antibiotic produced by *Bacillus subtilis*, *J. Antibiot.* 39 (1986) 888–901.
- [7] J. Micklefield, Daptomycin structure and mechanism of action revealed, *Chem. Biol.* 11 (2004) 887–888.
- [8] T. Neuhof, P. Schmieder, K. Preussel, R. Dieckmann, H. Pham, F. Bartl, H. von Dohren, Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium *Hassallia* sp., *J. Nat. Prod.* 68 (2005) 695–700.
- [9] T.H. Bui, V. Wray, M. Nimtz, T. Fossen, M. Preitsch, G. Schroeder, K. Wende, S.E. Heiden, S. Mundt, Baltidins A–D antifungal hassallidin-like lipopeptides from the Baltic Sea Cyanobacterium *Anabaena cylindrica* Bio33 (vol 77, pg 1287, 2014), *J. Nat. Prod.* 78 (2015), 345–345.
- [10] E.H. Duitman, L.W. Hamoen, M. Rembold, G. Venema, H. Seitz, W. Saenger, F. Bernhard, R. Reinhardt, M. Schmidt, C. Ullrich, T. Stein, F. Leenders, J. Vater, The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase an amino transferase, and a fatty acid synthase, *Proc. Natl. Acad. Sci. U. S. A.* 96 (1999) 13294–13299.
- [11] S.Y. Yao, X.W. Gao, N. Fuchsbaue, W. Hillen, J. Vater, J.S. Wang, Cloning sequencing, and characterization of the genetic region relevant to biosynthesis of the lipopeptides iturin A and surfactin in *Bacillus subtilis*, *Curr. Microbiol.* 47 (2003) 272–277.
- [12] J. Mares, J. Hajek, P. Urajova, J. Kopecky, P. Hrouzek, A hybrid non-ribosomal peptide/polyketide synthetase containing Fatty-Acyl Ligase (FAAL) synthesizes the  $\beta$ -Amino fatty acid lipopeptides Puwainaphycins in the cyanobacterium *Cylindrospermum alatosporum*, *PLoS One* 9 (2014) E111904, 10.1371.
- [13] A. Kakinuma, M. Hori, H. Sugino, I. Yoshida, M. Isono, G. Tamura, K. Arima, Determination of location of the lactone ring in surfactin, *Agric. Biol. Chem.* 33 (1969) 1523–1524.
- [14] T.K. Shishido, A. Humisto, J. Jokela, L.W. Liu, M. Wahlsten, A. Tamrakar, D.P. Fewer, P. Permi, A.P.D. Andreote, M.F. Fiore, K. Sivonen, Antifungal compounds from cyanobacteria, *Mar. Drugs* 13 (2015) 2124–2140.
- [15] J. Vestola, T.K. Shishido, J. Jokela, D.P. Fewer, O. Aitio, P. Permi, M. Wahlsten, H. Wang, L. Rouhiainen, K. Sivonen, Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway, *Proc. Natl. Acad. Sci. U. S. A.* 111 (2014) E1909, 10.1073.
- [16] W.P. Frankmölle, G. Knubel, R.E. Moore, G.M.L. Patterson, Antifungal cyclic peptides from terrestrial blue green alga *Anabaena laxa*. 2. Structures of laxaphycin A, B, D and E, *J. Antibiot.* 45 (1992) 1458–1466.
- [17] W.H. Gerwick, Z.D. Jiang, S.K. Agarwal, B.T. Farmer, Total structure of hormothamnin A, a toxic cyclic undecapeptide from the tropical marine cyanobacterium *Hormothamnion enteromorphoides*, *Tetrahedron* 48 (1992) 2313–2324.
- [18] H.S. Kang, M. Sturdy, A. Krunic, H. Kim, Q. Shen, S.M. Swanson, J. Orjala, Minutissamides E–L, antiproliferative cyclic lipodecapeptides from the cultured freshwater cyanobacterium cf. *Anabaena* sp., *Bioorgan. Med. Chem.* 20 (2012) 6134–6143.

- [19] T.Y. An, T.K.S. Kumar, M.L. Wang, L. Liu, J.O. Lay, R. Liyanage, J. Berry, M. Gantar, V. Marks, R.E. Gawley, K.S. Rein, Structures of pahayokolides A and B, cyclic peptides from a *Lyngbya* sp, *J. Nat. Prod.* 70 (2007) 730–735.
- [20] P. Hrouzek, M. Kuzma, J. Cerny, P. Novak, R. Fiser, P. Simek, A. Lukesova, J. Kopecky, The cyanobacterial cyclic lipopeptides puwainaphycins F/G are inducing necrosis via cell membrane permeabilization and subsequent unusual actin relocalization, *Chem. Res. Toxicol.* 25 (2012) 1203–1211.
- [21] P. Tomek, P. Hrouzek, M. Kuzma, J. Sykora, R. Fiser, J. Cerny, P. Novak, S. Bartova, P. Simek, M. Hof, D. Kavan, J. Kopecky, Cytotoxic lipopeptide muscotoxin A, isolated from soil cyanobacterium *desmonostoc muscorum* permeabilizes phospholipid membranes by reducing their fluidity, *Chem. Res. Toxicol.* 28 (2015) 216–224.
- [22] N. Maru, O. Ohno, D. Uemura, Lyngbyacyclamides A and B, novel cytotoxic peptides from marine cyanobacteria *Lyngbya* sp, *Tetrahedron Lett.* 51 (2010) 6384–6387.
- [23] J. Jokela, L. Oftedal, L. Herfindal, P. Permi, M. Wahlsten, S.O. Doskeland, K. Sivonen, Anabaenolysins, novel cytolytic lipopeptides from benthic *Anabaena* cyanobacteria, *PLoS One* 7 (2012) E41222, 10.1371.
- [24] J.B. MacMillan, M.A. Ernst-Russell, J.S. de Ropp, T.F. Molinski, Lobocyclamides A–C, lipopeptides from a cryptic cyanobacterial mat containing *Lyngbya confervoides*, *J. Org. Chem.* 67 (2002) 8210–8215.
- [25] A. Plaza, C.A. Bewley, Largamides A–H, unusual cyclic peptides from the marine cyanobacterium *oscillatoria* sp. (vol 71, pg 6898, 2006), *J. Org. Chem.* 74 (2009), 486–486.
- [26] L. Oftedal, L. Myhren, J. Jokela, G. Gausdal, K. Sivonen, S.O. Doskeland, L. Herfindal, The lipopeptide toxins anabaenolysin A and B target biological membranes in a cholesterol-dependent manner, *BBA-Biomembranes* 1818 (2012) 3000–3009.
- [27] T.K. Shishido, J. Jokela, C.T. Kolehmainen, D.P. Fewer, M. Wahlsten, H. Wang, L. Rouhiainen, E. Rizzi, G. De Bellis, P. Permi, K. Sivonen, Antifungal activity improved by coproduction of cyclodextrins and anabaenolysins in cyanobacteria, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015), 10.1073.
- [28] M.B. Allen, D.I. Arnon, Studies on nitrogen-fixing blue-green algae. 1. Growth and nitrogen fixation by *anabaena cylindrica* LEMM, *Plant Physiol.* 30 (1955) 366–372.
- [29] J. Kótai, Instructions for preparation of modified nutrient solution Z8 for algae, *Blindern B–11/69* (Norwegian Institute for Water Research, Oslo).
- [30] L. Spoof, P. Vesterkvist, T. Lindholm, J. Meriluoto, Screening for cyanobacterial hepatotoxins, microcystins and nodularin in environmental water samples by reversed-phase liquid chromatography–electrospray ionisation mass spectrometry, *J. Chromatogr. A* 1020 (2003) 105–119.
- [31] J. Kohoutek, O. Adamovský, M. Oravec, Z. Simek, M. Palíková, R. Kopp, L. Bláha, LC–MS analyses of microcystins in fish tissues overestimate toxin levels—critical comparison with LC–MS/MS, *Anal. Bioanal. Chem.* 398 (2010) 1231–1237.

## 4.2. Paper II.

## GENOME MINING REVEALS HIGH INCIDENCE OF PUTATIVE LIPOPEPTIDE BIOSYNTHESIS NRPS/PKS CLUSTERS CONTAINING FATTY ACYL-AMP LIGASE GENES IN BIOFILM-FORMING CYANOBACTERIA<sup>1</sup>

*Tomáš Galica*

Centre Algatech, Institute of Microbiology, The Czech Academy of Sciences (CAS), Novohradská 237, 37981 Třeboň, Czech Republic

Biology Centre of the CAS, Institute of Hydrobiology, Na Sádkach 7, 37005 České Budějovice, Czech Republic  
Faculty of Science, University of South Bohemia, Branišovská 1750, 37005 České Budějovice, Czech Republic

*Pavel Hrouzek*

Centre Algatech, Institute of Microbiology, The Czech Academy of Sciences (CAS), Novohradská 237, 37981 Třeboň, Czech Republic

and *Jan Mareš*<sup>2</sup> 

Centre Algatech, Institute of Microbiology, The Czech Academy of Sciences (CAS), Novohradská 237, 37981 Třeboň, Czech Republic

Biology Centre of the CAS, Institute of Hydrobiology, Na Sádkach 7, 37005 České Budějovice, Czech Republic  
Faculty of Science, University of South Bohemia, Branišovská 1750, 37005 České Budějovice, Czech Republic

**Cyanobacterial lipopeptides have antimicrobial and antifungal bioactivities with potential for use in pharmaceutical research. However, due to their hemolytic activity and cytotoxic effects on human cells, they may pose a health issue if produced in substantial amounts in the environment. In bacteria, lipopeptides can be synthesized via several well-evidenced mechanisms. In one of them, fatty acyl-AMP ligase (FAAL) initiates biosynthesis by activation of a fatty acyl residue. We have performed a bioinformatic survey of the cyanobacterial genomic information available in the public databases for the presence of FAAL-containing non-ribosomal peptide synthetase/polyketide synthetase (NRPS/PKS) biosynthetic clusters, as a genetic basis for lipopeptide biosynthesis. We have identified 79 FAAL genes associated with various NRPS/PKS clusters in 16% of 376 cyanobacterial genomic assemblies available, suggesting that FAAL is frequently incorporated in NRPS/PKS biosynthetases. FAAL was present either as a stand-alone protein or fused either to NRPS or PKS. Such clusters were more frequent in derived phylogenetic lineages with larger genome sizes, which is consistent with the general pattern of NRPS/PKS pathways distribution. The putative lipopeptide clusters were more frequently found in genomes of cyanobacteria that live attached to surfaces and are capable of forming microbial**

**biofilms. While lipopeptides are known in other bacterial groups to play a role in biofilm formation, motility, and colony expansion, their functions in cyanobacterial biofilms need to be tested experimentally. According to our data, benthic and terrestrial cyanobacteria should be the focus of a search for novel candidates for lipopeptide drug synthesis and the monitoring of toxic lipopeptide production.**

*Key index words:* cyanobacteria; fatty-acyl AMP ligase; genome mining; lipopeptides; microbial biofilm; non-ribosomal peptide synthesis

*Abbreviations:* FAAL, fatty acyl-AMP ligase; NRPS, non-ribosomal peptide synthase; PKS, polyketide synthase; PLC, putative lipopeptide biosynthesis clusters; PSSM, position-specific scoring matrix; WGS, whole-genome sequence

---

It is well known that cyanobacteria produce a wide array of secondary metabolites including alkaloids, polyketones, heterocyclic compounds, and peptides (Welker and von Döhren 2006, Van Wagoner et al. 2007). Cyanobacterial secondary metabolites frequently contain non-proteinogenic amino acids and other chemical modifications that stabilize and enable them to bind to various molecular targets (Welker and von Döhren 2006). Bioactivities of these compounds are as variable as their structures. Some of them are toxic (e.g., microcystin and anatoxin; Dittmann et al. 2013), while others exhibit therapeutic potential (e.g., dolastatin; Gerwick and Moore 2012).

<sup>1</sup>Received 19 December 2016. Accepted 22 May 2017. First Published Online 20 June 2017. Published Online 18 July 2017, Wiley Online Library (wileyonlinelibrary.com).

<sup>2</sup>Author for correspondence: e-mail jan.mares@centrum.cz  
Editorial Responsibility: J. Collier (Associate Editor)



One of the most prominent groups of cyanobacterial secondary metabolites is the lipopeptides, defined by the presence of a fatty acid residue attached to a peptide core (Fig. 1). Similar to lipopeptides produced by other prokaryotes, cyanobacterial lipopeptides exhibit antifungal, antimicrobial, and cytotoxic activities (Pergament and Carmeli 1994, Neuhof et al. 2005, Hrouzek et al. 2012). These activities are linked to their ability to disrupt biological membranes via their permeabilization (puwainaphycin, Hrouzek et al. 2012, anabaenolysin, Oftedal et al. 2012, muscotoxin, Tomek et al. 2015). Some lipopeptides produced by heterotrophic bacteria such as *Bacillus subtilis* (e.g., daptomycin) are, due to their specific effect on prokaryotic membranes, used as antibiotics (Fowler et al. 2006).

Understanding the chemical structure of secondary metabolites, their biosynthesis and genetic basis makes it possible to apply a genome-mining strategy in search for potential producers of novel compounds among strains with publicly available genomes (Liu et al. 2014, Micallef et al. 2015a,b). In lipopeptides, the peptide core is synthesized by a large protein complex consisting mostly of non-ribosomal peptide synthetases (NRPS; Duitman et al. 1999). For the function of NRPS, three protein domains are essential: the adenylation domain (A-domain), which recognizes and activates a specific amino acid prior to its incorporation into the nascent peptide, the thiolation domain (T-domain), which via the phosphopantetheine arm docks the growing peptide, and the condensation domain (C-domain), which catalyzes the peptide bond formation (Welker and von Döhren 2006). Bioinformatic tools allow prediction of A-domain specificity based on its amino acid sequence (Stachelhaus

et al. 1999). Aside from NRPS, the biosynthetic assembly line may involve polyketide synthases (PKS) as well as various tailoring enzymes (methyltransferases, aminotransferases, oxygenases) that modify the incorporated peptide/acyl chain, thus amplifying the structural diversity (Duitman et al. 1999, Edwards et al. 2004, Ramaswamy et al. 2007). Proteins of a particular NRPS/PKS biosynthetic pathway are coded by genes organized in large clusters. The order in which the proteins participate in the biosynthesis of a lipopeptide usually follows the arrangement of enzymatic domains within the NRPS/PKS modules and also roughly follows the order of the corresponding genes in the operon (the colinearity rule; Marahiel et al. 1997, Guenzi et al. 1998). Therefore, the number and identity of incorporated amino acids, as well as (less accurately) their sequence in the expected product can be inferred based solely on genetic information.

According to current knowledge, biosynthetic ligation of the fatty acyl residue to a nascent lipopeptide can be performed employing at least two different mechanisms. Typically, the binding of the fatty acid to the NRPS/PKS complex is the first step in the biosynthesis pathway (e.g., Duitman et al. 1999, Miao et al. 2005, Mareš et al. 2014). In one of the possible scenarios, a previously activated fatty acid in the form of an acyl-CoA is recognized by a substrate-specific starter C-domain and subsequently fused to the first amino acid of the emerging peptide (fengycin, Koumoutsis et al. 2004, surfactin, Kraas et al. 2010). Nevertheless, the acyl-CoA is also a common intermediate of basal fatty acid metabolism. Hence, genes encoding fatty acyl-CoA ligases, which activate the fatty acids, may either be associated with the biosynthetic gene cluster (fengycin; Koumoutsis et al. 2004) or encoded elsewhere in the

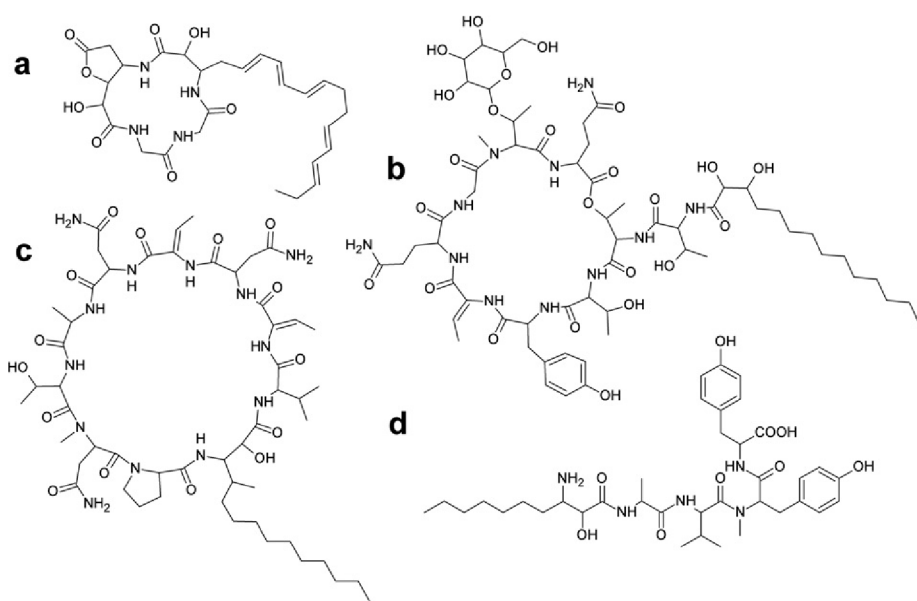


FIG. 1. Examples of cyanobacterial lipopeptides. Structures of common cyanobacterial lipopeptides: (a) anabaenolysin A, (b) hassallidin A, (c) puwainaphycin F, (d) microginin.

genome rather as part of the fatty acid metabolism (surfactin; Kraas et al. 2010).

In cyanobacteria, this type of lipopeptide biosynthesis has not yet been demonstrated, although a similar mechanism may be involved in the assembly of the lipidic part of glycosylated antifungal lipopeptides hassallidins (Vestola et al. 2014). Alternatively, in the second described mechanism the fatty acid is recognized, adenylated, and fused directly to an acyl carrier domain, thus bypassing the acyl-CoA intermediate. Domains catalyzing this reaction in the biosynthesis of jamaicamide and hectochlorin were named acyl-AMP synthases in the respective studies (Edwards et al. 2004, Ramaswamy et al. 2007). For a homologous domain performing an identical reaction in the biosynthesis of puwainaphycins, the designation fatty acyl-AMP ligase (FAAL) was proposed (Mareš et al. 2014). Finally, the same mechanism was reported in the biosynthesis of bacterial lipopeptides of the iturin family. Therein, the reaction is catalyzed by a domain originally designated as acyl synthase, which differs slightly in structure from the domains activating fatty acids in cyanobacteria (Duitman et al. 1999, Hansen et al. 2007). As the reaction catalyzed by all these domains (acyl-AMP synthase, FAAL, acyl synthase) is the same, we henceforth refer to them as FAAL. The FAAL enzymes, in association with PKS, participate also in hydrocarbon and lipid metabolism, e.g., in synthesis of olefins (Coates et al. 2014) or unusual lipids (Mohanty et al. 2011). In natural product gene clusters, the FAAL usually represents a starter unit followed by PKS/NRPS and tailoring genes (Duitman et al. 1999, Mareš et al. 2014).

The involvement of the FAAL domain in the production of lipopeptides was first observed in biosynthesis of mycosubtilin as a part of the MycA protein, along with PKS, aminotransferase and NRPS (Duitman et al. 1999). The biosynthesis starts when the fatty acid is recognized by FAAL, then fused to AMP and bound to the first T-domain. Subsequently, the carbon chain is elongated by PKS, which results in the addition of 2 carbons and a keto group. The original carbonyl oxygen of the fatty acid is replaced with an amino group by an aminotransferase. Seven amino acids are added by NRPS modules and the nascent peptide is released and cyclized by a thioesterase domain (Te-domain). Iturin, bacillomycin, paenilarvin and puwainaphycin are synthesized analogously (Tsuge et al. 2001, Koumoutsi et al. 2004, Mareš et al. 2014, Sood et al. 2014), although the architecture of the gene clusters shows some variability (e.g., a stand-alone FAAL in the puwainaphycin gene cluster). Aside from cyclic lipopeptides mentioned thus far, FAAL is involved in the synthesis of the linear lipopeptide microginin (designated as an adenylation domain for octanoic acid; Kramer 2010) and other cyanobacterial metabolites that contain a fatty acid moiety, namely jamaicamide, hectochlorin and hapalosin (designated as acyl-ACP

synthases in the respective publications; Edwards et al. 2004, Ramaswamy et al. 2007, Micallef et al. 2015b).

Considering the expected biosynthesis scenarios, genome mining for putative lipopeptide producers in cyanobacteria may focus on the specific components of the gene clusters such as either the starter C-domains or the FAAL domains. So far, four cyanobacterial secondary metabolites (puwainaphycins, jamaicamides, hectochlorins, hapalosins) have been shown to acquire their fatty acid moiety via FAAL while the mechanism of its incorporation remains unclear in others (e.g., hassallidins, Vestola et al. 2014, anabaenolysins, Shishido et al. 2015). On the basis of this knowledge, we have searched publicly available genomes for biosynthetic gene clusters that encode FAAL along with NRPS genes. We propose that such clusters will likely produce lipopeptides. Furthermore, we have investigated the distribution of such putative lipopeptide biosynthesis clusters (PLC) among cyanobacteria, considering both their phylogeny as well as ecology/habitat.

#### METHODS

FAAL and acyl synthases (PSSM-Id: 213276, Conserved Domain Database) were reported to activate and incorporate fatty acid into lipopeptides via the mechanism of acyl-adenylate formation and its direct transfer to phosphopantetheine arm of T-domain (Duitman et al. 1999, Hansen et al. 2007). Therefore, both types were considered for construction of position-specific scoring matrix (PSSM) and subsequent BLAST search. Amino-acid sequences of FAAL and acyl-synthase domains from 12 known biosynthetic clusters, namely from PuwC (AIW82280, puwainaphycins), DptE (AAX31555, daptomycin), microginin acyl synthase (WP\_026798330), HctA (AAY42393, hectochlorin), ColA (AKQ09578, columbamides), JamA (AAS98774, jamaicamides), Tar4 (AHH53502, taromycin), MycA (AAF08795, mycosubtilin), ItuA (BAB69698, iturin A), BmyA (ABS74181, bacillomycin), and PalA (AHD05679, paenilarvins) were extracted and aligned using MAFFT v. 7 (Katoh and Standley 2013) alignment tool in Geneious software (version 8.1.7, default settings; Biomatters Ltd., Auckland, New Zealand.). The resulting 586 amino acids long multiple sequence alignment was checked for presence of conserved residues required for fatty acid recognition (Goyal et al. 2012) and then used to generate PSSM using locally installed BLAST+ package. PSSM was blasted against the NCBI database of non-redundant protein sequences via PSI-BLAST web interface. Search space was restricted by taxonomy to sequences of cyanobacterial origin and the number of maximum target sequences was set to 5,000. Other search parameters were set to default. Obtained results were downloaded in .xml format and reordered according to e-values of particular

domains instead of e-value of the whole proteins. FAAL domains are homologous to A-domains of NRPS therefore those also give a weak signal that is multiplied if there are more of them in the same protein, resulting in a higher e-value of the whole protein. The cut-off value was set to  $1e^{-120}$ . At this point, the e-value dropped suddenly and domains that were not identified as FAAL started to occur (e.g., proteins HapC or BarE; Fig. S1 in the Supporting Information). To reliably evaluate PLC distribution across cyanobacteria, the results were inspected manually to exclude duplicates from the same strain deposited under different IDs, but at the same time to include identical proteins from different strains (tagged [multispecies] in the NCBI protein database).

The resulting list of proteins was processed by a custom script written in Python/BioPython (Cock et al. 2009). For every hit, the script obtained the coding sequence and its genomic surroundings (set to 60,000 nucleotides in both directions, if possible) and all the proteins encoded there. Downloaded protein sequences were examined for conserved domains known to be involved in NRPS/PKS pathways and if evaluated as positives their corresponding genes were highlighted in the nucleotide sequence. In the case of script failure for a particular hit, the process was finished manually. The pre-processed nucleotide sequences were inspected individually.

Candidate genomic regions were evaluated as negative if they did not contain genes coding for proteins with altogether at least three A-domains. The threshold of at least three A-domains was applied as the intention of this study was to screen for lipopeptides based on a true peptide core composed of several amino acids. This precaution was applied to eliminate the PLCs with hybrid products containing one or two amino acids lacking standard peptide bonds (e.g., hectochlorin; Ramaswamy et al. 2007), despite the presence of one or two A-domains encoded in the gene cluster. Putative gene clusters containing FAAL and PKS genes but lacking NRPS genes were omitted from further analysis. If in the vicinity of the FAAL-coding sequence, there were genes that coded for at least three A-domains the sequence was considered to contain a PLC. Evaluation was not possible if the contig was too short or when the FAAL-coding gene was located near an assembly gap; such cases were left unresolved.

*Classification and domain count in PLCs.* Some of the PLCs contained genes for a sufficient number of A-domains but either were found at the end of a contig, lacked a thioesterase enzyme, or contained minor sequence errors and therefore it was reasonable to assume that they are not complete. Such PLCs were omitted from further analysis. Complete PLCs were classified according to relative position of the FAAL coding region in the cluster and the presence of PKS. Proteins encoded by genes in the

PLCs were inspected for their domain content as provided in their annotation in the NCBI database. In the case of a dubious or outdated annotation, novel CD-search was performed in the NCBI conserved domains database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>; Marchler-Bauer and Bryant 2004, Marchler-Bauer et al. 2015).

*Ecology of cyanobacteria with available genome information.* The list of available assemblies was obtained from GenBank via Entrez Utilities. Search query was organism=cyanobacteria, date: from January 01, 1970 to March 23, 2016. Strains with multiple assemblies, or derived mutants were counted once. Literature was reviewed for the information regarding the original habitat of the organisms (Appendix S1 in the Supporting Information). This information was further used to categorize the strains into three categories: substrate-associated, planktic, and symbiotic. Strains with unclear origin (22) were not classified. If possible, substrate-associated strains were further divided in the following subcategories: marine, saline, freshwater, terrestrial, and thermophilic. Similarly, planktic strains were further classified as marine, saline, or freshwater.

*Phylogenetic analysis.* The phylogenetic tree was generated using a subset of 23 conserved single-copy orthologous protein-coding genes selected from loci previously tested for congruent cyanobacterial phylogenies (Shih et al. 2013, Komárek et al. 2014). The selection was made based on the presence of a full unambiguous sequence of the gene in all analyzed genomes (including drafts at various stages of finishing) and comprised these loci: *dnaG*, *frz*, *infC*, *nusA*, *pgk*, *pyrG*, *rpIB*, *rpIC*, *rpID*, *rpIE*, *rpIF*, *rpIM*, *rpIP*, *rpIT*, *rpOB*, *rpsC*, *rpsE*, *rpsI*, *rpsK*, *rpsM*, *rpsS*, and *tsf*. First, BLAST queries were derived from sequences of the respective proteins mined from the complete genome of *Synechocystis* sp. PCC 6803 (BA000022). Each of the 23 queries was then used in *tblastn* (cut-off value  $1e^{-15}$ ) algorithm searches against a custom database compiled from all the complete cyanobacterial genomes available and whole-genome sequence (WGS) contigs downloadable from NCBI (April 2016). Hits for each protein were aligned using MAFFT v. 7 FFT-NS-i algorithm and the alignments were manually reviewed to remove ambiguous sites and gap regions. The best substitution model was selected for each of the alignments using ProtTest 3.4 (Guindon and Gascuel 2003, Darriba et al. 2011), applying the Bayesian Information Criterion. All but one of the loci were assigned LG+I+G as the best model (and the single protein Pkg was assigned LG+I+G+F). A Maximum Likelihood tree was calculated using each of the 23 alignments with 1,000 bootstrap pseudoreplications in RaxML v. 8 (Stamatakis 2014). The resulting phylogenies were inspected manually for topological incongruence. As individual protein trees were generally consistent, all alignments were concatenated into a 5,679 amino-acid long matrix

containing 194 sequence lines. A Maximum Likelihood tree was produced using RaxML v. 8, employing a separate substitution model for each of the 23 partitions as selected using ProtTest. One thousand bootstrap pseudoreplications were calculated to evaluate the relative support of the branches. A Bayesian Inference tree was obtained using MrBayes v.3.2.6 (Ronquist et al. 2012). Two independent runs of eight Markov chains were performed for ~1.5 million generations, sampling every 100th tree until the likelihood values were stable and the divergence criterion was lower than 0.01. The Bayesian Inference calculation employed a common LG+I+G evolutionary model due to excessive computational demands of a calculation using partitioned models. Posterior probabilities were estimated from branch frequencies in the sampled trees, discarding the first 25% of the harvested data as burn-in. A neighbor-joining phylogenetic analysis was performed in SeaView v.4 (Gouy et al. 2010) using the BioNJ algorithm (Gascuel 1997) with Poisson distance model and 1,000 bootstrap pseudoreplications. The Maximum Likelihood tree was run via the CIPRES supercomputing facility (Miller et al. 2015) and Bayesian Inference calculations were run on MetaCentrum computer grid (www.metacentrum.cz).

Genome sizes of all 376 cyanobacterial isolates with whole-genome data available included in the PSI-BLAST search for PLCs were extracted from the NCBI databases. Comparison of average genome size between the PLC-containing and non-containing isolates was performed using a standard *t*-test in Statistica v. 13 software (Fig. S2 in the Supporting Information).

## RESULTS

*Bioinformatic search for putative lipopeptides.* The bioinformatic search yielded 400 proteins with FAAL domain from 150 isolates. Of these, 255 were evaluated as irrelevant to lipopeptide production. Further 60 candidate proteins were encoded on contigs too short or of bad quality and left unresolved (Appendix S2 in the Supporting Information). Three FAAL domain-containing proteins were part of known secondary metabolite producing pathways, namely jamaicamide, hectochlorin, and hapalosin. Although in literature these compounds were described as lipopeptides, their gene clusters did not meet our search criteria to contain at least three A-domains. Finally, each of the remaining 82 proteins was recovered as a part of FAAL-NRPS gene clusters and was considered further.

Three of the 82 clusters have already been described previously and submitted to NCBI as separate nucleotide entries representing biosynthetic gene clusters. One supposedly produces puwainaphycins and the other two are orphan gene clusters (KM078884, KC407996 and AB279593; Nishizawa et al. 2007, Kampa et al. 2013, Mareš et al. 2014).

The remaining 79 clusters were recovered from genome assemblies. Among these, there were three clusters (Fig. 2, M-1, M-3, M-4) previously suggested to code for microginin biosynthesis (Rouge et al. 2009, Humbert et al. 2013). In addition, our search yielded two previously unreported microginin biosynthetic gene clusters from *Planktothrix* sp. NIVA-CYA 406 (Fig. 2, M-2) and *Microcystis aeruginosa* TAIHU-98 (Fig. 2, M-5). Cluster M-5 was longer than the rest due to two additional NRPS modules. A stand-alone FAAL-encoding gene was found in all microginin clusters several kbp upstream from the first NRPS gene (*micA*).

Aside from microginin gene clusters, the PLCs were divided into three groups according to the organization of genes for different biosynthetic steps. Group 1 (Fig. 3 and Fig. S3 in the Supporting Information) contained FAAL accompanied solely by NRPS modules and tailoring enzymes (no PKS is present), in group 2 (Fig. 4 and Fig. S4 in the Supporting Information) FAAL was fused to an NRPS module, but the cluster contained at least one PKS, and in group 3 (Fig. 5 and Fig. S5 in the Supporting Information) FAAL was either fused to a PKS module or was immediately followed by a PKS. Another 18 clusters contained FAAL and a sufficient number of A-domains to be considered PLCs; however, they were not fully sequenced or contained minor gaps that prevented further classification.

*Group 1.* With a single exception, genes encoding biosynthetic steps in group 1 PLCs were coded in the same orientation (Fig. 3). In most of the clusters (8/10), the FAAL domain was encoded as a part of a longer protein along with an NRPS module. The clusters in this group ranged from 27 to 40 kbp in length and contained from 3 to 14 genes. The putative biosynthetic assembly lines consisted of 5–8 NRPS modules and a varying number of tailoring enzymes of which the dioxygenases were the most frequent (in seven of 10 clusters) followed by methyltransferases (4/10) and aminotransferases (3/10). In six pathways, there were two dioxygenases arranged in a particular pattern, so that there was one NRPS module surrounded by NRPS modules with associated dioxygenases.

*Group 2.* In 31 cases, FAAL was followed by NRPS and then by PKS (Fig. 4). None of the clusters in this group encoded a stand-alone FAAL. Putative clusters in this group ranged from 25 to 75 kbp in length and contained from 3 to 33 genes. Genes encoding FAAL, all NRPS and PKS modules as well as tailoring enzymes shared the same orientation except for a single gene in the clusters G2-3 and G2-9. In 26/31 clusters, there was a common arrangement where FAAL was followed by one NRPS module and by a PKS module (not taking into account the tailoring enzymes). The most abundant of tailoring enzymes were dioxygenases (26/31), aminotransferases (26/31), and methyltransferases (23/31). Furthermore, if also

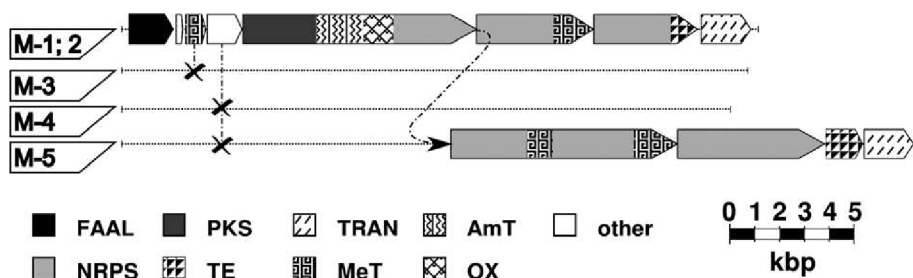


FIG. 2. Schematic representation of microginin gene clusters. The top sketch represents the reference microginin gene cluster as it is found in *Planktothrix prolifica* NIVA-CYA 98. The rest of the microginin gene clusters are depicted as dotted lines parallel to the reference with only the differences outlined. Missing genes are indicated by an "X" sign. Genes are shaded according to the protein domain they encode. FAAL, fatty acyl-AMP ligase; NRPS, non-ribosomal peptide synthetase; PKS, polyketide synthase; TE, thioesterase; TRAN, transporter; MeT, methyltransferase; AmT, aminotransferase; OX, oxygenase. For a full list of microginin gene clusters, see Figure S6 in the Supporting Information.

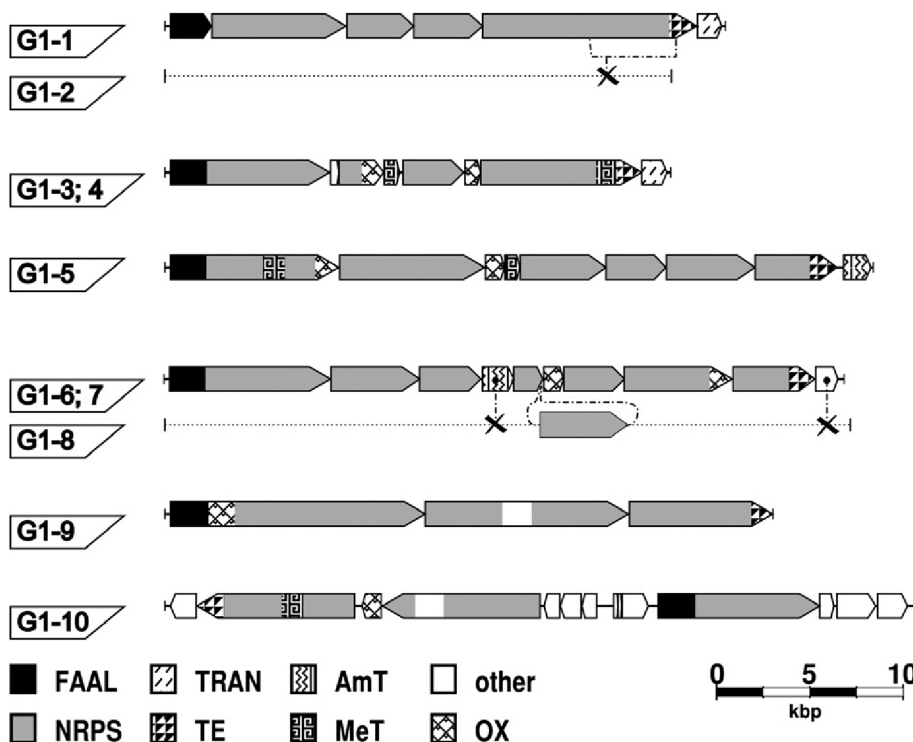


FIG. 3. Schematic representation of group 1 PLCs. Sketches represent PLCs that do not include gene(s) for PKS (group 1). Identical PLCs found in different strains are represented by a single sketch. Gene clusters with minor differences are grouped together, where the uppermost is the reference and the rest is depicted as dotted lines parallel to the reference with the differences outlined. Missing genes are indicated by an "X" sign. Genes depicted on the dotted lines represent insertions in comparison to the reference with the point of insertion indicated by a dashed line. Genes are shaded according to the protein domain they encode. FAAL, fatty acyl-AMP ligase; NRPS, non-ribosomal peptide synthetase; TE, thioesterase; TRAN, transporter; MeT, methyltransferase; AmT, aminotransferase; OX, oxygenase. For a full list of group 1 PLCs, see Figure S3.

monooxygenases were considered, there were only two clusters without any oxygenase. In five clusters, a gene encoding a glycosyltransferase was detected, suggesting glycosylation of the produced lipopeptide.

**Group 3.** PLCs of group 3 were the most variable (Fig. 5). In total, 18 clusters fell into this group. One of them was previously suggested to produce cyclic lipopeptides puwainaphycins. Another had

already been published as an orphan cluster, the *psm3* gene cluster, and G3-5 was very similar to it. In 10 of the 18 PLCs, the putative biosynthetic genes were bidirectional. In such cases, it was difficult to determine cluster boundaries and a rather wider range was taken. FAAL was encoded as a stand-alone protein in 3/18 cases, in one case, it was fused to a T-domain and in 4/18 it was fused to an acyl-CoA dehydrogenase-like domain and a T-domain. The

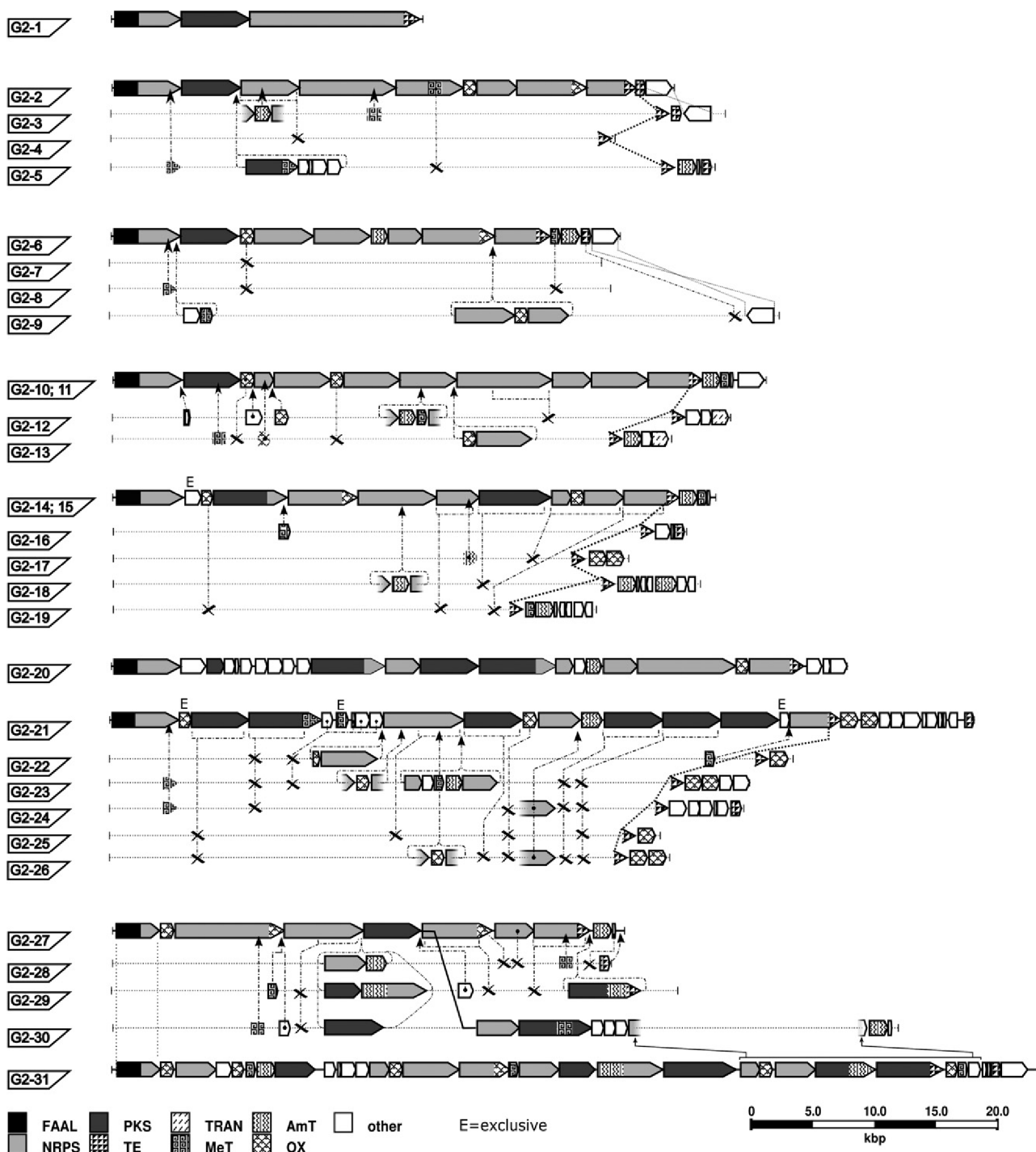


FIG. 4. Schematic representation of group 2 PLCs. Sketches represent PLCs that encode FAAL fused to NRPS, and a PKS (group 2). Topologically identical PLCs found in different strains are represented by a single sketch. Gene clusters with minor differences are grouped together where the uppermost is the reference and the rest is depicted as dotted lines parallel to the reference with the differences outlined. Missing genes, or their parts, are indicated by an "X" sign. Genes (with black border line) or domain coding regions (without border line) depicted on the dotted lines represent insertions in comparison to the reference with the point of insertion indicated by a dashed line. "E" symbol over a gene indicates a gene that is exclusively present in the reference gene cluster. A set of various tailoring enzymes is often observed downstream from the last NRPS module fused to the thioesterase (indicated by dotted vertical line connecting the thioesterases). Beginning of G2-30 shares the same topology with G2-27, yet its latter part (between the two small arrows) is topologically identical to G2-31. In G2-6 and G2-9 at the end of the cluster there is a single gene with opposite orientation (marked with gray dotted line). Genes are shaded according to protein domain, they encode: FAAL, fatty acyl-AMP ligase; NRPS, non-ribosomal peptide synthase; PKS, polyketide synthase; TE, thioesterase; TRAN, transporter; MeT, methyltransferase; AmT, aminotransferase; OX, oxygenase. For a full list of group 2 PLCs see Figure S4.

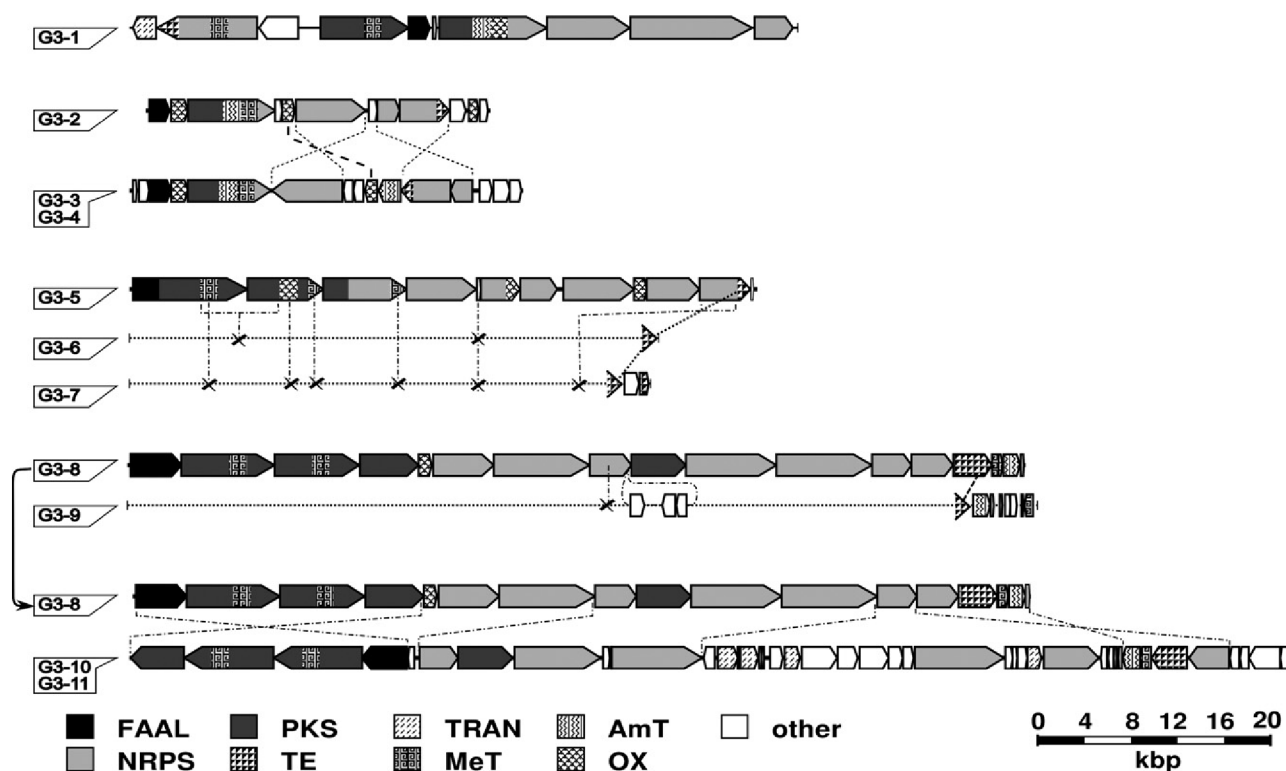


Fig. 5. Representatives of group 3 PLCs. Schematic representation of PLCs where FAAL is encoded either fused to PKS or separately but immediately followed by a gene encoding PKS (group 3). G3-1 is provided as an example of PLCs with proposed product (puwainaphycins). Other group 3 PLCs depicted are those that were found in multiple species or had topologically similar counterparts in other strains. For a full list of group 3 PLCs see Figure S5. Topologically identical PLCs found in different strains are represented by a single sketch. Among similar gene clusters dotted lines indicate related genes that however differ in orientation. Where possible gene clusters with minor differences are grouped together where the uppermost is the reference and the rest is depicted as dotted lines parallel to the reference with the differences outlined. Missing genes, or their parts, are indicated by an “X” sign. Genes depicted on the dotted lines parallel to the reference represent insertions with the point of insertion indicated by a dashed line. Dotted vertical line indicates the position of the thioesterase, downstream of which a region encoding variable tailoring enzymes is usually present. G3-8 is depicted twice. Once in relation to G3-9, with which it shares the same topology. It is depicted once more in relation to G3-10 and G3-11, where the respective genes are arranged differently although they encode the same biosynthetic steps. Genes are shaded according to the protein domain they encode. FAAL, fatty acyl-AMP ligase; NRPS, non-ribosomal peptide synthase; PKS, polyketide synthase; TE, thioesterase; TRAN, transporter; MeT, methyltransferase; AmT, aminotransferase; OX, oxygenase.

most frequent tailoring enzymes were again dioxygenases (11/18), aminotransferases (13/18), and methyltransferases (12/18).

*Distribution of putative lipopeptide-producing NRPS pathways. Phylogenetic and taxonomic distribution of PLCs:* A total of 79 PLCs were found in 59 isolates representing 31 cyanobacterial genera as designated in the NCBI database, with the highest abundances in strains of *Fischerella* (eight strains), *Cyanotheca* (five strains), *Microcystis* (five strains), *Nostoc* (five strains) and *Leptolyngbya* (four strains). To evaluate the distribution of PLCs in different cyanobacterial phylogenetic lineages, the PLCs were mapped on the multilocus phylogenetic tree (Fig. 6). The lineages resolved in the tree were taxonomically designated to correspond to the current system of cyanobacteria (Komárek et al. 2014).

In general, the PLCs were recovered in numerous independent lineages throughout the phylogenetic tree, with no evidence for their common

monophyletic origin. Nevertheless, several pairs and groups of PLCs with a highly similar organization were found in closely related strains, such as *Cyanotheca* (G1-1, G1-2), *Chlorogloeopsis* (G2-14, G2-15), or *Xenococcus/Stanieria* (Fig. 4, G2-6, G2-7, G2-8). Interestingly, clusters G1-4 and G1-5 from *Tolypothrix* sp. PCC 7601 and *Anabaena variabilis* ATCC 29413 share the same cluster topology, and even exhibit high sequence identity on DNA level (pairwise blast, 98% query coverage, 96% identity), yet the strains are not closely related. Analogously, clusters G2-2, G2-3, G2-4, and G2-5 from strains *Leptolyngbya boryana* IAM M-101, *Microchaete* sp. PCC 7126, *Calothrix* sp. PCC 6303 and *Aliterella atlantica* CENA 595, respectively, are topologically similar, although the strains are phylogenetically distant.

More than a half (56%) of PLCs obtained in this study were found in genomes of heterocytous cyanobacteria (Nostocales). Reciprocally, a high frequency of PLCs was observed (PLC present in 61%

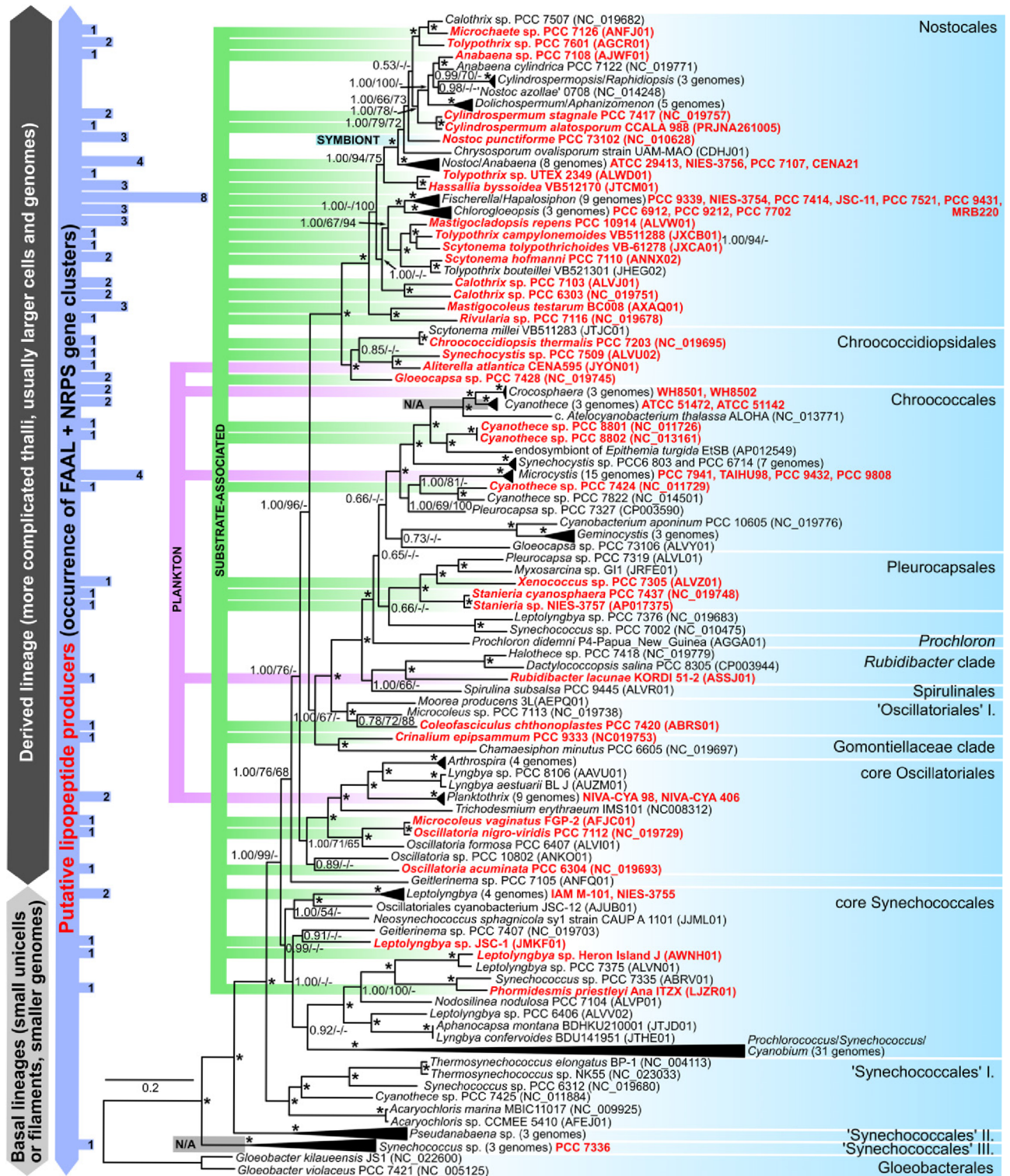


Fig. 6. Phylogenetic tree (Bayesian Inference) of cyanobacteria based on 23 housekeeping proteins from 194 representative genomes. Branch support values are shown near nodes in this form: Bayesian Inference/Maximum Likelihood/Neighbor-Joining; full supports from all methods are indicated by asterisks. Cyanobacterial strains/isolates with genomes containing putative lipopeptide gene clusters (PLCs) are printed in bold. The abundance of PLCs in individual clades is indicated by numbers and bars of a respective length on the left side of the tree. General life strategies (plankton, substrate-associated, symbiont) are visualized for PLC-containing cyanobacteria where available. The clades are classified into operational taxonomic categories corresponding to the current botanical system of cyanobacteria. PLCs are more frequent in later branching cyanobacteria exhibiting a substrate-associated mode of life. [Color figure can be viewed at wileyonlinelibrary.com]



of genomes) among nostocalean isolates included in the tree. PLCs were also very frequent in the genomes of two phylogenetically independent coccoid lineages including baeocyte-forming strains, Pleurocapsales, and Chroococciopsidales (present in 60% and 80% of genomes, respectively), although the number of sequenced isolates in these groups is still rather low and is not sufficient for a representative evaluation. A relatively lower incidence of PLCs was observed in the non-baeocyte-forming coccoid lineage of Chroococcales (present in 28% of genomes). Derived filamentous non-branched lineages of Oscillatoriales were resolved as polyphyletic in two independent clusters in our tree (Fig. 6), and altogether their PLCs had a frequency comparable to Chroococcales (25%). PLCs were relatively scarce in the basal paraphyletic branches of cyanobacterial strains covered under the traditional order Synechococcales. Only six PLCs (8% of the total) appeared each in a single synechococcalean strain (10% of synechococcalean strains positive for PLC). A single PLC was found in *Rubidibacter lacunae* KORDI-52 and *Crinalium epipsammum* PCC 9333, strains that are outside any of these groups.

Generally, a majority of the PLC-containing strains clustered in derived cyanobacterial lineages exhibiting larger cells, more diversified thalli, and non-parietal thylakoid patterns. PLCs were especially frequent in heterocytous and derived coccoid cyanobacteria. These types typically possessed genomes of a greater size. The average genome size of PLC-containing isolates was around 7 Mb compared to an average size of ~3 Mb in the genomes that lacked PLCs (Fig. S2). This difference was statistically highly significant ( $P < 0.001$ ). While some of the genome assemblies lacking PLCs were large (>11 Mb in extreme), none of the PLC-containing ones had a size under 4 Mb, including the few positive hits in synechococcalean strains.

**Ecological distribution of PLCs:** To find a possible relationship between lipopeptide production and strain ecology, we evaluated the origin of the strains with available genomes. Genome information, either in the form of a complete genome assembly or WGS, for 376 different cyanobacterial taxa was extracted from GenBank (see Methods). Of these, 59 were positive for the presence of PLC (16%; Table 1). A high frequency of PLCs (48%) was found among substrate-associated cyanobacteria. PLCs were especially abundant in strains of terrestrial origin (67%), followed by substrate-associated saline strains (50%), substrate-associated freshwater strains (46%), substrate-associated thermophilic strains (43%), and substrate-associated marine strains (25%). In contrast, only 4% of total 237 planktic cyanobacterial strains were positive for PLC presence. This comprised 13% of total freshwater planktic strains, only 2% among isolates of marine plankton origin and none among saline planktic strains. Only one of the 16 symbiotic strains (i.e., 6%) contained a PLC.

With one exception, PLCs of freshwater planktic cyanobacteria were highly similar to the microginin-producing cluster of *Planktothrix prolifica* (NIVA-CYA 98). The only exception was a cluster found in *Microcystis aeruginosa* PCC 9808, which is, however, related to a known *psm3* orphan gene cluster from *M. aeruginosa* K-139. Four clusters were found in marine planktic cyanobacteria, two of them in related *Crocospaera* strains (WH8501 and WH8502); however, contigs were not long enough to classify the PLC. Another two were found in *Aliterella atlantica* CENA595 and *Rubidibacter* sp. KORDI-52 and classified according to cluster topology as in group 2 and group 3, respectively.

PLCs of substrate-associated cyanobacteria exhibited a great diversity, particularly in A-domain count, the content of tailoring enzymes and relative position of FAAL encoding region.

#### DISCUSSION

In cyanobacteria, FAAL is thus far the only enzymatic domain proved to incorporate fatty acids into secondary metabolites produced by PKS-NRPS machineries, although sometimes designated as acyl-ACP synthase (Edwards et al. 2004, Ramaswamy et al. 2007). Absence of FAAL in biosynthetic gene clusters of hassalidins (Vestola et al. 2014) and anaenolysins (Shishido et al. 2015) suggests that cyanobacteria may employ alternative mechanisms of FA recruitment into nascent lipopeptide, such as the one observed in surfactin biosynthesis by *Bacillus subtilis* (Kraas et al. 2010). Nevertheless, our approach in this study addresses biosynthetic pathways that employ FAAL, since its presence in a biosynthetic gene cluster alongside a sufficient

TABLE 1. Summary of the original habitat (isolation source) information on cyanobacterial strains with available whole-genome data. Proportion of strains/isolates that possess putative lipopeptide clusters in their genomes is indicated.

Category	Sub-category	Number of strains	Number of positive strains (%)
Substrate-associated	Freshwater	24	11 (46)
	Marine	16	4 (25)
	Saline	2	1 (50)
	Terrestrial	24	16 (67)
	Thermal	21	9 (43)
	Unclassified	5	3 (60)
	Total	92	44 (48)
Planktic	Freshwater	50	7 (14)
	Marine	191	4 (2)
	Saline	8	0 (0)
	Unclassified	1	0 (0)
	Total	250	11 (4)
Symbiotic		16	1 (6)
Unclassified		18	4 (22)
Total		376	60 (16)

number of NRPS modules provides the basis for lipopeptide production.

A similar approach was previously used by Liu and colleagues, who screened 283 cyanobacterial genomes available via NCBI, to identify biosynthetic gene clusters that encode pathways for production of secondary metabolites with 4-methylproline (4mPro), by a BLAST search for homologues of genes *nosE* and *nosF*, essential for 4mPro synthesis (Liu et al. 2014). In this study, this method was elaborated and partially automated. To put more emphasis on conserved regions of the FAAL amino acid sequence, a constructed FAAL-PSSM was used as a query for the BLAST search. Furthermore, a Python/Biopython script/module was written to automate data acquisition from public domains and provide efficient visualization of putative clusters. Such modifications allowed distant functional homologues to score better in the BLAST search and the prepared script greatly facilitated the processing of PLCs.

In the present contribution, we have demonstrated that FAAL is frequently included in cyanobacterial NRPS gene clusters. For easier orientation, we have classified the PLCs in four groups. The first one included PLCs with a topology virtually identical to the published microginin gene cluster (Kramer 2010). We assume that these clusters (M-2 to M-5) produce structural variants of microginin. The rest of the found clusters were classified according to the topology, with special focus on FAAL and adjacent genes.

In cyanobacteria, FAAL is considered to be involved in biosynthesis of jamaicamide, hectochlorin, hapalosin, microginin, puwainaphycin, and colombamides. In all of the mentioned clusters, FAAL transfers the activated acyl residue (acyl-adenylate) to T-domain for subsequent elongation by PKS. Only afterwards NRPS is employed in the biosynthetic pathway. In these gene clusters, at least one PKS gene is always encoded downstream of the FAAL gene and upstream of the first NRPS gene. Such an arrangement corresponds to the topology found in group 3 and microginin gene clusters. Interestingly, none of the known pathways has the arrangement corresponding to group 1 and group 2, where FAAL is immediately followed by NRPS genes.

Furthermore, FAAL in all known cyanobacterial biosynthesis pathways is encoded as a stand-alone protein. The only exception to this is the hapalosin biosynthetic gene cluster, in which the *hapA* gene encodes FAAL fused to an acyl-carrier domain (Micallef et al. 2015b). In addition, FAAL is encoded as a stand-alone protein in the taromycin biosynthetic gene cluster from the marine actinomycete *Saccharomonospora* sp. CNQ-490 (Yamanaka et al. 2014) and also in the daptomycin biosynthetic cluster from the genus *Bacillus* (Miao et al. 2005). Interestingly, in the PLCs presented herein FAAL is

encoded as a stand-alone protein only in six cases (G1-1, G1-2, G3-2, G3-3, G3-4, and S-G3-4, not counting microginin-like PLCs and G3-19 which is *puwC*). Most frequently it is fused to a thiolation/acyl-carrier domain, followed by the NRPS domain or PKS. The latter kind of topology, where the fatty acid activating enzyme is fused to the next carrier domain and PKS is observed in biosynthesis pathways of iturin, bacillomycin, mycosubtilin, and paenilarvins (Duitman et al. 1999, Tsuge et al. 2001, Koumoutsis et al. 2004, Sood et al. 2014).

To the best of our knowledge, there is no mechanistic evidence on the function of the frequently observed FAAL/NRPS hybrids (group 2, group 1 except G1-1, G1-2) and their involvement in lipopeptide biosynthesis.

Of the tailoring enzymes, the most frequent were dioxygenases, found in 44 of 59 clusters. Furthermore, multiple dioxygenases per cluster were frequently observed. Enzymes with a dioxygenase signature facilitate various kinds of reactions such as hydroxylation, halogenation, ring formation, epimerization, and desaturation (Martinez and Hausinger 2015). Relevant to non-ribosomally synthesized peptides are studies of SyrB2, a dioxygenase from *Pseudomonas syringae*, which was shown to chlorinate threonine residues during non-ribosomal synthesis of syringomycin (Vaillancourt et al. 2005). Proteins that contain dioxygenase domain were also suggested to act as halogenases in cyanobacteria, e.g., HctB, BarB1, and BarB2 (Chang et al. 2002, Edwards et al. 2004, Ramaswamy et al. 2007). As a result of their functional variability, it is not straightforward to predict the actual tailoring activity of the protein.

The database mining performed in this study revealed the occurrence of PLC in ~16% of publicly available genomes (59 of 376 genomes). Previous surveys reported an overall frequency of NRPS and hybrid NRPS/PKS gene clusters in cyanobacterial genomes between 51% and 56% (Shih et al. 2013, Wang et al. 2014). Calteau and colleagues have recently shown that only around 10% of NRPS and PKS/NRPS gene clusters in cyanobacterial genomes have known products (Calteau et al. 2014). Our results suggest that non-ribosomal secondary metabolite clusters containing FAAL are frequent and their products may contribute to a large extent to the chemical diversity of cyanobacterial compounds.

The relationship between genome size and abundance of NRPS clusters in general is well documented in both bacteria and cyanobacteria (Shih et al. 2013, Wang et al. 2014). Wang et al. (2014) have reported that bacterial genomes smaller than 2 Mbp always lack this kind of gene clusters, while they are most frequent in genomes >4 Mbp. In cyanobacteria, the NRPS and PKS/NRPS clusters are most abundant in heterocytous (Nostocales) and baeocytous (Chroococcioidiales),

Pleurocapsales) lineages, followed by Chroococcales and Oscillatoriales (Shih et al. 2013, Calteau et al. 2014). These are the phylogenetically later branching clades of cyanobacteria. Their genome size is generally high, reflecting their great morphological and ultrastructural complexity and variability in life strategies. The distribution of PLCs in this study obviously follows this general trend (Fig. 6).

Aside from this general trend, our bioinformatic search revealed that the distribution of PLCs reflects the life history of the strain and that PLCs are especially abundant in substrate-associated cyanobacteria. From publications describing novel lipopeptide structures in cyanobacteria (Table S1 in the Supporting Information), we can conclude that a much greater diversity of lipopeptide structures was isolated from various substrate-associated cyanobacteria (mainly from marine environment) in comparison to planktic isolates. The majority of reports of lipopeptide production in plankton communities and isolates are connected to microginins, which comprises 42 of 59 described cases of lipopeptide production in plankton (15/36 of strains). The second class of lipopeptides frequently observed in plankton is hassalidins, which are also glycosylated and a part of their structural variability comes from different saccharides used for glycosylation (12 variants, not considering glycosylation). Furthermore, five cyanopeptolins and four aeruginosins with octanoic acid are known to be produced by planktic cyanobacteria. Taking into consideration this limited structural variability, the variability in gene clusters responsible for lipopeptide biosynthesis in planktic cyanobacteria may also be restricted. This is in agreement with previously published data on microginins (Ishida et al. 2000, Humbert et al. 2013) and hassalidins (Vestola et al. 2014). On the contrary, 60 biosynthetic gene clusters with a gene coding for FAAL were found in genomes of substrate-associated cyanobacteria. Although based on our data, we are not able to assess the number of chemical classes produced by PLCs found in substrate-associated cyanobacteria, they are clearly more frequent and more diverse in their domain organization than PLCs in planktic cyanobacteria. This is in agreement with the broad chemical diversity of lipopeptides isolated from various substrate-associated cyanobacteria (Table S1).

Lipopeptides have not yet been examined for their eco-physiological role in cyanobacteria. However, there is evidence on the role of lipopeptides in Firmicutes (bacteria), genus *Bacillus*. Surfactin, a lipopeptide produced among others by *B. subtilis* reduces the surface tension of water as well as the friction between bacterial cells and their substrate, allowing them to form bundles and spread out over the surrounding surface (Kinsinger et al. 2003, van Gestel et al. 2015). Reduction in surface tension is not the only mechanism by which this lipopeptide influences its producer. It acts also as a signaling

molecule that governs the development of structured biofilms on agar plates as well as in the rhizosphere of *Arabidopsis thaliana* (Bais et al. 2004, Hofemeister et al. 2004, Lopez et al. 2009). Both as a surfactant and a signaling molecule, surfactin is important to biofilm development in *Bacillus*.

Although cyanobacteria positive in PLC content are widely distributed across many cyanobacterial phylogenetic lineages, it is striking that the majority of them were identified in cyanobacteria isolated from or capable of forming structured biofilms, mats or colonies (e.g., *Nostoc*, *Tolypothrix*, *Gloeocapsa*, *Microcoleus*, *Leptolyngbya*). Thus, we can hypothesize that based on their ecological distribution lipopeptides can play an analogous role as lipopeptides in Firmicutes or other bacteria. Experimental testing of the actual lipopeptide production and their eco-physiological functions in biofilm-forming cyanobacteria that contain FAAL-NRPS clusters is a highly encouraged direction of further research.

Finally, many lipopeptides are known to possess antibacterial or antifungal properties (schizotrin, Pergament and Carmeli 1994, hassalidins, Neuhof et al. 2005, Vestola et al. 2014). *Bacillus subtilis*, aside from surfactin, also produces fengycins, bacilomycins, and iturins, which are not essential to biofilm formation but possess antifungal or antimicrobial capabilities. Furthermore, their production prevents infections of plants with which the bacterium forms symbiotic relationships (Romero et al. 2007). Results presented herein show that diverse PLCs are abundant in substrate-associated cyanobacteria, which highlights the importance of this ecological group in screening for novel molecules of interest such as potential pharmaceuticals and environmental toxins.

Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the program "Projects of Large Research, Development, and Innovations Infrastructures." The study was financially supported by the Czech Science Foundation project No. 16-09381S, the Grant Agency of the University of South Bohemia project no. 04-158/2016/P and the Ministry of Education, Youth and Sports of the Czech Republic - National Programme of Sustainability I, ID: LO1416 and Project Algamic (CZ.1.05/2.1.00/19.0392.). The authors hereby declare that no competing interests exist.

- Bais, H. P., Fall, R. & Vivanco, J. M. 2004. Biocontrol of *Bacillus subtilis* against infection of *Arabidopsis* roots by *Pseudomonas syringae* is facilitated by biofilm formation and surfactin production. *Plant Physiol.* 134:307–19.
- Calteau, A., Fewer, D. P., Laïfi, A., Coursin, T., Laurent, T., Jokela, J., Kerfeld, C. A., Sivonen, K., Piel, J. & Gugger, M. 2014. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. *BMC Genom.* 15:14.
- Chang, Z. X., Flatt, P., Gerwick, W. H., Nguyen, V. A., Willis, C. L. & Sherman, D. H. 2002. The barbamide biosynthetic gene cluster: a novel marine cyanobacterial system of mixed polyketide synthase (PKS)-non-ribosomal peptide synthetase

- (NRPS) origin involving an unusual trichloroleucyl starter unit. *Gene* 296:235–47.
- Coates, R. C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D. H., Allen, E. E., Gerwick, L. & Gerwick, W. H. 2014. Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS ONE* 9:e85140.
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B. & de Hoon, M. J. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25:1422–3.
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. 2011. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27:1164–5.
- Dittmann, E., Fewer, D. P. & Neilan, B. A. 2013. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. *FEMS Microbiol. Rev.* 37:23–43.
- Duitman, E. H., Hamoen, L. W., Rembold, M., Venema, G., Seitz, H., Saenger, W., Bernhard, F. et al. 1999. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc. Natl. Acad. Sci. USA* 96:13294–9.
- Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A. & Gerwick, W. H. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* 11:817–33.
- Fowler, V. G., Boucher, H. W., Corey, G. R., Abrutyn, E., Karchmer, A. W., Rupp, M. E., Levine, D. P. et al. 2006. Daptomycin versus standard therapy for bacteremia and endocarditis caused by *Staphylococcus aureus*. *N. Engl. J. Med.* 355:653–65.
- Gascuel, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14:685–95.
- Gerwick, W. H. & Moore, B. S. 2012. Lessons from the past and charting the future of marine natural products drug discovery and chemical biology. *Chem. Biol.* 19:85–98.
- van Gestel, J., Vlamakis, H. & Kolter, R. 2015. From cell differentiation to cell collectives: *Bacillus subtilis* uses division of labor to migrate. *PLoS Biol.* 13:29.
- Gouy, M., Guindon, S. & Gascuel, O. 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* 27:221–4.
- Goyal, A., Verma, P., Anandhakrishnan, M., Gokhale, R. S. & Sankaranarayanan, R. 2012. Molecular basis of the functional divergence of fatty acyl-AMP ligase biosynthetic enzymes of *Mycobacterium tuberculosis*. *J. Mol. Biol.* 416:221–38.
- Guenzi, E., Galli, G., Grgurina, I., Pace, E., Ferranti, P. & Grandi, G. 1998. Coordinate transcription and physical linkage of domains in surfactin synthetase are not essential for proper assembly and activity of the multienzyme complex. *J. Biol. Chem.* 273:14403–10.
- Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52:696–704.
- Hansen, D. B., Bumpus, S. B., Aron, Z. D., Kelleher, N. L. & Walsh, C. T. 2007. The loading module of mycosubtilin: an adenylation domain with fatty acid selectivity. *J. Am. Chem. Soc.* 129:6366–7.
- Hofemeister, J., Conrad, B., Adler, B., Hofemeister, B., Feesche, J., Kucheryava, N., Steinborn, G. et al. 2004. Genetic analysis of the biosynthesis of non-ribosomal peptide- and polyketide-like antibiotics, iron uptake and biofilm formation by *Bacillus subtilis* AI/3. *Mol. Genet. Genomics* 272:363–78.
- Hrouzek, P., Kuzma, M., Černý, J., Novák, P., Fišer, R., Šimek, P., Lukešová, A. & Kopecký, J. 2012. The cyanobacterial cyclic lipopeptides puwainaphycins F/G are inducing necrosis via cell membrane permeabilization and subsequent unusual actin relocalization. *Chem. Res. Toxicol.* 25:1203–11.
- Humbert, J. F., Barbe, V., Latifi, A., Gugger, M., Calteau, A., Coursin, T., Lajus, A. et al. 2013. A tribute to disorder in the genome of the bloom-forming freshwater cyanobacterium *Microcystis aeruginosa*. *PLoS ONE* 8:e70747.
- Ishida, K., Kato, T., Murakami, M., Watanabe, M. & Watanabe, M. F. 2000. Microginins, zinc metalloproteases inhibitors from the cyanobacterium *Microcystis aeruginosa*. *Tetrahedron* 56:8643–56.
- Kampa, A., Gagunashvili, A. N., Gulder, T. A. M., Morinaka, B. I., Daolio, C., Godejohann, M., Miao, V. P. W., Piel, J. & Andresson, O. S. 2013. Metagenomic natural product discovery in lichen provides evidence for a family of biosynthetic pathways in diverse symbioses. *Proc. Natl. Acad. Sci. USA* 110: E3129–37.
- Katoh, K. & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30:772–80.
- Kinsinger, R. F., Shirk, M. C. & Fall, R. 2003. Rapid surface motility in *Bacillus subtilis* is dependent on extracellular surfactin and potassium ion. *J. Bacteriol.* 185:5627–31.
- Komárek, J., Kaštovský, J., Mareš, J. & Johansen, J. R. 2014. Taxonomic classification of cyanoprokaryotes (cyanobacterial genera) 2014, using a polyphasic approach. *Preslia* 86:295–335.
- Koumoutsis, A., Chen, X. H., Henne, A., Liesegang, H., Hitzeroth, G., Franke, P., Vater, J. & Borriss, R. 2004. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J. Bacteriol.* 186:1084–96.
- Kraas, F. I., Helmetag, V., Wittmann, M., Strieker, M. & Marahiel, M. A. 2010. Functional dissection of surfactin synthetase initiation module reveals insights into the mechanism of lipoinitiation. *Chem. Biol.* 17:872–80.
- Kramer, D. 2010. Microginin producing proteins and nucleic acids encoding a microginin gene cluster as well as methods for creating novel microginins. Patent US 7846686 B2.
- Liu, L., Jokela, J., Herfindal, L., Wahlsten, M., Sinkkonen, J., Permi, P., Fewer, D. P., Döskeland, S. O. & Sivonen, K. 2014. 4-Methylproline guided natural product discovery: co-occurrence of 4-hydroxy- and 4-methylprolines in nostoweipeptins and nostopeptolides. *ACS Chem. Biol.* 9:2646–55.
- Lopez, D., Fischbach, M. A., Chu, F., Losick, R. & Kolter, R. 2009. Structurally diverse natural products that cause potassium leakage trigger multicellularity in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* 106:280–5.
- Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* 97:2651–74.
- Marchler-Bauer, A. & Bryant, S. H. 2004. CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* 32:W327–31.
- Marchler-Bauer, A., Derbyshire, M. K., Gonzales, N. R., Lu, S., Chitsaz, F., Geer, L. Y., Geer, R. C. et al. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res.* 43:D222–6.
- Mareš, J., Hájek, J., Urajová, P., Kopecký, J. & Hrouzek, P. 2014. A hybrid non-ribosomal peptide/polyketide synthetase containing fatty-acyl ligase (FAAL) synthesizes the  $\beta$ -amino fatty acid lipopeptides puwainaphycins in the cyanobacterium *Cylindrospermum alatosporum*. *PLoS ONE* 9:e111904.
- Martinez, S. & Hausinger, R. P. 2015. Catalytic mechanisms of Fe (II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* 290:20702–11.
- Miao, V., Coëffet-LeGal, M. F., Brian, P., Brost, R., Penn, J., Whiting, A., Martin, S. et al. 2005. Daptomycin biosynthesis in *Streptomyces roseosporus*: cloning and analysis of the gene cluster and revision of peptide stereochemistry. *Microbiology SGM* 151:1507–23.
- Micallef, M. L., D'Agostino, P. M., Al-Sinawi, B., Neilan, B. A. & Moffitt, M. C. 2015a. Exploring cyanobacterial genomes for natural product biosynthesis pathways. *Mar. Genomics* 21:1–12.
- Micallef, M. L., D'Agostino, P. M., Sharma, D., Viswanathan, R. & Moffitt, M. C. 2015b. Genome mining for natural product

- biosynthetic gene clusters in the Subsection V cyanobacteria. *BMC Genom.* 16:669.
- Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., Passarotti, M., Kaufman, S. & O'Leary, M. A. 2015. A RESTful API for access to phylogenetic tools via the CIPRES Science Gateway. *Evol. Bioinform.* 11:43–8.
- Mohanty, D., Sankaranarayanan, R. & Gokhale, R. S. 2011. Fatty acyl-AMP ligases and polyketide synthases are unique enzymes of lipid biosynthetic machinery in *Mycobacterium tuberculosis*. *Tuberculosis* 91:448–55.
- Neuhof, T., Schmieder, P., Preussel, K., Dieckmann, R., Pham, H., Bartl, F. & von Döhren, H. 2005. Hassallidin A, a glycosylated lipopeptide with antifungal activity from the cyanobacterium *Hassallia* sp. *J. Nat. Prod.* 68:695–700.
- Nishizawa, A., Bin Arshad, A., Nishizawa, T., Asayama, M., Fujii, K., Nakano, T., Harada, K. & Shirai, M. 2007. Cloning and characterization of a new hetero-gene cluster of nonribosomal peptide synthetase and polyketide synthase from the cyanobacterium *Microcystis aeruginosa* K-139. *J. Gen. Appl. Microbiol.* 53:17–27.
- Oftedal, L., Myhren, L., Jokela, J., Gausdal, G., Sivonen, K., Døskeland, S. O. & Herfindal, L. 2012. The lipopeptide toxins anabaenolysin A and B target biological membranes in a cholesterol-dependent manner. *Biochim. Biophys. Acta* 1818:3000–9.
- Pergament, I. & Carmeli, S. 1994. Schizotrin A, a novel antimicrobial cyclic peptide from a cyanobacterium. *Tetrahedron Lett.* 35:8473–6.
- Ramaswamy, A. V., Sorrels, C. M. & Gerwick, W. H. 2007. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lyngbya majuscula*. *J. Nat. Prod.* 70:1977–86.
- Romero, D., de Vicente, A., Rakotoaly, R. H., Dufour, S. E., Veening, J. W., Arrebola, E., Cazorla, F. M., Kuipers, O. P., Paquot, M. & Perez-Garcia, A. 2007. The iturin and fengycin families of lipopeptides are key factors in antagonism of *Bacillus subtilis* toward *Podosphaera fusca*. *Mol. Plant Microbe Interact.* 20:430–40.
- Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A. & Huelsenbeck, J. P. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61:539–42.
- Rouge, T. B., Rohrlack, T., Nederbragt, A. J., Kristensen, T. & Jakobsen, K. S. 2009. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. *BMC Genom.* 10:396.
- Shih, P. M., Wu, D., Latifi, A., Axen, S. D., Fewer, D. P., Talla, E., Calteau, A. et al. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci. USA* 110:1053–8.
- Shishido, T. K., Jokela, J., Kolehmainen, C. T., Fewer, D. P., Wahlsten, M., Wang, H., Rouhiainen, L., Rizzi, E., De Bellis, G., Permi, P. & Sivonen, K. 2015. Antifungal activity improved by coproduction of cyclodextrins and anabaenolysins in Cyanobacteria. *Proc. Natl. Acad. Sci. USA* 112:13669–74.
- Sood, S., Steinmetz, H., Beims, H., Mohr, K. I., Stadler, M., Djukic, M., von der Ohe, W., Steinert, M., Daniel, R. & Müller, R. 2014. Paenilarvins: iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. *ChemBioChem* 15:1947–55.
- Stachelhaus, T., Mootz, H. D. & Marahiel, M. A. 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chem. Biol.* 6:493–505.
- Stamatakis, A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–3.
- Tomek, P., Hrouzek, P., Kuzma, M., Sýkora, J., Fišer, R., Černý, J., Novák, P. et al. 2015. Cytotoxic lipopeptide muscotoxin A, isolated from soil cyanobacterium *Desmonostoc muscorum*, permeabilizes phospholipid membranes by reducing their fluidity. *Chem. Res. Toxicol.* 28:216–24.
- Tsuge, K., Akiyama, T. & Shoda, M. 2001. Cloning, sequencing, and characterization of the iturin A operon. *J. Bacteriol.* 183:6265–73.
- Vaillancourt, F. H., Yin, J. & Walsh, C. T. 2005. SyrB2 in syringomycin E biosynthesis is a nonheme FeII alpha-ketoglutarate- and O<sub>2</sub>-dependent halogenase. *Proc. Natl. Acad. Sci. USA* 102:10111–6.
- Van Wagoner, R. M., Drummond, A. K. & Wright, J. L. C. 2007. Biogenetic diversity of cyanobacterial metabolites. *Adv. Appl. Microbiol.* 61:89–217.
- Vestola, J., Shishido, T. K., Jokela, J., Fewer, D. P., Aitio, O., Permi, P., Wahlsten, M., Wang, H., Rouhiainen, L. & Sivonen, K. 2014. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. *Proc. Natl. Acad. Sci. USA* 111: E1909–17.
- Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L. & Sivonen, K. 2014. Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. USA* 111:9259–64.
- Welker, M. & von Döhren, H. 2006. Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiol. Rev.* 30:530–63.
- Yamanaka, K., Reynolds, K. A., Kersten, R. D., Ryan, K. S., Gonzalez, D. J., Nizet, V., Dorrestein, P. C. & Moore, B. S. 2014. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proc. Natl. Acad. Sci. USA* 111:1957–62.

### Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

**Figure S1.** PSI-BLAST results.

**Figure S2.** Full list of microginin gene clusters.

**Figure S3.** Full list of group 1 putative lipopeptide gene clusters.

**Figure S4.** Full list of group 2 putative lipopeptide gene clusters.

**Figure S5.** Full list of group 3 putative lipopeptide gene clusters.

**Figure S6.** Genome size comparison between cyanobacterial genomes containing and lacking putative lipopeptide gene clusters (PLC).

**Table S1.** List of lipopeptides isolated from cyanobacteria with fully characterized chemical structure based on the combination of NMR and MS methods.

**Appendix S1.** Supplementary references.

**Appendix S2.** BLAST hits analysis.

### **4.3. Paper III.**



# Alternative Biosynthetic Starter Units Enhance the Structural Diversity of Cyanobacterial Lipopeptides

Jan Mareš,<sup>a,b,c</sup> Jan Hájek,<sup>b,c</sup> Petra Urajová,<sup>b</sup> Andreja Kust,<sup>a,b,c</sup> Jouni Jokela,<sup>d</sup> Kumar Saurav,<sup>b</sup> Tomáš Galica,<sup>b,c</sup> Kateřina Čapková,<sup>a</sup> Antti Mattila,<sup>d</sup> Esa Haapaniemi,<sup>e,f</sup> Perttu Permi,<sup>e</sup> Ivar Myrsterud,<sup>g</sup> Olav M. Skulberg,<sup>h</sup> Jan Karlsen,<sup>i</sup> David P. Fewer,<sup>d</sup> Kaarina Sivonen,<sup>d</sup> Hanne Hjorth Tønnesen,<sup>i</sup> Pavel Hrouzek<sup>b,c</sup>

<sup>a</sup>The Czech Academy of Sciences, Biology Centre, Institute of Hydrobiology, České Budějovice, Czech Republic

<sup>b</sup>The Czech Academy of Sciences, Institute of Microbiology, Center Algatech, Třeboň, Czech Republic

<sup>c</sup>University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic

<sup>d</sup>Department of Microbiology, Viikki Biocenter, University of Helsinki, Helsinki, Finland

<sup>e</sup>Department of Chemistry, University of Jyväskylä, Jyväskylä, Finland

<sup>f</sup>Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, Jyväskylä, Finland

<sup>g</sup>Department of Biosciences, University of Oslo, Oslo, Norway

<sup>h</sup>Norwegian Institute for Water Research (NIVA), Oslo, Norway

<sup>i</sup>School of Pharmacy, University of Oslo, Oslo, Norway

**ABSTRACT** Puwainaphycins (PUWs) and minutissamides (MINs) are structurally analogous cyclic lipopeptides possessing cytotoxic activity. Both types of compound exhibit high structural variability, particularly in the fatty acid (FA) moiety. Although a biosynthetic gene cluster responsible for synthesis of several PUW variants has been proposed in a cyanobacterial strain, the genetic background for MINs remains unexplored. Herein, we report PUW/MIN biosynthetic gene clusters and structural variants from six cyanobacterial strains. Comparison of biosynthetic gene clusters indicates a common origin of the PUW/MIN hybrid nonribosomal peptide synthetase and polyketide synthase. Surprisingly, the biosynthetic gene clusters encode two alternative biosynthetic starter modules, and analysis of structural variants suggests that initiation by each of the starter modules results in lipopeptides of differing lengths and FA substitutions. Among additional modifications of the FA chain, chlorination of minutissamide D was explained by the presence of a putative halogenase gene in the PUW/MIN gene cluster of *Anabaena minutissima* strain UTEX B 1613. We detected PUW variants bearing an acetyl substitution in *Symplocastrum muelleri* strain NIVA-CYA 644, consistent with an *O*-acetyltransferase gene in its biosynthetic gene cluster. The major lipopeptide variants did not exhibit any significant antibacterial activity, and only the PUW F variant was moderately active against yeast, consistent with previously published data suggesting that PUWs/MINs interact preferentially with eukaryotic plasma membranes.

**IMPORTANCE** Herein, we deciphered the most important biosynthetic traits of a prominent group of bioactive lipopeptides. We reveal evidence for initiation of biosynthesis by two alternative starter units hardwired directly in the same gene cluster, eventually resulting in the production of a remarkable range of lipopeptide variants. We identified several unusual tailoring genes potentially involved in modifying the fatty acid chain. Careful characterization of these biosynthetic gene clusters and their diverse products could provide important insight into lipopeptide biosynthesis in prokaryotes. Some of the variants identified exhibit cytotoxic and antifungal properties, and some are associated with a toxigenic biofilm-forming strain. The findings may prove valuable to researchers in the fields of natural product discovery and toxicology.

**KEYWORDS** biosynthesis, cyanobacteria, fatty acyl-AMP ligase, lipopeptides, nonribosomal peptide synthetase

**Citation** Mareš J, Hájek J, Urajová P, Kust A, Jokela J, Saurav K, Galica T, Čapková K, Mattila A, Haapaniemi E, Permi P, Myrsterud I, Skulberg OM, Karlsen J, Fewer DP, Sivonen K, Tønnesen HH, Hrouzek P. 2019. Alternative biosynthetic starter units enhance the structural diversity of cyanobacterial lipopeptides. *Appl Environ Microbiol* 85:e02675-18. <https://doi.org/10.1128/AEM.02675-18>.

**Editor** Marie A. Elliot, McMaster University

**Copyright** © 2019 American Society for Microbiology. All Rights Reserved.

Address correspondence to Pavel Hrouzek, [hrouzekp@gmail.com](mailto:hrouzekp@gmail.com).

**Received** 5 November 2018

**Accepted** 28 November 2018

**Accepted manuscript posted online** 30 November 2018

**Published** 6 February 2019

**B**acterial lipopeptides are a prominent group of secondary metabolites with pharmaceutical potential as antibacterial, antifungal, anticancer, and antiviral agents (1). Compounds like fengycin, the iturin family antibiotics, octapeptins, and daptomycin are important pharmaceutical leads, the latter of which is already in clinical use (1–3). Their biological activity is the result of an amphipathic molecular structure that allows micellar interaction within the cell membranes of target organisms (4).

Lipopeptides are widespread in cyanobacteria and possess cytotoxic and antifungal activities (5–8). Puwainaphycins (PUWs) and minutissamides (MINs) are lipopeptides featuring a  $\beta$ -amino fatty acid and a 10-membered peptide ring (5, 9–11). Both classes exhibit considerable structural variability in terms of length and functionalization of the fatty acyl (FA) side chain attached to the stable peptide core (10–14). Only minor discrepancies in the lengths and substitutions of the FA chain separate these two types of lipopeptides. A wide array of bioactivities has been reported for these compounds. PUW C is a cardioactive compound (15), as demonstrated by positive inotropic activity in mouse atria, while PUW F/G exhibit cytotoxicity against human cells *in vitro* through cell membrane permeabilization (5). MIN A to L exhibited antiproliferative effects over a range of concentrations when tested against human cancer cell lines, similarly to PUWs (10, 11). The overall structural similarity suggests that PUWs and MINs share a similar biosynthetic origin. However, the biosynthetic mechanisms generating the conspicuous chemical variability remain unknown.

PUWs are synthesized by a hybrid polyketide/nonribosomal peptide synthetase (PKS/NRPS) accompanied by tailoring enzymes (12). A characteristic feature of the PUW synthetase is the fatty acyl-AMP ligase (FAAL) starter unit (12). This enzyme specifically binds and adenylates FAs and passes the activated acyl-adenylate to a downstream phosphopantetheine arm of the PKS acyl carrier protein (ACP) for further processing (12). The whole process bears a resemblance to the biosynthesis of iturin family lipopeptides (16–19), as well as small lipopeptide-like cyanobacterial metabolites like hectochlorin (20), hapalosin (21), and jamaicamide (22), as discussed previously (23). Bacterial FAAL enzymes originate from basal cell metabolism and likely evolved from fatty acyl-coenzyme A (CoA) ligases (FACLs) following a specific insertion that hampered subsequent ligation to reduced CoA (CoASH) (24) or altered the catalytic conformation (25). FAAL enzymes play an important role in the assembly of other metabolites, including olefins (26) and unusual lipids (27), in addition to lipopeptide synthesis. The exact substrate-binding mechanism employed by FAALs was demonstrated experimentally in *Mycobacterium tuberculosis* using several homologous FAAL enzymes and FA substrates as models (28). The substrate specificity of these enzymes corresponds to the structure of the substrate-binding pocket (25, 28), although it overlaps among homologs.

Herein, we combined recently developed bioinformatics and high-performance liquid chromatography combined with high-resolution tandem mass spectrometry (HPLC–HRMS-MS) approaches (13, 23) to identify biosynthesis gene clusters for PUWs/MINs in five new cyanobacterial strains and characterized the chemical variability of their products. We discuss the specific structural properties of the lipopeptide variants identified and compare the predicted functions of synthetase enzymes.

## RESULTS AND DISCUSSION

**Structural variability versus common biosynthetic origin of PUWs and MINs.** In the present study, we collected all known PUW/MIN producers (except for *Anabaena* sp. strain UIC10035). The strains, referred to herein as strains 1 to 6, were originally isolated from various soil habitats (Table 1). HPLC–HRMS-MS analysis detected multiple PUW and MIN variants in each of the strains studied (Fig. 1), ranging from 13 to 26 in strains 3 and 1, respectively (Table S1 in the supplemental material).

The MS-MS data acquired for crude extracts were used to create a molecular network (Fig. 2), analysis of which demonstrated that *Cylindrospermum* strains 1 to 3 and *Anabaena* strains 4 and 5 formed a single group with MIN A as the only variant common to all the strains (Fig. 2a). All major structural variants of these strains shared



**TABLE 1** Strains analyzed for PUW/MIN production

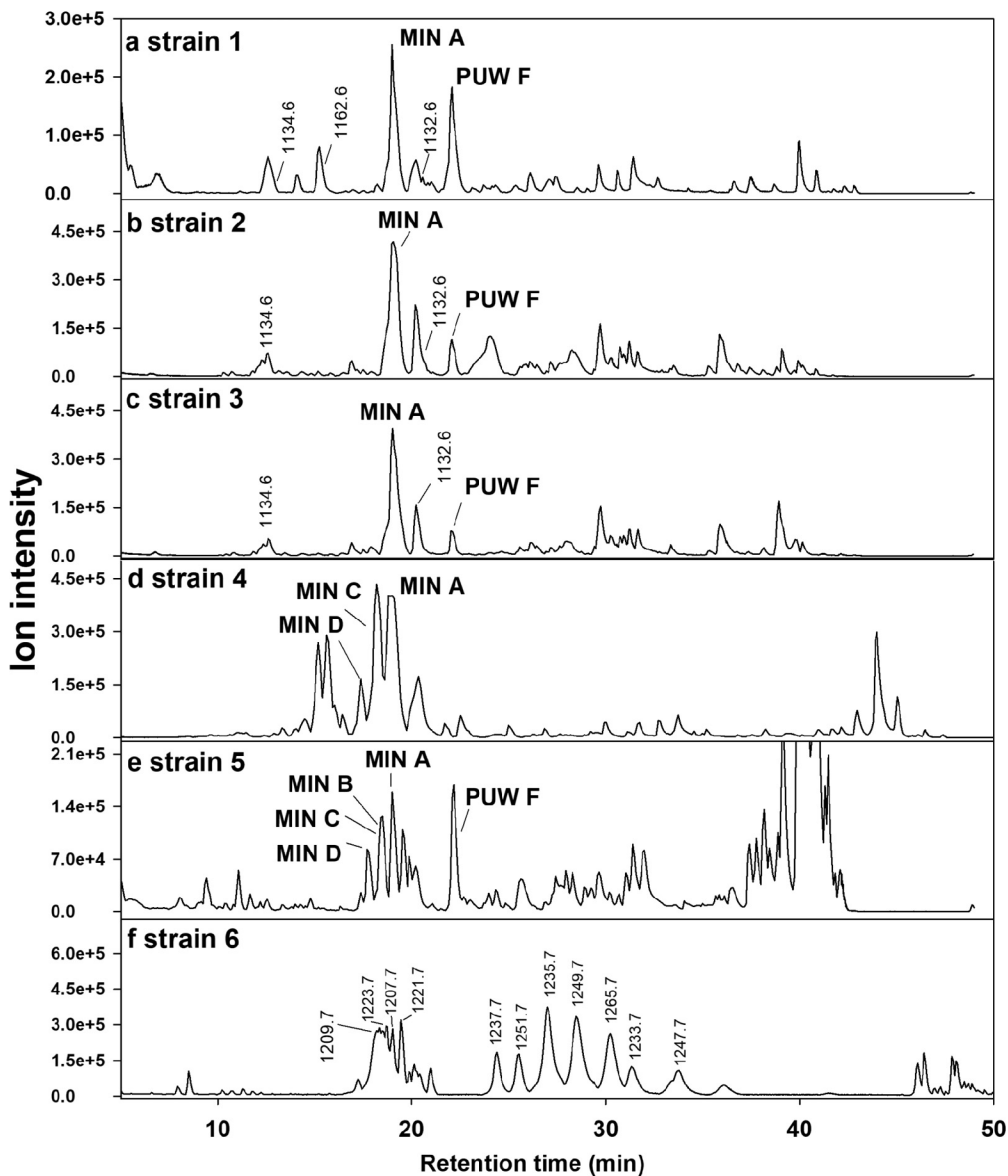
Strain no. herein	Strain	Person who isolated strain	Date	Locality, habitat	Reference
1	<i>Cylindrospermum alatosporum</i> CCALA 988	A. Lukešová	1989	Canada, Manitoba, Riding Mountain National Park, soil	55
2	<i>Cylindrospermum moravicum</i> CCALA 993	A. Lukešová	2008	Czech Republic, South Moravia, Moravian Karst, Amaterska Cave, cave sediment	55
3	<i>Cylindrospermum alatosporum</i> CCALA 994	A. Lukešová	2011	Czech Republic, Moravian Karst, earthworms collected from soil above Amaterska Cave, earthworm casings	55
4	<i>Anabaena</i> sp. UHCC-0399	M. Wahlsten	N/A	Finland, Jurmo, Southwestern Archipelago National Park, copepods	56
5	<i>Anabaena minutissima</i> UTEX B1613	T. Kantz	1967	South Texas, USA, soil	57
6	<i>Symplocastrum muelleri</i> NIVA-CYA 644	O. M. Skulberg	2009	Norway, Møre og Romsdal county, Halså municipality, western slope of Slettfjellet mountain in semiterrestrial alpine habitat, biofilm on turf in ombrotrophic blanket bog	40

the common peptide sequence FA<sup>1</sup>-Val<sup>2</sup>-dehydroaminobutyric acid (Dhb)<sup>3</sup>-Asn<sup>4</sup>-Dhb<sup>5</sup>-Asn<sup>6</sup>-Ala<sup>7</sup>-Thr<sup>8</sup>-*N*-methyl-L-asparagine (NMeAsn)<sup>9</sup>-Pro<sup>10</sup> (Fig. 3), described previously for PUW F and MIN A (5, 10). The pattern of variant production was almost identical in *Cylindrospermum* strains 2 and 3, which in addition to MIN A contained PUW F (Fig. 1, Table S1). In contrast, *Anabaena* strains 4 and 5 produced MIN C and D in addition to the major variant, MIN A (Fig. 1). The peptide core of the molecule was different in *Symplocastrum muelleri* strain 6 (Fig. 3), forming a separate group in the molecular network (Fig. 2b), with the general peptide sequence FA<sup>1</sup>-Val<sup>2</sup>-Dhb<sup>3</sup>-Thr<sup>4</sup>-Thr/Val<sup>5</sup>-Gln<sup>6</sup>-Ala<sup>7</sup>-*O*-methyl-L-threonine (OMeThr)<sup>8</sup>-NMeAsn<sup>9</sup>-Pro<sup>10</sup> (Fig. 3), identical to those of PUW A to D and MIN I, K, and L isolated previously from *Anabaena* sp. (9, 11).

The peptide cores of the variants included in the network differed to some degree, but most variation was detected in the FA moiety (Fig. 4) when crude extracts were analyzed for the presence of characteristic FA immonium fragments (13).

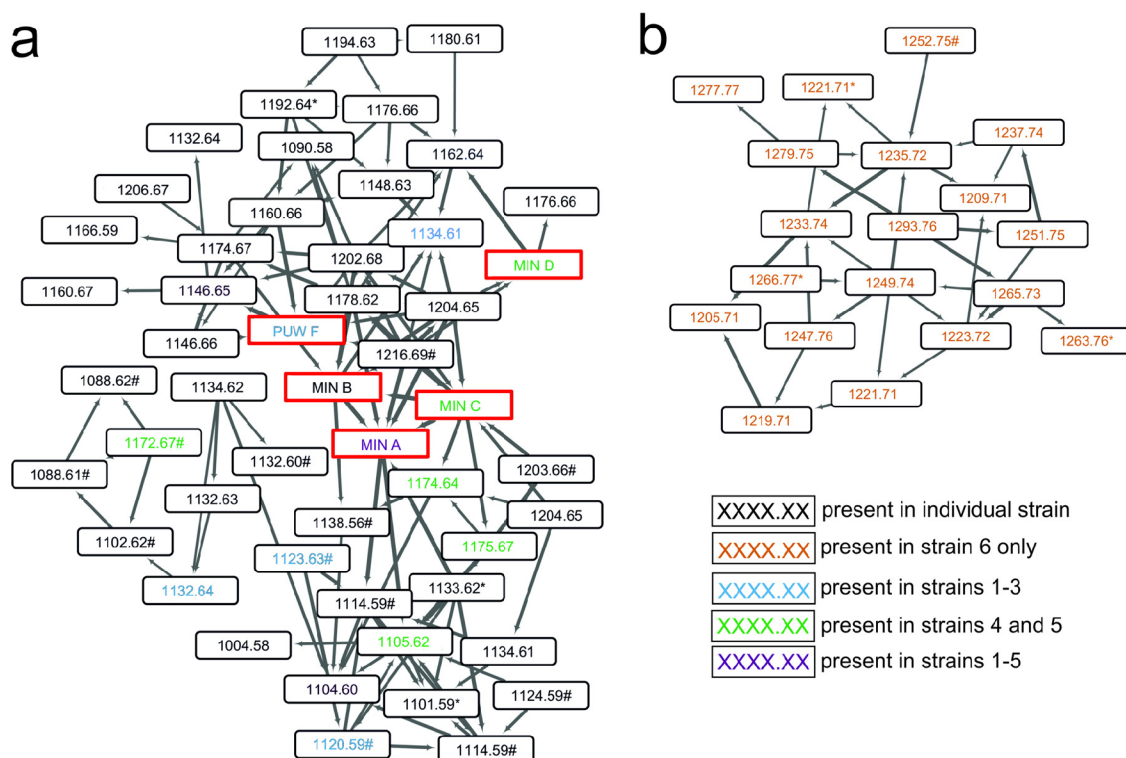
Accordingly, bioinformatic analysis identified putative PUW and MIN gene clusters in each of the five newly sequenced strains (Fig. 5, Table 2). Based on BLASTp, CDD, and AntiSMASH searches, these gene clusters exhibited synteny and functional homology with the previously characterized *puw* biosynthesis gene cluster in strain 1 (12) (Fig. 5). Therefore, our results strongly indicate a common biosynthetic origin of PUWs and MINs in cyanobacteria.

**Variability in the peptide core.** A common set of NRPS genes (*puwA* and *puwE* to *-H*) (Fig. 5) encoding a sequence of nine amino acid-incorporating modules (Fig. 6) was detected in all strains analyzed. Individual NRPS modules displayed variability in amino acid adenylation and tailoring domains that was generally congruent with the PUW/MIN peptide cores inferred using HPLC-HRMS-MS (Fig. 3). The two major types of peptide cores observed (represented by PUW A and PUW F, respectively) differed in the amino acids at positions 4 (Thr→Asn), 5 (Thr→Dhb), 6 (Gln→Asn), 7 (Ala→Gly), and 8 (Thr→OMe-Thr), as shown in Fig. 3 and Table S1. This was reflected in the predicted substrates of the corresponding adenylation domains (A domains) and by the presence of an *O*-methyltransferase domain in *PuwH* of *S. muelleri* strain 6, which is responsible for the methoxylation of Thr<sup>8</sup> (Fig. 6, Table S2). In contrast to the variability observed at the previously noted amino acid positions, the two positions adjacent to both sides of the modified fatty acid [NMeAsn<sup>9</sup>-Pro<sup>10</sup>-(FA<sup>1</sup>)-Val<sup>2</sup>-Dhb<sup>3</sup>] are conserved in all known PUW/MIN variants described here and previously (5, 9, 13–15) (Fig. 3, Table S1). Accordingly, no functional variation in A domains corresponding to these positions was observed within the deduced *PuwA*, *PuwE*, and *PuwF* proteins (Table S2). This is interesting because these four hydrophobic amino acids surround the FA moiety, which is likely responsible for the membrane disruption effect suggested previously (5). Thus, we hypothesize that such an arrangement could further support hydrophobic interactions with the lipid layer of the plasma membrane.



**FIG 1** HPLC-HRMS-MS analysis of crude extracts from the strains investigated. Major puwainaphycin (PUW) and minitissamide (MIN) variants are highlighted. For variants without complete structural information, only  $m/z$  values are shown.

For some of the other positions, minor variants were observed involving substitution of amino acids similar in structure and hydrophobicity, including Asn-Gln at position 4, Thr-Val at position 5, Ala-Gly at position 7, and Thr-Ser at position 8 (Fig. 3, Table S1), indicative of probable substrate promiscuity in their respective adenylation domains (29). The A6 domains in strains 4 and 5 activated Ala as a major substrate and Gly to a lesser extent, even though *in silico* analysis predicted Gly as their main substrate (Table S2). In strain 6, Gly was incorporated, in agreement with the predicted substrate specificity. An epimerase domain was present in each of the sixth NRPS modules (i.e., C-A6-PCP) of the pathways (Fig. 6), indicating the probable formation of a D-amino acid enantiomer at position 7 of the peptide core. Indeed, the presence of D-Ala was previously confirmed in PUW F (5) and MIN A to H (10, 11), and D-Gly was identified in MIN I to L (10, 11). In two cases, the adenylation domains A3 (PuwF) and A6 (PuwG) are capable of incorporating significantly different amino acids, such as Asn<sup>4</sup>-Thr<sup>4</sup> and Ala<sup>7</sup>-Ser<sup>7</sup>, respectively (Fig. 6). This degree of substrate promiscuity is relatively uncom-

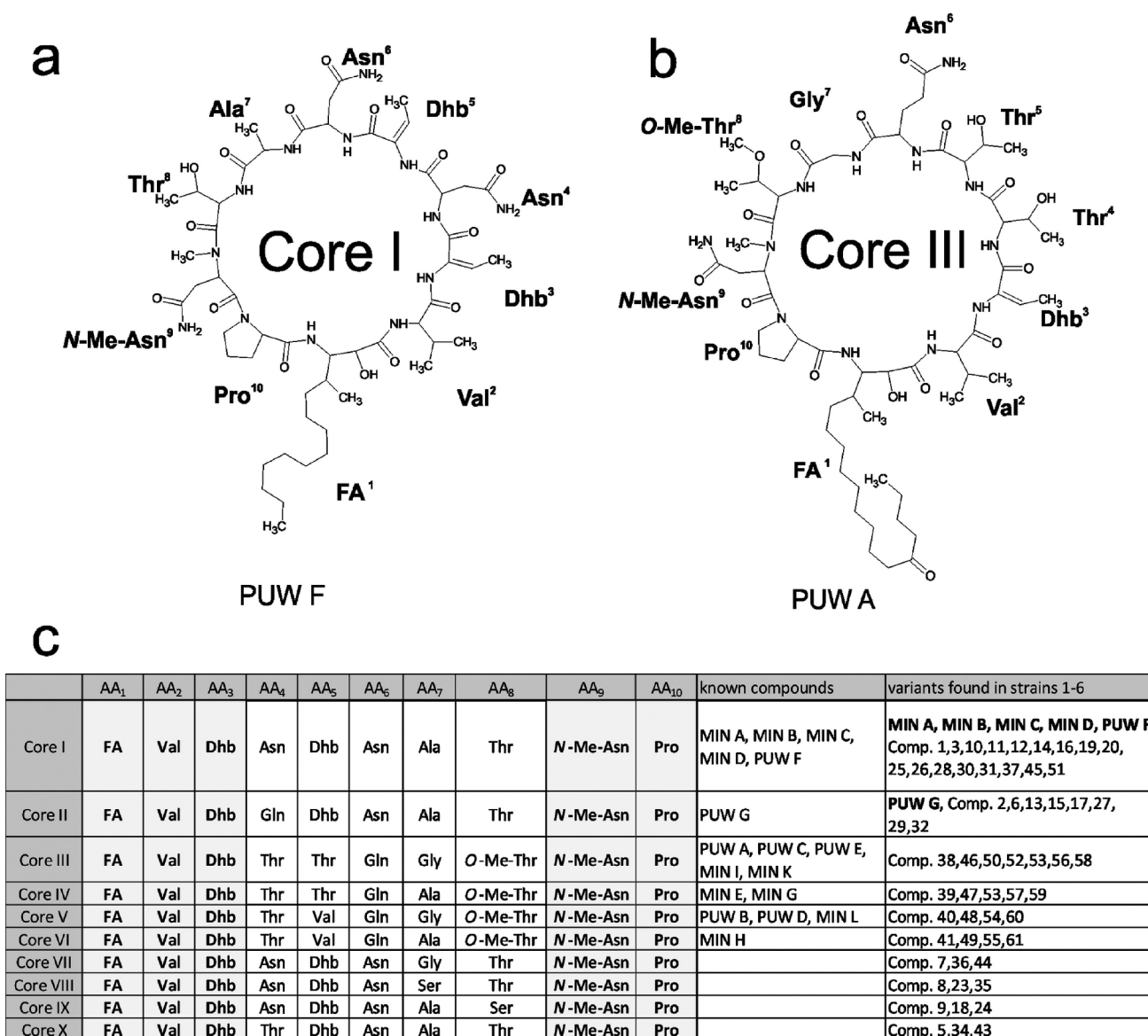


**FIG 2** Molecular network created using the Global Natural Products Social Molecular Networking (GNPS) web platform. Two separate networks were obtained during GNPS analysis: a group containing *Cylindrospermum* strains 1 to 3 and *Anabaena* strains 4 and 5 (a), and a group containing only variants detected in *Symplocastrum muelleri* strain 6 (b). The separate groups differ mainly in the peptide core of the molecule. For variants without complete structural information, only  $m/z$  values are shown. \*, compound present in trace amounts; #, compound for which MS-MS data failed to resolve the structural information.

mon. The activation of two divergent amino acids (Arg/Tyr) by a single adenylation domain, based on point mutations in just three codons, was previously demonstrated in the anabaenopeptin synthetase from the cyanobacterium *Planktothrix agardhii* (30). The substrate exchange of Ala versus Ser was previously reported from fungal class IV adenylation-forming reductases that contain A domains homologous to those of NRPS enzymes (31).

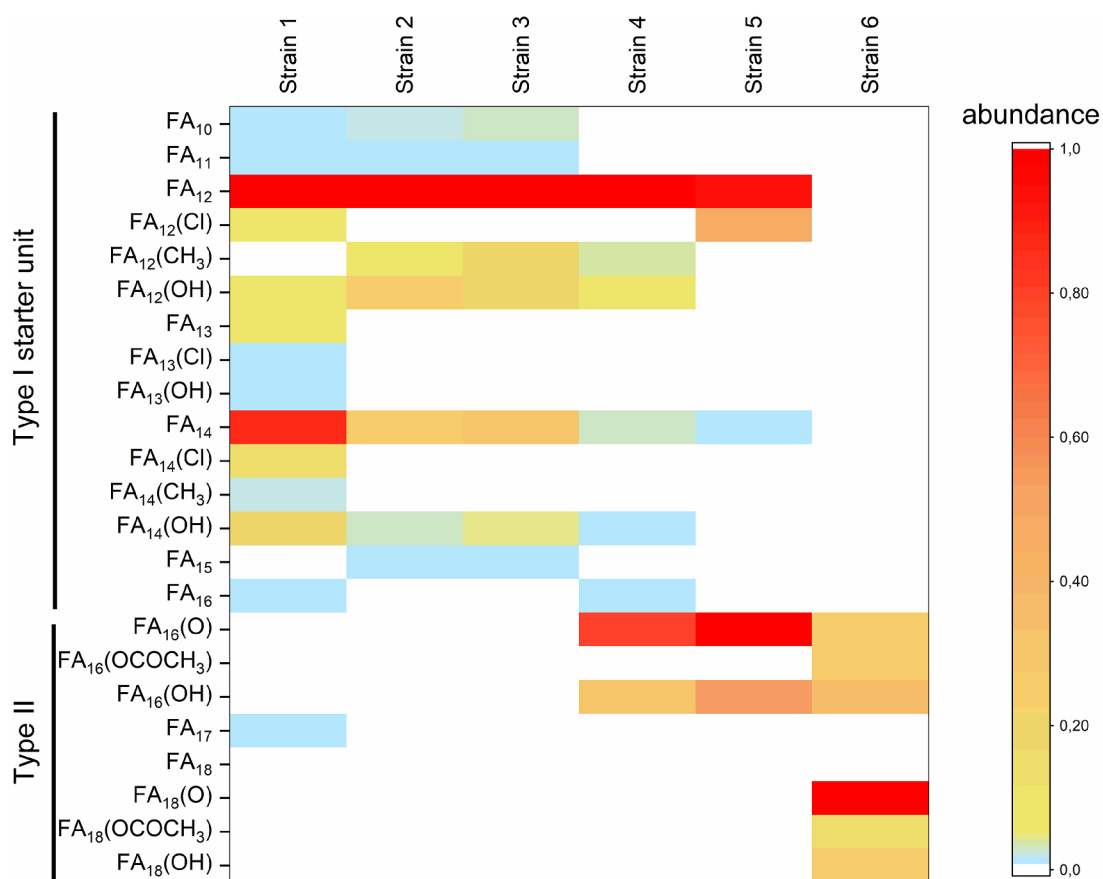
The last synthetase enzyme in the pathway (PuwA) is equipped with a terminal thioesterase domain (Fig. 6), which presumably catalyzes cleavage of the final product and formation of the cycle via a peptide bond between the terminal prolyl and the  $\beta$ -amino group of the FA chain, as previously suggested (12).

**Two hypothetical starter units and their substrate range.** The biosynthesis of bacterial lipopeptides is typically commenced by FA-activating enzymes (16, 18). Initiation of the biosynthesis of PUW/MIN is performed by a FAAL enzyme (12) and allows a much broader array of activated substrates than the relatively conserved oligopeptide core (Fig. 4) (13). We identified three alternative arrangements of the putative FAAL starter units (Fig. 5 and 6), each corresponding to a different array of FA side chains detected by HPLC–HRMS–MS, which presumably reflects the range of FA substrates activated during their biosynthesis (Fig. 4). *Cylindrospermum* species strains 1 to 3 possess the type I putative starter unit consisting of a standalone FAAL enzyme, PuwC, and a separate ACP, PuwD (Fig. 5, Table 2). In contrast, the biosynthetic gene cluster of *S. muelleri* strain 6 contains the type II putative starter unit (PuwI) consisting of a FAAL fused to an ACP (Fig. 5, Table 2). *Anabaena* species strains 4 and 5 combine both type I and type II putative starter units in their biosynthetic gene clusters (Fig. 5, Table 2). Although the functions and substrate ranges of these hypothetical starter units require further confirmation by gene manipulation experiments, they are supported by the



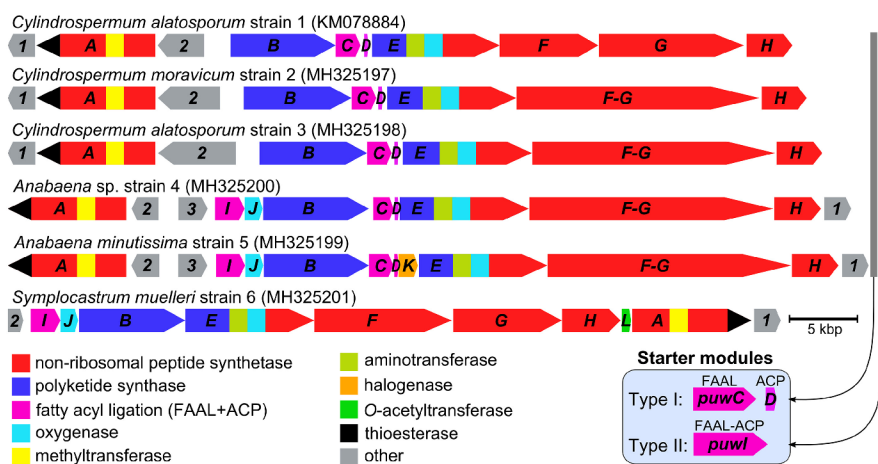
**FIG 3** Structural variability of the peptide cores of PUW/MIN variants. Examples of structural variants PUW F (a) and PUW A (b) with designated amino acid positions representing the two major peptide cores. (c) Table summarizing all types of the PUW/MIN peptide core found in known compounds reported in literature and compounds (Comp.) detected in studied strains. Columns shaded in gray highlight the conserved amino acid positions.

patterns of lipopeptide variants detected by HPLC–HRMS–MS (Fig. 4, Table S1). In *Cylindrospermum* strains 1 to 3, which exclusively contain the type I starter unit, the PUW/MIN products exhibited an almost-continuous FA distribution of between C<sub>10</sub> and C<sub>15</sub> (up to C<sub>17</sub> in negligible trace amounts) (Fig. 4). In *S. muelleri* strain 6, the presence of the type II loading module resulted in the production of PUW/MIN variants with discrete FA lengths of C<sub>16</sub> and C<sub>18</sub>. Strains containing both type I and type II starter units (*Anabaena* strains 4 and 5) produced two sets of PUW/MIN products with no overlap (C<sub>12</sub> to C<sub>14-15</sub> for the type I pathway and C<sub>16</sub> for the type II pathway) but exhibited a slightly shifted length distribution (Fig. 4). Based on these results, it seems plausible that PuwC–D and PuwI represent two alternative FAAL starter modules capable of initiating PUW/MIN biosynthesis (Fig. 5 and 6). An analogous situation was previously described for the alternative NRPS starter modules in the anabaenopeptin synthetase (32).



**FIG 4** Structural variability of the FA moiety of PUW/MIN variants. The relative proportions of variants with differences in FA lengths and substitutions are depicted using a color scale. For comparison, the peak area of a given variant was normalized against the peak area of the major variant present in the strain (MIN A for strains 1 to 5 and  $m/z$  1,235.7 for strain 6).

In the FA residue of the lipopeptide, proximal carbons in the linear aliphatic chain are incorporated into the nascent product by PKS enzymes (12). The PKS domains of PuwB and PuwE (Fig. 6, Table 2) catalyze two elongation steps. Therefore, the fatty acid is expected to be extended by four carbons. The substrate length specificities of the



**FIG 5** Structures of the *puw* gene cluster in the six cyanobacterial strains investigated. Gene arrangement and functional annotation of *puwA* to *-L* genes and selected PKS/NRPS tailoring domains are indicated by colored arrows. The distribution of the two types of putative starter modules (shaded boxes) observed is indicated by bars.

**TABLE 2** Deduced proteins encoded by the *puw* gene cluster in six cyanobacterial strains, including length and functional annotation<sup>a</sup>

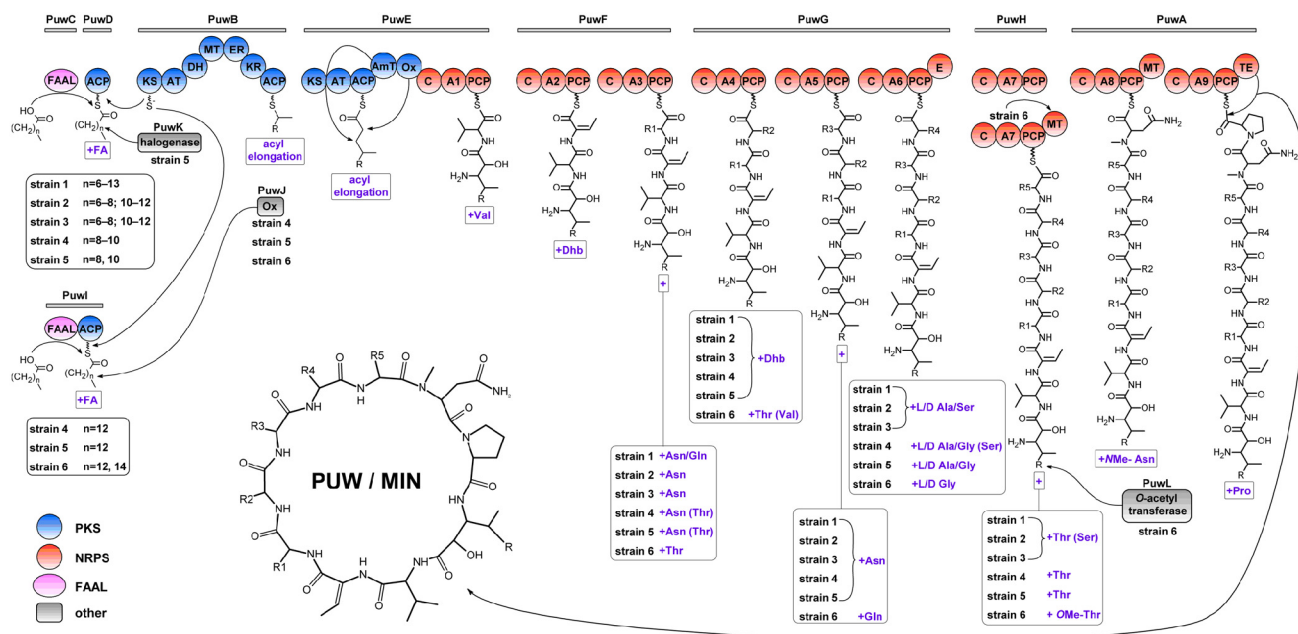
Protein	Length (aa) in strain no.:						Predicted function(s) <sup>a</sup>
	1	2	3	4	5	6	
ORF1	659	664	664	643	643	647	ABC transporter
PuwA	2,870	2,870	2,870	2,854	2,854	2,866	NRPS
ORF2	1,116	1,499	1,875	643	670	376	Patatin-like phospholipase
ORF3				696	696		Dynamin family protein
PuwI				709	702		FAAL, ACP
PuwJ				427	427	529	Cytochrome-like protein
PuwB	2,534	2,592	2,592	2,549	2,537	2,555	Hybrid PKS/NRPS, aminotransferase, oxygenase
PuwC	597	590	590	597	589		FAAL
PuwD	101	104	96	93	92		ACP
PuwK					465		Halogenase
PuwE	3,077	3,121	3,121	3,099	3,112	3,113	NRPS
PuwF	2,370	5,851 <sup>b</sup>	5,851 <sup>b</sup>	5,877 <sup>b</sup>	5,871 <sup>b</sup>	3,310	NRPS
PuwG	3,492					2,620	NRPS
PuwH	1,102	1,081	1,102	1,121	1,121	1,408	NRPS
PuwL						217	O-Acetyltransferase

<sup>a</sup>ACP, acyl carrier protein; FAAL, fatty acyl-AMP ligase; PKS, polyketide synthase; NRPS, nonribosomal peptide synthetase.

<sup>b</sup>The proteins PuwF and PuwG are encoded in a single ORF in this strain.

<sup>c</sup>The empty fields in the table indicate the absence of a particular protein.

FAAL enzymes in *Mycobacterium tuberculosis* were recently shown to be determined by the size and position of specific amino acid residues protruding into the FA-binding pocket (28). Experimental replacement of Gly or Ile by a larger Trp residue in the upper and middle parts of the pocket blocked the binding of the original C<sub>12</sub> substrate, but shorter chains (C<sub>2</sub> or C<sub>10</sub>, respectively) were still activated (28). Experimental data on FAAL substrate specificity in cyanobacteria are currently lacking. Alignment of amino acid residues from all putative PuwC and PuwI proteins demonstrates overall homology (Fig. S2a), including the positions corresponding to the FA-binding pocket, as previously shown in *Mycobacterium* (Fig. S2b) (28). Experimental evidence, such as *in vitro*



**FIG 6** Schematic view of the proposed biosynthesis assembly line of puwainaphycins and minutissamides. Variable amino acid positions and the ranges of fatty acyl lengths incorporated by the two putative alternative starter units are listed for individual strains. A, adenylation domain; ACP, acyl carrier protein; AmT, aminotransferase; AT, acyltransferase; C, condensation domain; DH, dehydratase; E, epimerase; ER, enoylreductase; FAAL, fatty acyl-AMP ligase; MT, methyltransferase; NRPS, nonribosomal peptide synthetase; KR, ketoreductase; KS, ketosynthetase; Ox, monooxygenase; PCP, peptidyl carrier protein; PKS, polyketide synthetase; TE, thioesterase.

activity assays and crystallization of protein-ligand complexes, is required to explain the varying substrate specificities of PuwC versus those of PuwL. Also, we cannot exclude the possibility that the FA substrate length range is partially determined by the pool of free FAs available to the FAAL enzyme. Indeed, this possibility is supported by observations of *Cylindrospermum* strains 1 to 3, which share highly conserved PuwC proteins (Fig. S2a) with identical residues in the predicted FA-binding pockets (Fig. S2b) but display slightly different ranges and ratios of incorporated FAs in the PUW/MIN variants produced (Fig. 1 and 4).

**FA tailoring reactions: oxidation, halogenation, and acetylation.** Intriguingly, all products originating from biosynthesis initiated by the type II starter unit (variants with C<sub>16</sub> and C<sub>18</sub> FA tails in *Anabaena* strains 4 and 5 and *S. muelleri* strain 6) include the substitution of a hydroxy or oxo moiety (Fig. 6). For minutissamides C and D, this substitution takes place on the third carbon from the FA terminus (C<sub>14</sub>), as described previously (10), and this position was confirmed by nuclear magnetic resonance (NMR) in variants produced by *Anabaena* sp. strain 4 in our study (Table S3, Fig. S3 to S6). In agreement with these hydroxy and oxo substitutions, the respective gene clusters each encode PuwJ, a putative cytochrome P450-like oxidase (Table 2), immediately downstream from the gene encoding the type II starter module. We therefore hypothesize that the PuwJ enzyme is responsible for hydroxylation of FA residues activated by PuwL (Fig. 6). However, the formation of the keto variant remains unexplained by our data.

Another gene, *puwK*, encoding a putative halogenase, was associated with the type II starter module in *Anabaena* sp. strain 5 (Table 2). Although no conserved enzymatic domain was detected in the deduced protein, it shares similarity with proteins postulated to be involved in the halogenation of cyanobacterial chlorinated acyl amides known as columbamides (33) and *N*-oxygenases similar to the *p*-aminobenzoate *N*-oxygenase AurF (34–36). The possible functional designation of this enzyme as a halogenase is further supported by the fact that the  $\omega$ -chlorinated product MIN B, originally described in *Anabaena* sp. strain 5 (10), was also detected in this study (Table S1) as one of the major variants, while no MIN B or any other chlorinated PUW/MIN products were detected in *Anabaena* sp. strain 4 (Fig. 1). *Anabaena* species strains 4 and 5 share identical organization across the entire gene cluster, and the lack of the putative halogenase gene *puwK* is the only difference between these two clusters in terms of the presence of genes (Fig. 5).

In *Cylindrospermum* species strains 1 to 3, which exclusively possess the type I starter unit, the presence of minor amounts of hydroxylated and chlorinated variants (Fig. 4) suggests the involvement of another biosynthetic mechanism unexplained by the current data. This ambiguity warrants experimental research, such as gene knockout experiments, to confirm the proposed functions of *puwJ* and *puwK*.

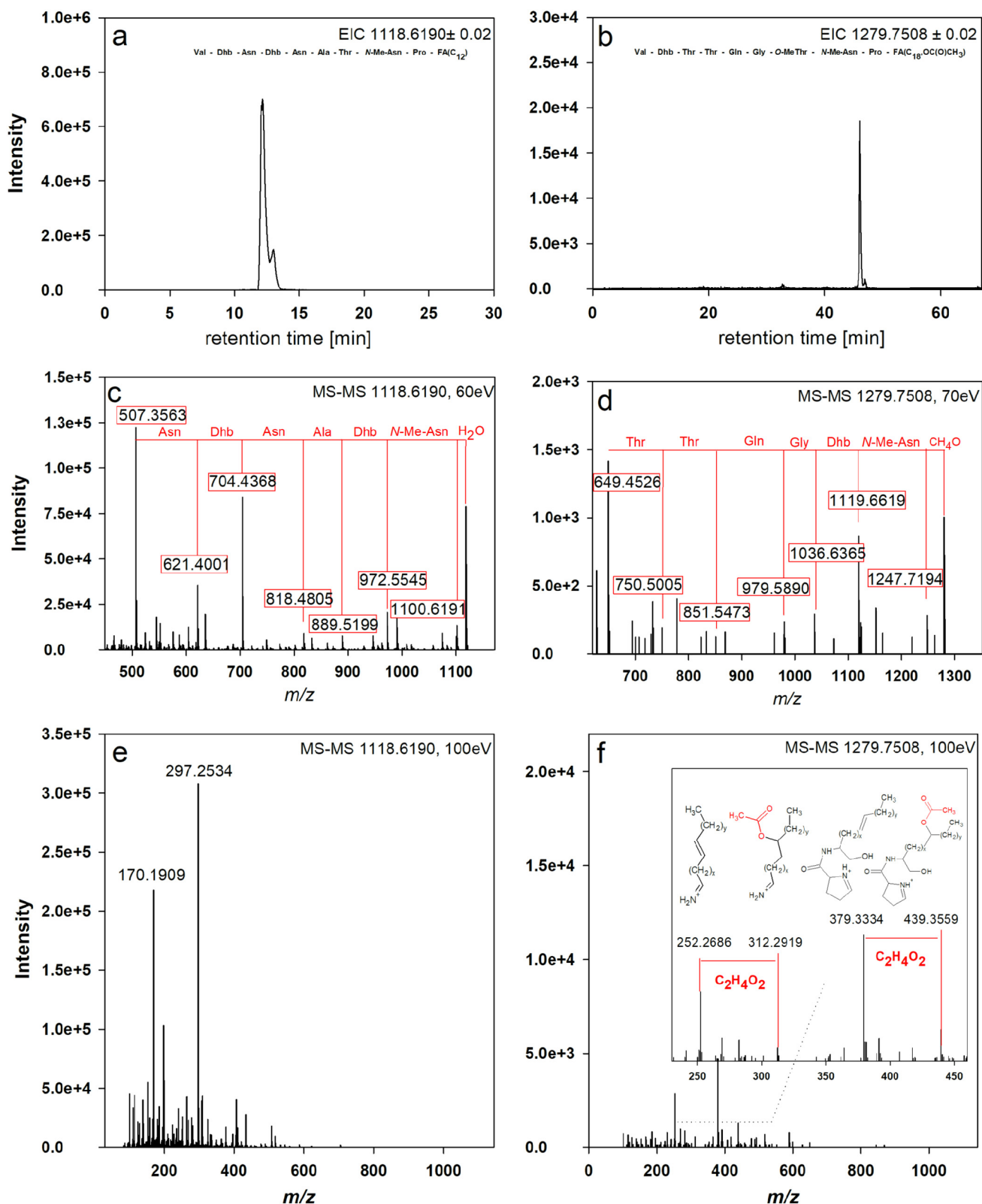
Finally, the gene cluster identified in *S. muelleri* strain 6 was the only one containing gene *puwL*. The deduced product of this gene shares 53.4% similarity with the *O*-acetyltransferase McyL (Table 2), which is involved in acetylation of the aliphatic chain of microcystin in cyanobacteria (37). Additionally, this gene is similar to those encoding chloramphenicol and streptogramin A *O*-acetyltransferases that serve as antibiotic resistance agents in various bacteria (38). The functional annotation of PuwL as a putative *O*-acetyltransferase is consistent with the detection of *O*-acetylated lipopeptide variants in *S. muelleri* strain 6 (Table 3, Fig. 7). Five PUW variants (*m/z* 1,265.7338, 1,279.7496, 1,277.7695, 1,291.7870, and 1,293.7654) that yield high-energy fragments proving the presence of an acetyl group bonded to the FA moiety have been detected. In the *m/z* 1,279.8 and 1,293.8 peaks, the high-energy-fragment ion at *m/z* 312 corresponds to the FA immonium ion bearing an acetyl group, and the fragment ion at *m/z* 439 corresponds to the prolyl-FA-acetyl fragment. The subsequent loss of an acetyl group resulted in the presence of ions at *m/z* 252 and 379, respectively (Table 3, Fig. 7). Similarly, analysis of the *m/z* 1,265.7 peak revealed analogous fragments at *m/z* 284/411 and 351/224 (Table 3).

**TABLE 3** Fragmentation and amino acid composition of PUW variants from *Symplocastrum muelleri* strain 6 bearing acetyl substitutions on the FA moiety<sup>a</sup>

Level of fragmentation energy, peptide sequence or fragment obtained	Value when X, Y, and FA are as indicated														
	X, Ala; Y, Thr; FA, C <sub>16</sub>			X, Gly; Y, Thr; FA, C <sub>18</sub>			X, Ala; Y, Thr; FA, C <sub>18</sub>			X, Gly; Y, Val; FA, C <sub>18</sub>			X, Ala; Y, Val; FA, C <sub>18</sub>		
	m/z	Δ(ppm)	Sum formula	m/z	Δ(ppm)	Sum formula	m/z	Δ(ppm)	Sum formula	m/z	Δ(ppm)	Sum formula	m/z	Δ(ppm)	Sum formula
Low fragmentation energy (60 eV)															
[M+H] <sup>+</sup>	1,265.7338	+0.7	C <sub>59</sub> H <sub>101</sub> N <sub>12</sub> O <sub>18</sub>	1,279.7496	+0.9	C <sub>60</sub> H <sub>103</sub> N <sub>12</sub> O <sub>18</sub>	1,293.7654	+0.8	C <sub>61</sub> H <sub>105</sub> N <sub>12</sub> O <sub>18</sub>	1,277.7695	+1.6	C <sub>61</sub> H <sub>105</sub> N <sub>12</sub> O <sub>17</sub>	1,291.7870	+0.1	C <sub>62</sub> H <sub>107</sub> N <sub>12</sub> O <sub>17</sub>
[M-CH <sub>3</sub> OH] <sup>+</sup>	1,233.7170	-6.6	C <sub>58</sub> H <sub>97</sub> N <sub>12</sub> O <sub>17</sub>	1,247.7194	+4.1	C <sub>59</sub> H <sub>99</sub> N <sub>12</sub> O <sub>17</sub>	1,261.7494	-7.3	C <sub>60</sub> H <sub>101</sub> N <sub>12</sub> O <sub>17</sub>	Low int.		C <sub>60</sub> H <sub>101</sub> N <sub>12</sub> O <sub>16</sub>	Low int.		C <sub>61</sub> H <sub>103</sub> N <sub>12</sub> O <sub>16</sub>
[M-CH <sub>3</sub> OH-NMeAsn] <sup>+</sup>	1,105.6558	-4.9	C <sub>53</sub> H <sub>89</sub> N <sub>10</sub> O <sub>15</sub>	1,119.6619	+3.7	C <sub>54</sub> H <sub>91</sub> N <sub>10</sub> O <sub>15</sub>	1,133.681	+0.6	C <sub>55</sub> H <sub>93</sub> N <sub>10</sub> O <sub>15</sub>	1,117.6924	-5.0	C <sub>55</sub> H <sub>93</sub> N <sub>10</sub> O <sub>14</sub>	1,131.7307	-25.0	C <sub>56</sub> H <sub>95</sub> N <sub>10</sub> O <sub>14</sub>
[M-CH <sub>3</sub> OH-NMeAsn-Dhb] <sup>+</sup>	1,022.6180	-4.7	C <sub>49</sub> H <sub>84</sub> N <sub>9</sub> O <sub>14</sub>	1,036.6365	-7.3	C <sub>50</sub> H <sub>86</sub> N <sub>9</sub> O <sub>14</sub>	1,050.6478	-3.1	C <sub>51</sub> H <sub>88</sub> N <sub>9</sub> O <sub>14</sub>	1,134.6603	-10.3	C <sub>51</sub> H <sub>88</sub> N <sub>9</sub> O <sub>13</sub>	1,048.6671	-1.7	C <sub>52</sub> H <sub>90</sub> N <sub>9</sub> O <sub>13</sub>
[M-CH <sub>3</sub> OH-NMeAsn-Dhb-X] <sup>+</sup>	951.5785	-2.5	C <sub>48</sub> H <sub>83</sub> N <sub>8</sub> O <sub>13</sub>	979.589	+18.8	C <sub>48</sub> H <sub>83</sub> N <sub>8</sub> O <sub>13</sub>	979.6041	+3.4	C <sub>48</sub> H <sub>83</sub> N <sub>8</sub> O <sub>13</sub>	977.6481	-20.5	C <sub>49</sub> H <sub>85</sub> N <sub>8</sub> O <sub>12</sub>	977.6518	-24.1	C <sub>49</sub> H <sub>85</sub> N <sub>8</sub> O <sub>12</sub>
[M-CH <sub>3</sub> OH-NMeAsn-Dhb-X-Gln] <sup>+</sup>	823.5253	-9.4	C <sub>41</sub> H <sub>71</sub> N <sub>6</sub> O <sub>11</sub>	851.5473	+1.8	C <sub>43</sub> H <sub>75</sub> N <sub>6</sub> O <sub>11</sub>	851.5478	+1.2	C <sub>43</sub> H <sub>75</sub> N <sub>6</sub> O <sub>11</sub>	849.5838	-16.7	C <sub>44</sub> H <sub>76</sub> N <sub>6</sub> O <sub>10</sub>	849.5589	+12.5	C <sub>44</sub> H <sub>77</sub> N <sub>6</sub> O <sub>10</sub>
[M-CH <sub>3</sub> OH-NMeAsn-Dhb-X-Gln-Y] <sup>+</sup>	722.4729	-4.2	C <sub>37</sub> H <sub>64</sub> N <sub>5</sub> O <sub>9</sub>	750.5005	+0.9	C <sub>39</sub> H <sub>68</sub> N <sub>5</sub> O <sub>9</sub>	750.5147	-18.1	C <sub>39</sub> H <sub>68</sub> N <sub>5</sub> O <sub>9</sub>	Low int.		C <sub>40</sub> H <sub>72</sub> N <sub>5</sub> O <sub>8</sub>	Low int.		C <sub>40</sub> H <sub>72</sub> N <sub>5</sub> O <sub>8</sub>
[M-CH <sub>3</sub> OH-NMeAsn-Dhb-X-Gln-Y-Thr] <sup>+</sup>	621.4223	-0.2	C <sub>33</sub> H <sub>57</sub> N <sub>4</sub> O <sub>7</sub>	649.4526	+1.4	C <sub>35</sub> H <sub>61</sub> N <sub>4</sub> O <sub>7</sub>	649.4539	-0.6	C <sub>35</sub> H <sub>61</sub> N <sub>4</sub> O <sub>7</sub>	649.465	-17.8	C <sub>35</sub> H <sub>61</sub> N <sub>4</sub> O <sub>7</sub>	649.4483	+8.0	C <sub>35</sub> H <sub>61</sub> N <sub>4</sub> O <sub>7</sub>
High fragmentation energy (100 eV)															
Fragment 1	411.3208	+2.2	C <sub>23</sub> H <sub>43</sub> N <sub>2</sub> O <sub>4</sub>	439.3559	-6.5	C <sub>25</sub> H <sub>47</sub> N <sub>2</sub> O <sub>4</sub>	439.3556	-5.8	C <sub>25</sub> H <sub>47</sub> N <sub>2</sub> O <sub>4</sub>	439.3556	-5.8	C <sub>25</sub> H <sub>47</sub> N <sub>2</sub> O <sub>4</sub>	439.3508	+5.1	C <sub>25</sub> H <sub>47</sub> N <sub>2</sub> O <sub>4</sub>
Fragment 1, C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	351.3006	+0.0	C <sub>21</sub> H <sub>39</sub> N <sub>2</sub> O <sub>2</sub>	379.3334	-4.0	C <sub>23</sub> H <sub>43</sub> N <sub>2</sub> O <sub>2</sub>	379.3329	-2.6	C <sub>23</sub> H <sub>43</sub> N <sub>2</sub> O <sub>2</sub>	379.3328	-2.4	C <sub>23</sub> H <sub>43</sub> N <sub>2</sub> O <sub>2</sub>	379.3360	-10.8	C <sub>23</sub> H <sub>43</sub> N <sub>2</sub> O <sub>2</sub>
Fragment 2	284.2583	+0.4	C <sub>17</sub> H <sub>34</sub> NO <sub>2</sub>	312.2919	-6.9	C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>	312.2892	+1.6	C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>	Low int.		C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>	Low int.		C <sub>19</sub> H <sub>38</sub> NO <sub>2</sub>
Fragment 2, C <sub>2</sub> H <sub>4</sub> O <sub>2</sub>	224.2367	+2.6	C <sub>15</sub> H <sub>30</sub> N	252.2686	0.0	C <sub>17</sub> H <sub>34</sub> N	252.2687	-0.5	C <sub>17</sub> H <sub>34</sub> N	252.2684	+0.7	C <sub>17</sub> H <sub>34</sub> N	252.2677	+3.5	C <sub>17</sub> H <sub>34</sub> N

<sup>a</sup>Low int., low intensity (defined here as a signal-to-noise ratio lower than 2). For detailed methods, see Materials and Methods.





**FIG 7** MS-MS fragmentation of MIN A (a, c, e) and the PUW variant at  $m/z$  1,279 bearing an acetyl substitution of the fatty acid chain (b, d, f). (a, b) Base peak chromatograms. (c, d) Fragmentation of the protonated molecule at low fragmentation energy, yielding  $b$  series of ions, corresponding to the losses of particular amino acid residues. (e, f) Fragmentation of the protonated molecule at high energy (100 eV), yielding fragments characteristic for the  $\beta$ -amino fatty acid.

**TABLE 4** Bacterial and yeast strains used for antimicrobial testing of PUW F and MIN A, C, and D

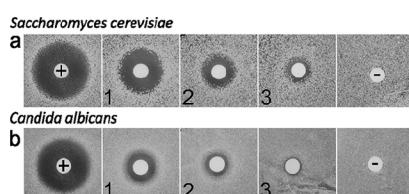
Test organism (HAMBI no.) <sup>a</sup>	Medium <sup>b</sup>	Incubation temp (°C)	Incubation time (h)	Gram stain reaction <sup>c</sup>
<i>Pseudomonas</i> sp. (2796)	TGY	28	24	–
<i>Micrococcus luteus</i> (2688)	TGY	28	24	+
<i>Bacillus subtilis</i> (251)	TGY	28	24	+
<i>Pseudomonas aeruginosa</i> (25)	TGY	37	24	–
<i>Escherichia coli</i> (396)	TGY	37	24	–
<i>Bacillus cereus</i> (1881)	TSA	28	24	+
<i>Burkholderia cepacia</i> (2487)	TSA	37	24	–
<i>Staphylococcus aureus</i> (11)	TSA	37	24	+
<i>Xanthomonas campestris</i> (104)	NA	28	24	–
<i>Burkholderia pseudomallei</i> (33)	NA	37	24	–
<i>Salmonella enterica</i> serovar Typhi (1306)	NA	37	24	–
<i>Arthrobacter globiformis</i> (1863)	NA	28	24	–
<i>Kocuria varians</i> (40)	NA	28	24	+
<i>Candida albicans</i> (261)	YM agar	37	24	Yeast
<i>Cryptococcus albidus</i> (264)	YM agar	28	24	Yeast
<i>Saccharomyces cerevisiae</i> (1164)	YM agar	28	24	Yeast

<sup>a</sup>HAMBI, culture collection of University of Helsinki, Faculty of Agriculture and Forestry, Department of Microbiology.

<sup>b</sup>The compositions of all media were obtained from the American Type Culture Collection (ATCC). TGY, tryptone glucose yeast; TSA, tryptic soy agar; NA, nutrient agar; YM agar, yeast malt agar.

<sup>c</sup>–, negative; +, positive.

**Antimicrobial activity.** Both PUWs and MINs possess cytotoxic activity against human cells *in vitro* (5, 10, 11). In the current study, we demonstrated that the major PUW/MIN variants (PUW F and MIN A, C, and D) did not exert antibacterial effects against either Gram-positive or Gram-negative bacteria in a panel of 13 selected strains (Table 4). PUW F was the only variant tested that manifested antagonistic activity against two yeast strains utilized in our experiment, namely, *Candida albicans* strain HAMBI 261 and *Saccharomyces cerevisiae* strain HAMBI 1164, with inhibition zones of 14 and 18 mm, respectively, and MICs of 6.3  $\mu\text{g ml}^{-1}$  (5.5  $\mu\text{M}$ ) (Fig. 8). No antifungal activity was recorded for the MIN C and D variants, and only weak inhibition of the two yeast strains was recorded for MIN A (Fig. S7). PUW F differs only slightly from MIN A, by a  $-\text{CH}_2-\text{CH}_3$  extension of the FA moiety, indicating that the FA length affects bioactivity. Furthermore, the lack of bioactivity for MIN C and MIN D suggests that hydroxy and oxo substitutions also compromise antifungal efficacy. As previously demonstrated, cytotoxicity is due to membrane permeabilization activity accompanied by calcium flux into the cytoplasm (5), consistent with the membrane effects documented for other bacterial lipopeptides (4). However, as is apparent from our data (Fig. 8), PUW/MIN products appear to be effective solely against eukaryotes (thus far tested only on human and yeast cells). This finding is in contrast to the typical antibacterial activity frequently described for many lipopeptides produced by Gram-positive bacteria (4). Analogously, the cyanobacterial lipopeptides anabaenolysin A and hassalidins



**FIG 8** Antifungal activities of PUW F against yeast strains *Saccharomyces cerevisiae* HAMBI 1164 (a) and *Candida albicans* HAMBI 261 (b). Discs were treated with a range of concentrations from 25.2  $\mu\text{g ml}^{-1}$  to 0.0394  $\mu\text{g/ml}$  to determine the MIC. Numbers and symbols represent concentrations and controls, as follows: 1, 25.2  $\mu\text{g ml}^{-1}$ ; 2, 12.6  $\mu\text{g ml}^{-1}$ ; 3, 6.3  $\mu\text{g ml}^{-1}$ ; +, positive control (10  $\mu\text{g}$  nystatin); –, negative control (10  $\mu\text{l}$  methanol).

preferentially interact with cholesterol-containing membranes, hence their predisposition for activity against eukaryotic cells (6, 8).

**Distribution of PUWs and MINs in cyanobacteria.** PUWs and MINs form one of the most frequently reported groups of lipopeptides in cyanobacteria and have been isolated from heterocytous cyanobacteria, particularly members of the genera *Anabaena* and *Cylindrospermum* that inhabit soil (5, 9–11). Only a single study has mentioned the probable occurrence of puwainaphycins in a planktonic cyanobacterium (*Sphaerospermopsis*) (39). Our current comprehensive analysis of these lipopeptides and their biosynthetic genes further supports the hypothesis that lipopeptides occur predominantly in nonplanktonic biofilm-forming cyanobacteria (23). In this context, it is worth mentioning that *S. muelleri* strain 6 was isolated from a wetland bog in alpine mountains in coastal Norway (40). This strain is a toxigenic member of a biofilm microbiome and suspected to play a role in the development of severe hemolytic Alveld disease among outfield grazing sheep (41, 42). Biomass harvested from pure cultures of this strain exhibited strong cytotoxic activity toward primary rat hepatocytes (43, 44), which indicates the production of secondary metabolites with cytotoxic properties. Thus, the possible toxic potential of cyanobacterial lipopeptides, such as PUWs and MINs, in the environment warrants further attention.

**Conclusions.** Our study highlights and explores the extensive structural versatility of cyanobacterial lipopeptides from the PUW/MIN family by introducing previously unknown variants and newly sequenced biosynthetic gene clusters. Intriguingly, all variants are synthesized by a relatively conserved PKS/NRPS machinery with a common genetic origin. We hypothesize that chemical diversity is generated largely by the presence of two alternative fatty acyl-AMP ligase starter units, one of which exhibits an unusually broad specificity for FA substrates of various lengths. Additionally, putative halogenase and *O*-acetyltransferase genes were present in some gene clusters. This knowledge provides novel insight into the genetic background underpinning the biosynthesis of bacterial lipopeptides. The proposed biosynthetic mechanisms allow the studied microbes to generate a large pool of products that can be readily expanded by introducing relatively small genetic changes. This is consistent with the so-called “screening hypothesis” (45, 46), which predicts an evolutionary benefit for organisms producing a large chemical diversity of secondary metabolites at minimum cost. Accessory antimicrobial tests on bacteria and yeasts, together with previously published results, suggest a specific toxic effect of PUWs against eukaryotic cells. Thus, their toxic potential for humans and other animals clearly warrants further investigation, and their possible use as antifungal agents is ripe for exploration.

## MATERIALS AND METHODS

**Cultivation of cyanobacterial strains.** Six cyanobacterial strains were included in the present study, *Cylindrospermum moravicum* CCALA 993 (strain 1), *Cylindrospermum alatosporum* CCALA 994 (strain 2), *Cylindrospermum alatosporum* CCALA 988 (strain 3), *Anabaena* sp. strain UHCC-0399 (previously *Anabaena* sp. strain SMIX 1; strain 4), *Anabaena minutissima* UTEX B 1613 (strain 5), and *Symplocastrum muelleri* NIVA-CYA 644 (strain 6). The origins of the strains are listed in Table 1. For chemical analysis, strains 1 to 5 were cultivated in BG-11 medium (47) in glass columns (300 ml) bubbled with air enriched in 1.5% CO<sub>2</sub> at a temperature of 28°C and constant illumination of 100 μmol photons m<sup>-2</sup> s<sup>-1</sup>. Strain 6 was maintained in culture using a custom liquid medium obtained by mixing 200 ml of Z8 medium (48), 800 ml distilled water, 30 ml soil extract, and common vitamin premix (according to SAG [Sammlung von Algenkulturen der Universität Göttingen], but without biotin). Cultivation was performed in 100- to 200-ml Erlenmeyer flasks at 20°C with a 16:8 light/dark photoperiod under static conditions. Cultures were kept at low irradiance (4 μmol m<sup>-2</sup> s<sup>-1</sup> of photosynthetically active radiation [PhAR], generated using red-green-blue light-emitting-diode [RGB LED] strips). Cells were harvested by centrifugation (3,125 × *g*), stored at –80°C, and subsequently lyophilized. Strain 4 was cultivated at a larger scale for purification of major lipopeptide variants in a 10-liter tubular photobioreactor under the above-mentioned conditions in BG-11 medium.

**Molecular and bioinformatic analyses.** Single filaments of strains 2, 3, 5, and 6 were isolated for whole-genome amplification (WGA) and subsequent preparation of a whole-genome sequencing (WGS) library as described previously (12). Briefly, the glass capillary technique was used to isolate filaments and exclude minor bacterial contaminants. A set of 20 filaments from each strain was then used as a template for WGA. Multiple displacement amplification (MDA) using a Repli-g minikit (Qiagen, Hilden, Germany) was followed by PCR and sequencing to monitor the cyanobacterial 16S rRNA gene using primers 16S387F and 16S1494R (49). Positive samples (7 to 10 MDA products yielding clear 16S rRNA gene

sequences of the respective genera) were then pooled to create a template for WGS. DNA samples were sent for commercial *de novo* genome sequencing (EMBL Genomics Core Facility, Heidelberg, Germany) using the Illumina MiSeq platform (Illumina, San Diego, CA, USA) with an ~350-bp average-insert-length paired-end library and 250-bp reads (~1.4-Gbp data yield per strain). Raw data from *de novo* WGS were assembled using CLC Bio Genomics Workbench version 7.5 (CLC Bio, Aarhus, Denmark). Genomic DNA was isolated from strain 4 as previously described (37), and the quality was assessed using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and an Agilent TapeStation (Agilent Technologies, Santa Clara, CA, USA). High-molecular-weight DNA was used to construct an Illumina TruSeq PCR-free 350-bp library and sequenced using an Illumina HiSeq 2500 platform with a paired-end 100-cycle run. Genome data (1 Gb for each strain) were first checked using SPAdes version 3.7.1 (50) for read correction and removal of erroneous reads and then assembled using Newbler version 3.0 (454 Life Sciences, Branford, CT, USA). Genomic scaffolds were loaded into Geneious pro R10 (Biomatters) and investigated for FAAL and NRPS genes using BLASTp searches to identify putative lipopeptide synthetase gene clusters (23). FAAL and NRPS adenylation domains (A domains) from the single known PUW gene cluster (strain 1; KM078884) were used as queries since homologous gene clusters were expected. Contigs yielding high-similarity hits (E value of  $<10^{-20}$ ) were then analyzed using the Glimmer 3 (51) algorithm to discover putative open-reading frames (ORFs). Functional annotation of ORFs was conducted by applying a combination of BLASTp/CDD searches against the NCBI database and using the antiSMASH 4.0 secondary metabolite gene cluster annotation pipeline (52, 53). Pairwise sequence identities and the presence of conserved residues in homologous putative proteins encoded in the gene clusters were assessed using Geneious pro software based on amino acid alignment (MAFFT plugin, default parameters). Minor assembly gaps were identified in the genomic scaffolds of all strains investigated, either directly after paired-end read assembly or based on mapping to the reference gene cluster from *C. alatosporum* CICALA 988. Gaps in PUW/MIN gene clusters were closed by PCR, and subsequent Sanger sequencing of PCR products was performed using custom primer annealing to regions adjacent to the assembly gaps.

**Extraction and analysis of PUWs/MINs.** To obtain comparable results, each strain was extracted using an identical ratio of lyophilized biomass (200 mg) to extraction solvent (10 ml of 70% methanol [MeOH], vol/vol). Extracts were evaporated using a rotary vacuum evaporator at 35°C and concentrated to 1 ml of 70% MeOH. The methanolic extracts were analyzed using a Thermo Scientific Dionex UltiMate 3000 ultra-high-performance liquid chromatography plus (UHPLC+) instrument equipped with a diode array detector connected to a Bruker Impact HD (Bruker, Billerica, MA, USA) high-resolution mass spectrometer with electrospray ionization. Separation of extracts was performed on a reversed-phase Phenomenex Kinetex C<sub>18</sub> column (150 by 4.6 mm, 2.6 μm) using H<sub>2</sub>O (A)-acetonitrile (B) containing 0.1% HCOOH as a mobile phase, at a flow rate of 0.6 ml min<sup>-1</sup>. The gradient was as follows: A/B 85/15 (0 min), 85/15 (over 1 min), 0/100 (over 20 min), 0/100 (over 25 min), and 85/15 (over 30 min). For better resolution of minor PUW variants, another analytical method with a longer gradient (67 min) adopted from our previous study (13) was applied. The peptide sequence was reconstructed from the *b* ion series obtained after opening of the ring between the proline and *N*-methylasparagine residues, followed by the sequential loss of water and all the amino acids with exception of the last residue (Pro). The numbers of carbons in the FA moiety in PUW/MIN variants containing nonsubstituted and hydroxy-/chloro-substituted FA were determined using a method described previously by our team (13). Characteristic FA immonium fragments in oxo-substituted PUW/MIN variants were identified by applying this method to crude extracts of *Anabaena* strain 5 containing the oxo-substituted MIN D variant (10). Since a stable, prominent, and characteristic FA immonium fragment with the sum formula C<sub>15</sub>H<sub>30</sub>NO<sup>+</sup> was obtained for MIN D (Fig. S1), analogous fragments with the general formula C<sub>x</sub>H<sub>2x</sub>NO<sup>+</sup> were used to identify oxo-substituted components in unknown PUW/MIN variants from other strains investigated.

**Molecular networking.** A molecular network was created using the Global Natural Products Social Molecular Networking (GNPS) online workflow (54). Data were filtered by removing all MS-MS peaks within ±17 Da of the precursor *m/z*. MS-MS spectra were window filtered by choosing only the top six peaks in the ±50-Da window throughout the spectrum. Data were then clustered with MS-Cluster with a parent mass tolerance of 0.1 Da and an MS-MS fragment ion tolerance of 0.025 Da to create consensus spectra. Additionally, consensus spectra comprised of fewer than two spectra were discarded. A network was then created in which edges were filtered using a cosine score above 0.75 and more than three matched peaks. Additional edges between pairs of nodes were retained in the network only when both nodes were included in each other's respective top 10 most similar nodes. Spectra in the network were then searched against the GNPS spectral libraries, and library spectra were filtered in the same manner as the input data. All matches obtained between network spectra and library spectra were retained only when the score was above 0.7 and at least four peaks matched. Analog searching was performed against the library with a maximum mass shift of 200 Da.

**Purification of MINs from *Anabaena* sp. strain 4 and its NMR analysis.** Freeze-dried biomass of strain 4 (10 g) was extracted with 70% MeOH (500 ml). The extract was evaporated using a rotary vacuum evaporator to reduce the MeOH content, and the sample was subsequently diluted with distilled water to reach a final MeOH concentration of >5%. The diluted extract was prepurified using a Supelco C<sub>18</sub> SPE cartridge (10 g, 60 ml) preequilibrated with 60 ml of MeOH and 120 ml of H<sub>2</sub>O. After loading, retained components were eluted with 60 ml of pure MeOH, concentrated to dryness, and resuspended in 10 ml of pure MeOH. MIN A, C, and D were purified in two HPLC purification steps. The first step was performed on a preparative chromatographic system (Agilent 1260 Infinity series) equipped with a multiwavelength detector and automatic fraction collector. A preparative Reprosil 100 C<sub>18</sub> column (252 by 25 mm) was employed for separation at a flow rate of 10 ml min<sup>-1</sup> using the following gradient

of MeOH containing 0.1% HCOOH (A) and 10% MeOH containing 0.1% HCOOH (B): 0 min (100% B), 6 min (100% B), 15 min (43% B), 43 min (12% B), 45 min (0% B), 58 min (0% B), 60 min (100% B), and 64 min (100% B). Fractions were collected using an automatic fraction collector at 1-min intervals, and fractions were analyzed for MIN A, C, and D using the method described above. Fractions containing MIN A, C, and D were collected in separate vials and concentrated using a rotary evaporator. The second purification step was performed on a semipreparative HPLC (Agilent 1100 Infinity series) using a Reprosil 100 Phenyl column (250 by 8 mm) with acetonitrile containing 0.1% HCOOH (A) and water containing 0.1% HCOOH (B) using the following gradient: 0 min (60% B), 2 min (60% B), 6 min (50% B), 28 min (18% B), 30 min (0% B), 30 min (0% B), 32 min (0% B), 31 min (60% B), and 36 min (60% B). The flow rate was 1 ml min<sup>-1</sup> throughout, fractions were collected manually, and the purity was analyzed using the HPLC-HRMS method described above. NMR spectra of minutissamides were measured in dimethyl sulfoxide (DMSO)-*d*<sub>6</sub> at 30°C. All NMR spectra were collected using a Bruker Avance III 500 MHz NMR spectrometer, equipped with a 5-mm-diameter broadband probe (BBI) head with actively shielded z gradient.

**Antibacterial and antifungal assays.** The antimicrobial activities of four major variants (PUW F and MIN A, C, and D) were tested against 13 bacterial and two yeast strains (Table 4) using disc diffusion assays (8) in three independent experiments with kanamycin/nystatin and MeOH as positive and negative controls, respectively. The antifungal activity of PUW F was further evaluated by determining the MICs against *Candida albicans* (HAMBI 261) and *Saccharomyces cerevisiae* (HAMBI 1164) as described previously (8). PUW F was isolated from *Cylindrospermum* strain 1 according to a protocol described previously (5), and isolation of MIN A, C, and D was performed as described above. The variants produced by *S. muelleri* strain 6 were impossible to isolate due to the slow growth of the cyanobacterium, which resulted in low biomass yields during the study period.

**Accession number(s).** The accession numbers for the newly sequenced complete putative biosynthetic gene clusters uploaded to the European Nucleotide Archive are [MH325197](https://doi.org/10.1128/M325197) to [MH325201](https://doi.org/10.1128/M325201).

## SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/AEM.02675-18>.

**SUPPLEMENTAL FILE 1**, PDF file, 4.8 MB.

## ACKNOWLEDGMENTS

This work was supported by Czech Science Foundation grant no. 16-09381S (Bioactive cyanobacterial lipopeptides: genome mining, detection, and structure-activity relationships), and by the Ministry of Education, Youth and Sports of the Czech Republic, National Program of Sustainability I, ID: LO1416 project ALGAMIC (ID: CZ.1.05/2.1.00/19.0392) and MSCA IF II project (CZ.02.2.69/0.0/0.0/18\_070/0010493) as well as by Ministry of Regional Development of the Czech Republic—Cross-Border cooperation Czech-Bavaria project no. 41. Student participation (A. Kust and T. Galica) in the project was supported by The Grant Agency of the Faculty of Science, University of South Bohemia, grant GAJU 158/2016/P. Access to instruments and other facilities was supported by the Czech Research Infrastructure for Systems Biology (C4SYS; project no. LM2015055). This research was also supported by a grant from the NordForsk NCoE program NordAqua (project no. 82845) and by the Jane and Aatos Erkko Foundation (K. Sivonen and J. Jokela), which mostly supported the structural elucidation of PUW variants. The Norwegian participation was supported by grants from the Department of Agriculture and Forestry, the County Governor of Møre og Romsdal, the County Governor of Sogn og Fjordane, the University of Oslo, and the Norwegian Institute for Water Research.

The authors declare no conflict of interest.

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

1. Cochrane SA, Vederas JC. 2016. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: a gold mine of antibiotic candidates. *Med Res Rev* 36:4–31. <https://doi.org/10.1002/med.21321>.
2. Taylor SD, Palmer M. 2016. The action mechanism of daptomycin. *Bioorg Med Chem* 24:6253–6268. <https://doi.org/10.1016/j.bmc.2016.05.052>.
3. Velkov T, Roberts KD, Li J. 2017. Rediscovering the octapeptins. *Nat Prod Rep* 34:295–309. <https://doi.org/10.1039/c6np00113k>.
4. Ines M, Dhouha G. 2015. Lipopeptide surfactants: production, recovery and pore forming capacity. *Peptides* 71:100–112. <https://doi.org/10.1016/j.peptides.2015.07.006>.
5. Hrouzek P, Kuzma M, Černý J, Novák P, Fišer R, Šimek P, Lukešová A, Kopecký J. 2012. The cyanobacterial cyclic lipopeptides puwainaphycins F/G are inducing necrosis via cell membrane permeabilization and subsequent unusual actin relocalization. *Chem Res Toxicol* 25:1203–1211. <https://doi.org/10.1021/tx300044t>.
6. Oftedal L, Myhren L, Jokela J, Gausdal G, Sivonen K, Doskeland SO, Herfindal L. 2012. The lipopeptide toxins anabaenolysin A and B target biological membranes in a cholesterol-dependent manner. *Biochim Biophys Acta Biomembr* 1818:3000–3009. <https://doi.org/10.1016/j.bbmem.2012.07.015>.
7. Tomek P, Hrouzek P, Kuzma M, Sýkora J, Fišer R, Černý J, Novák P, Bártová S, Šimek P, Hof M, Kavan D, Kopecký J. 2015. Cytotoxic

- lipopeptide muscotoxin A, isolated from soil cyanobacterium *Desmonostoc muscorum*, permeabilizes phospholipid membranes by reducing their fluidity. *Chem Res Toxicol* 28:216–224. <https://doi.org/10.1021/tx500382b>.
8. Vestola J, Shishido TK, Jokela J, Fewer DP, Aitio O, Permi P, Wahlsten M, Wang H, Rouhiainen L, Sivonen K. 2014. Hassallidins, antifungal glycolipopeptides, are widespread among cyanobacteria and are the end-product of a nonribosomal pathway. *Proc Natl Acad Sci U S A* 111: E1909–E1917. <https://doi.org/10.1073/pnas.1320913111>.
9. Gregson JM, Chen JL, Patterson GML, Moore RE. 1992. Structures of puwainaphycins A-E. *Tetrahedron* 48:3727–3734. [https://doi.org/10.1016/S0040-4020\(01\)92264-1](https://doi.org/10.1016/S0040-4020(01)92264-1).
10. Kang HS, Kronic A, Shen Q, Swanson SM, Orjala J. 2011. Minutissamides A-D, antiproliferative cyclic decapeptides from the cultured cyanobacterium *Anabaena minutissima*. *J Nat Prod* 74:1597–1605. <https://doi.org/10.1021/np2002226>.
11. Kang HS, Sturdy M, Kronic A, Kim H, Shen Q, Swanson SM, Orjala J. 2012. Minutissamides E-L, antiproliferative cyclic lipodecapeptides from the cultured freshwater cyanobacterium cf. *Anabaena* sp. *Biorg Med Chem* 20:6134–6143. <https://doi.org/10.1016/j.bmc.2012.08.017>.
12. Mareš J, Hájek J, Urajová P, Kopecký J, Hrouzek P. 2014. A hybrid non-ribosomal peptide/polyketide synthetase containing fatty-acyl ligase (FAAL) synthesizes the beta-amino fatty acid lipopeptides puwainaphycins in the cyanobacterium *Cylindrospermum alatosporum*. *PLoS One* 9:e111904. <https://doi.org/10.1371/journal.pone.0111904>.
13. Urajová P, Hájek J, Wahlsten M, Jokela J, Galica T, Fewer DP, Kust A, Zapomělová-Kozlíková E, Delawska K, Sivonen K, Kopecký J, Hrouzek P. 2016. A liquid chromatography-mass spectrometric method for the detection of cyclic beta-amino fatty acid lipopeptides. *J Chromatogr A* 1438:76–83. <https://doi.org/10.1016/j.chroma.2016.02.013>.
14. Cheel J, Urajová P, Hájek J, Hrouzek P, Kuzma M, Bouju E, Faure K, Kopecký J. 2017. Separation of cyclic lipopeptide puwainaphycins from cyanobacteria by countercurrent chromatography combined with polymeric resins and HPLC. *Anal Bioanal Chem* 409:917–930. <https://doi.org/10.1007/s00216-016-0066-z>.
15. Moore RE, Bornemann V, Niemczura WP, Gregson JM, Chen JL, Norton TR, Patterson GML, Helms GL. 1989. Puwainaphycin C, a cardioactive cyclic peptide from the blue-green alga *Anabaena* BQ-16-1. Use of two dimensional carbon-13-carbon-13 and carbon-13-nitrogen-15 correlation spectroscopy in sequencing the amino acid units. *J Am Chem Soc* 111:6128–6132. <https://doi.org/10.1021/ja00198a021>.
16. Duitman EH, Hamoen LW, Rembold M, Venema G, Seitz H, Saenger W, Bernhard F, Reinhardt R, Schmidt M, Ullrich C, Stein T, Leenders F, Vater J. 1999. The mycosubtilin synthetase of *Bacillus subtilis* ATCC6633: a multifunctional hybrid between a peptide synthetase, an amino transferase, and a fatty acid synthase. *Proc Natl Acad Sci U S A* 96: 13294–13299. <https://doi.org/10.1073/pnas.96.23.13294>.
17. Tsuge K, Akiyama T, Shoda M. 2001. Cloning, sequencing, and characterization of the iturin A operon. *J Bacteriol* 183:6265–6273. <https://doi.org/10.1128/JB.183.21.6265-6273.2001>.
18. Koumoutsis A, Chen XH, Henne A, Liesegang H, Hitzeroth G, Franke P, Vater J, Borriss R. 2004. Structural and functional characterization of gene clusters directing nonribosomal synthesis of bioactive cyclic lipopeptides in *Bacillus amyloliquefaciens* strain FZB42. *J Bacteriol* 186: 1084–1096. <https://doi.org/10.1128/JB.186.4.1084-1096.2004>.
19. Sood S, Steinmetz H, Beims H, Mohr KI, Stadler M, Djukic M, von der Ohe W, Steinert M, Daniel R, Müller R. 2014. Paenilarvins: iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. *Chembiochem* 15:1947–1955. <https://doi.org/10.1002/cbic.201402139>.
20. Ramaswamy AV, Sorrelis CM, Gerwick WH. 2007. Cloning and biochemical characterization of the hectochlorin biosynthetic gene cluster from the marine cyanobacterium *Lynngbya majuscula*. *J Nat Prod* 70: 1977–1986. <https://doi.org/10.1021/np0704250>.
21. Micallef ML, D'Agostino PM, Sharma D, Viswanathan R, Moffitt MC. 2015. Genome mining for natural product biosynthetic gene clusters in the subsection V cyanobacteria. *BMC Genomics* 16:669. <https://doi.org/10.1186/s12864-015-1855-z>.
22. Edwards DJ, Marquez BL, Nogle LM, McPhail K, Goeger DE, Roberts MA, Gerwick WH. 2004. Structure and biosynthesis of the jamaicamides, new mixed polyketide-peptide neurotoxins from the marine cyanobacterium *Lynngbya majuscula*. *Chem Biol* 11:817–833. <https://doi.org/10.1016/j.chembiol.2004.03.030>.
23. Galica T, Hrouzek P, Mareš J. 2017. Genome mining reveals high incidence of putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP ligase genes in biofilm-forming cyanobacteria. *J Phycol* 53:985–998. <https://doi.org/10.1111/jpy.12555>.
24. Arora P, Goyal A, Natarajan VT, Rajakumara E, Verma P, Gupta R, Yousuf M, Trivedi OA, Mohanty D, Tyagi A, Sankaranarayanan R, Gokhale RS. 2009. Mechanistic and functional insights into fatty acid activation in *Mycobacterium tuberculosis*. *Nat Chem Biol* 5:166–173. <https://doi.org/10.1038/nchembio.143>.
25. Liu Z, Iøerger TR, Wang F, Sacchetti JC. 2013. Structures of *Mycobacterium tuberculosis* FadD10 protein reveal a new type of adenylate-forming enzyme. *J Biol Chem* 288:18473–18483. <https://doi.org/10.1074/jbc.M113.466912>.
26. Coates RC, Podell S, Korobeynikov A, Lapidus A, Pevzner P, Sherman DH, Allen EE, Gerwick L, Gerwick WH. 2014. Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS One* 9:e85140. <https://doi.org/10.1371/journal.pone.0085140>.
27. Mohanty D, Sankaranarayanan R, Gokhale RS. 2011. Fatty acyl-AMP ligases and polyketide synthases are unique enzymes of lipid biosynthetic machinery in *Mycobacterium tuberculosis*. *Tuberculosis (Edinb)* 91:448–455. <https://doi.org/10.1016/j.tube.2011.04.006>.
28. Goyal A, Verma P, Anandhakrishnan M, Gokhale RS, Sankaranarayanan R. 2012. Molecular basis of the functional divergence of fatty acyl-AMP ligase biosynthetic enzymes of *Mycobacterium tuberculosis*. *J Mol Biol* 416:221–238. <https://doi.org/10.1016/j.jmb.2011.12.031>.
29. Villiers BRM, Hollfelder F. 2009. Mapping the limits of substrate specificity of the adenylation domain of TycA. *Chembiochem* 10:671–682. <https://doi.org/10.1002/cbic.200800553>.
30. Christiansen G, Philmus B, Hemscheidt T, Kurmayer R. 2011. Genetic variation of adenylation domains of the anabaenopeptin synthesis operon and evolution of substrate promiscuity. *J Bacteriol* 193: 3822–3831. <https://doi.org/10.1128/JB.00360-11>.
31. Brandenburger E, Braga D, Kombrink A, Lackner G, Gressler J, Künzler M, Hoffmeister D. 2018. Multi-genome analysis identifies functional and phylogenetic diversity of basidiomycete adenylate-forming reductases. *Fungal Genet Biol* 112:55–63. <https://doi.org/10.1016/j.fgb.2016.07.008>.
32. Rouhiainen L, Jokela J, Fewer DP, Urmann M, Sivonen K. 2010. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). *Chem Biol* 17:265–273. <https://doi.org/10.1016/j.chembiol.2010.01.017>.
33. Kleigrewe K, Almaliti J, Tian IY, Kinnel RB, Korobeynikov A, Monroe EA, Duggan BM, Di Marzo V, Sherman DH, Dorresteijn PC, Gerwick L, Gerwick WH. 2015. Combining mass spectrometric metabolic profiling with genomic analysis: a powerful approach for discovering natural products from cyanobacteria. *J Nat Prod* 78:1671–1682. <https://doi.org/10.1021/acs.jnatprod.5b00301>.
34. Voráčková K, Hájek J, Mareš J, Urajová P, Kuzma M, Cheel J, Villunger A, Kapuscik A, Bally M, Novák P, Kabeláč M, Krumschnabel G, Lukeš M, Voloshko L, Kopecký J, Hrouzek P. 2017. The cyanobacterial metabolite nocuolin A is a natural oxadiazine that triggers apoptosis in human cancer cells. *PLoS One* 12:e0172850. <https://doi.org/10.1371/journal.pone.0172850>.
35. He J, Hertweck C. 2004. Biosynthetic origin of the rare nitroaryl moiety of the polyketide antibiotic aureothin: involvement of an unprecedented N-oxygenase. *J Am Chem Soc* 126:3694–3695. <https://doi.org/10.1021/ja039328t>.
36. Choi YS, Zhang HJ, Brunzelle JS, Nair SK, Zhao HM. 2008. In vitro reconstitution and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in aureothin biosynthesis. *Proc Natl Acad Sci U S A* 105:6858–6863. <https://doi.org/10.1073/pnas.0712073105>.
37. Fewer DP, Wahlsten M, Osterholm J, Jokela J, Rouhiainen L, Kaasalainen U, Rikkinen J, Sivonen K. 2013. The genetic basis for O-acetylation of the microcystin toxin in cyanobacteria. *Chem Biol* 20:861–869. <https://doi.org/10.1016/j.chembiol.2013.04.020>.
38. Murray IA, Shaw WV. 1997. O-acetyltransferases for chloramphenicol and other natural products. *Antimicrob Agents Chemother* 41:1–6.
39. Zapomělová E, Jezberová J, Hrouzek P, Hisem D, Řeháková K, Komárková J. 2009. Polyphasic characterization of three strains of *Anabaena reniformis* and *Aphanizomenon aphanizomenoides* (cyanobacteria) and their reclassification to *Sphaerospermum* gen. nov. (incl. *Anabaena kisseleviana*). *J Phycol* 45:1363–1373. <https://doi.org/10.1111/j.1529-8817.2009.00758.x>.
40. Skulberg OM, Mysterud I, Karlsen J, Tønnesen HH, Laane CMM, Schumacher T. 2012. *Alveld* research per annum 2012: searchlight on cyanobacteria we have minor knowledge of. *Biolog* 30:32–41. (In Norwegian.)
41. Tønnesen HH, Mysterud I, Karlsen J, Skulberg OM, Laane CMM, Schum-

- acher T. 2013. Identification of singlet oxygen photosensitizers in lambs drinking water in an alveld risk area in West Norway. *J Photochem Photobiol B* 119:37–45. <https://doi.org/10.1016/j.jphotobiol.2012.12.003>.
42. Hegge AB, Mysterud I, Karlsen J, Skulberg OM, Laane CMM, Schumacher T, Tønnesen HH. 2013. Impaired secondary oxidant deactivation capacity and enhanced oxidative stress in serum from alveld affected lambs. *J Photochem Photobiol B* 126:126–134. <https://doi.org/10.1016/j.jphotobiol.2013.07.005>.
43. Heinze R. 1996. A biotest for hepatotoxins using primary rat hepatocytes. *Phycologia* 35:89–93. <https://doi.org/10.2216/i0031-8884-35-65-89.1>.
44. Skulberg OM. 1996. Toxins produced by cyanophytes in Norwegian inland waters—health and environment, p 197–216. In Låg J (ed), *Chemical data as a basis of geomedical investigations*. The Norwegian Academy of Science and Letters, Oslo, Norway.
45. Jones CG, Firn RD. 1991. On the evolution of plant secondary chemical diversity. *Philos Trans R Soc Lond B* 333:273–280. <https://doi.org/10.1098/rstb.1991.0077>.
46. Firn RD, Jones CG. 2000. The evolution of secondary metabolism: a unifying model. *Mol Microbiol* 37:989–994.
47. Rippka R, Deruelles J, Waterbury JB, Herdman M, Stanier RY. 1979. Generic assignments, strain histories and properties of pure cultures of cyanobacteria. *J Gen Microbiol* 111:1–61. <https://doi.org/10.1099/00221287-111-1-1>.
48. Skulberg R, Skulberg OM. 1990. Research with algal cultures—NIVA's Culture Collection of Algae. NIVA report. Norwegian Institute for Water Research, Oslo, Norway.
49. Taton A, Grubisic S, Brambilla E, De Wit R, Wilmotte A. 2003. Cyanobacterial diversity in natural and artificial microbial mats of Lake Fryxell (McMurdo dry valleys, Antarctica): a morphological and molecular approach. *Appl Environ Microbiol* 69:5157–5169. <https://doi.org/10.1128/AEM.69.9.5157-5169.2003>.
50. Bankevich A, Nurk S, Antipov D, Gurevich A, Dvorkin M, Kulikov AS, Lesin V, Nikolenko S, Pham S, Prjibelski A, Pyshkin A, Sirotkin A, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–497. <https://doi.org/10.1089/cmb.2012.0021>.
51. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics* 23: 673–679. <https://doi.org/10.1093/bioinformatics/btm009>.
52. Weber T, Blin K, Duddela S, Krug D, Kim HU, Brucoleri R, Lee SY, Fischbach MA, Muller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43:W237–W243. <https://doi.org/10.1093/nar/gkv437>.
53. Blin K, Wolf T, Chevrette MG, Lu XW, Schwalen CJ, Kautsar SA, Duran HGS, Santos E, Kim HU, Nave M, Dickschat JS, Mitchell DA, Shelest E, Breitling R, Takano E, Lee SY, Weber T, Medema MH. 2017. antiSMASH 4.0—improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 45:W36–W41. <https://doi.org/10.1093/nar/gkx319>.
54. Wang M, Carver JJ, Phelan VV, Sanchez LM, Garg N, Peng Y, Nguyen DD, Watrous J, Kapono CA, Luzzatto-Knaan T, Porto C, Bouslimani A, Melnik AV, Meehan MJ, Liu W-T, Crüsemann M, Boudreau PD, Esquenazi E, Sandoval-Calderón M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu C-C, Floros DJ, Gavilan RG, Kleigrewe K, Northen T, Dutton RJ, Parrot D, Carlson EE, Aigle B, Michelsen CF, Jelsbak L, Sohlenkamp C, Pevzner P, Edlund A, McLean J, Piel J, Murphy BT, Gerwick L, Liaw C-C, Yang Y-L, Humpf H-U, Maansson M, Keyzers RA, Sims AC, et al. 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* 34:828–837. <https://doi.org/10.1038/nbt.3597>.
55. Johansen JR, Bohunická M, Lukešová A, Hřčková K, Vaccarino MA, Chesarino NM. 2014. Morphological and molecular characterization within 26 strains of the genus *Cylindrospermum* (Nostocaceae, Cyanobacteria) with descriptions of three new species. *J Phycol* 50:187–202. <https://doi.org/10.1111/jpy.12150>.
56. Tamrakar A. 2016. Isolation of benthic cyanobacteria and screening of bioactivities and natural products from culture collection strains. Master's Thesis. University of Helsinki, Helsinki, Finland.
57. Kantz T, Bold HC. 1969. Phycological studies IX. Morphological and taxonomic investigations of *Nostoc* and *Anabaena* in culture. Publication no. 6924. University of Texas, Austin, TX.

#### **4.4. Paper IV.**



# Cyanochelins, an Overlooked Class of Widely Distributed Cyanobacterial Siderophores, Discovered by Silent Gene Cluster Awakening

Tomáš Galica<sup>a,b</sup>, Nicola Borbone,<sup>c</sup> Jan Mareš<sup>a,b,d</sup>, Andreja Kust<sup>a,d</sup>, Alessia Caso,<sup>c</sup> Germana Esposito,<sup>c</sup> Kumar Saurav,<sup>a</sup> Jan Hájek<sup>a,b</sup>, Klára Řeháková,<sup>d,e</sup> Petra Urajová,<sup>a</sup> Valeria Costantino,<sup>c</sup> Pavel Hrouzek<sup>a</sup>

<sup>a</sup>Centre Algatech, Institute of Microbiology, The Czech Academy of Sciences (CAS), Třeboň, Czech Republic

<sup>b</sup>Faculty of Science, University of South Bohemia, České Budějovice, Czech Republic

<sup>c</sup>Dipartimento di Farmacia, Università degli Studi di Napoli Federico II, Naples, Italy

<sup>d</sup>Biology Centre of the CAS, Institute of Hydrobiology, České Budějovice, Czech Republic

<sup>e</sup>Institute of Botany of the Czech Academy of Sciences, Třeboň, Czech Republic

**ABSTRACT** Cyanobacteria require iron for growth and often inhabit iron-limited habitats, yet only a few siderophores are known to be produced by them. We report that cyanobacterial genomes frequently encode polyketide synthesis (PKS)/nonribosomal peptide synthetase (NRPS) biosynthetic pathways for synthesis of lipopeptides featuring  $\beta$ -hydroxyaspartate ( $\beta$ -OH-Asp), a residue known to be involved in iron chelation. Iron starvation triggered the synthesis of  $\beta$ -OH-Asp lipopeptides in the cyanobacteria *Rivularia* sp. strain PCC 7116, *Leptolyngbya* sp. strain NIES-3755, and *Rubidibacter lacunae* strain KORDI 51-2. The induced compounds were confirmed to bind iron by mass spectrometry (MS) and were capable of  $\text{Fe}^{3+}$  to  $\text{Fe}^{2+}$  photoreduction, accompanied by their cleavage, when exposed to sunlight. The siderophore from *Rivularia*, named cyanochelin A, was structurally characterized by MS and nuclear magnetic resonance (NMR) and found to contain a hydrophobic tail bound to phenolate and oxazole moieties followed by five amino acids, including two modified aspartate residues for iron chelation. Phylogenomic analysis revealed 26 additional cyanochelin-like gene clusters across a broad range of cyanobacterial lineages. Our data suggest that cyanochelins and related compounds are widespread  $\beta$ -OH-Asp-featuring cyanobacterial siderophores produced by phylogenetically distant species upon iron starvation. Production of photolabile siderophores by phototrophic cyanobacteria raises questions about whether the compounds facilitate iron monopolization by the producer or, rather, provide  $\text{Fe}^{2+}$  for the whole microbial community via photoreduction.

**IMPORTANCE** All living organisms depend on iron as an essential cofactor for indispensable enzymes. However, the sources of bioavailable iron are often limited. To face this problem, microorganisms synthesize low-molecular-weight metabolites capable of iron scavenging, i.e., the siderophores. Although cyanobacteria inhabit the majority of the Earth's ecosystems, their repertoire of known siderophores is remarkably poor. Their genomes are known to harbor a rich variety of gene clusters with unknown function. Here, we report the awakening of a widely distributed class of silent gene clusters by iron starvation to yield cyanochelins,  $\beta$ -hydroxy aspartate lipopeptides involved in iron acquisition. Our results expand the limited arsenal of known cyanobacterial siderophores and propose products with ecological function for a number of previously silent gene clusters.

**KEYWORDS** cyanobacteria, iron acquisition, lipopeptides, secondary metabolism, siderophores

Iron, the fourth most abundant element in the Earth's crust, is an important metal cofactor to many metabolic enzymes and an essential nutrient to all known living

**Citation** Galica T, Borbone N, Mareš J, Kust A, Caso A, Esposito G, Saurav K, Hájek J, Řeháková K, Urajová P, Costantino V, Hrouzek P. 2021. Cyanochelins, an overlooked class of widely distributed cyanobacterial siderophores, discovered by silent gene cluster awakening. *Appl Environ Microbiol* 87:e03128-20. <https://doi.org/10.1128/AEM.03128-20>.

**Editor** Hideaki Nojiri, University of Tokyo

**Copyright** © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Pavel Hrouzek, [hrouzek@alga.cz](mailto:hrouzek@alga.cz).

**Received** 26 December 2020

**Accepted** 4 June 2021

**Accepted manuscript posted online** 2021

**Published**

organisms. Yet, due to the prevalent oxidative conditions on the surface of our planet, it is frequently present in its oxidized  $\text{Fe}^{3+}$  form that is generally insoluble and not easily acquired by living organisms (1, 2). Microorganisms have evolved various strategies to cope with this problem, one of which is the production and excretion of siderophores, which are low-molecular-weight metabolites able to reversibly bind an iron atom (1, 3). Excreted siderophores are able to pick up even trace amounts of iron from the environment and facilitate cellular iron uptake or prevent precipitation and loss of its bioavailable forms (1).

Cyanobacteria depend on iron as do all other microorganisms and live in various habitats, including those with a low availability of iron as a nutrient. Several mechanisms of cyanobacterial iron stress response have been described, such as remodeling of photosynthetic complexes, upregulation of iron transport systems, and siderophore production (2, 3). Yet, intriguingly, few cyanobacterial siderophores are known; only two hydroxamate siderophores (schizokinen and synechobactin) and a single catecholate siderophore (anachelin) have been structurally characterized (2). Cyanobacterial genomes are known to harbor a large number of gene clusters (4) that are silent in luxuriant cultivation medium under laboratory conditions. These clusters may be activated if an appropriate trigger is present and, possibly, some may encode production of unknown siderophores, as shown in other bacteria (5).

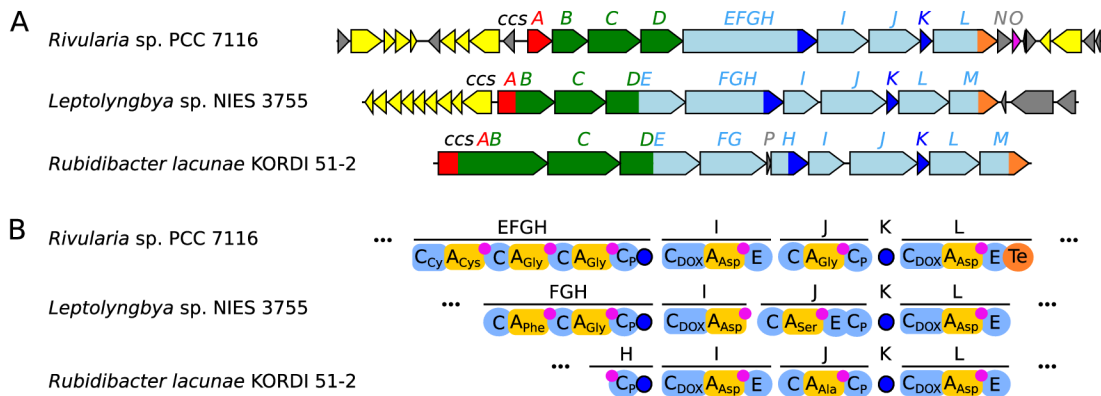
Knowledge of siderophores from other bacteria shows there are a number of structural motifs capable of iron chelation, some of which can be expected to also occur in cyanobacteria (1, 6). The carboxylate moiety of  $\beta$ -hydroxy aspartate ( $\beta$ -OH-Asp) is one such a case (1, 6). Several bacterial siderophores, e.g., serobactin, cupriachelin, and pacifibactin, were found to employ  $\beta$ -OH-Asp for iron chelation (7–9). Their synthesis is driven by nonribosomal peptide synthetases (NRPS) with specialized modules for selective incorporation of  $\beta$ -OH-Asp into the nascent chain (7–9). The modules consist of an adenylation domain (A-domain) that activates aspartic acid (Asp), a dioxygenase that hydroxylates the activated Asp on its  $\beta$  carbon, and a specific condensation domain (C-domain) that does not effectively proceed to form a peptide bond prior to successful hydroxylation of Asp to  $\beta$ -OH-Asp (7, 10).

A common feature of  $\beta$ -OH-Asp siderophores is the presence of a prolonged hydrocarbon chain on the N terminus (1). The hydrocarbon chain is thought to play a role in adjusting diffusion rates via an enhanced association with biological membranes and so limit siderophore loss, especially in the case of microbes living in aquatic environments (1, 11). A prominent mechanism of N-terminal acylation of cyanobacterial nonribosomal peptides is the employment of fatty acyl-AMP ligase (FAAL) at the beginning of synthesis (12).

The genetic signature for both the NRPS module for  $\beta$ -OH-Asp and the FAAL domain can be identified in nucleotide sequences, and the growing number of publicly available cyanobacterial genomes provides suitable data for targeted genome mining. Our recent mining of cyanobacterial gene clusters that possibly encode biosynthesis of lipopeptides yielded, among others, a group of 10 similar clusters with features that suggest the presence of  $\beta$ -OH-Asp in the product and offered a direction for further research (13). In the present study, we performed a bioinformatic analysis of selected silent gene clusters, induced the production of a new type of cyanobacterial siderophores by iron deprivation, described the structure of a selected representative and, finally, took a closer look at the distribution across the cyanobacterial phylum.

## RESULTS

**Identification of  $\beta$ -OH-Asp lipopeptide gene clusters.** Three hybrid gene clusters that combine elements of polyketide synthesis (PKS) and NRPS (PKS/NRPS) putatively coding for  $\beta$ -OH-Asp-containing peptides were found (Fig. 1A, Table S1 to S3 in the supplemental material). One cluster was located on the chromosome of *Rivularia* sp. PCC 7116 (CP003549, 8.7 Mbp), the second was on a plasmid from *Leptolyngbya* sp. NIES-3755 (AP017310.1, 97 kbp), and the third was on a whole-genome scaffold from



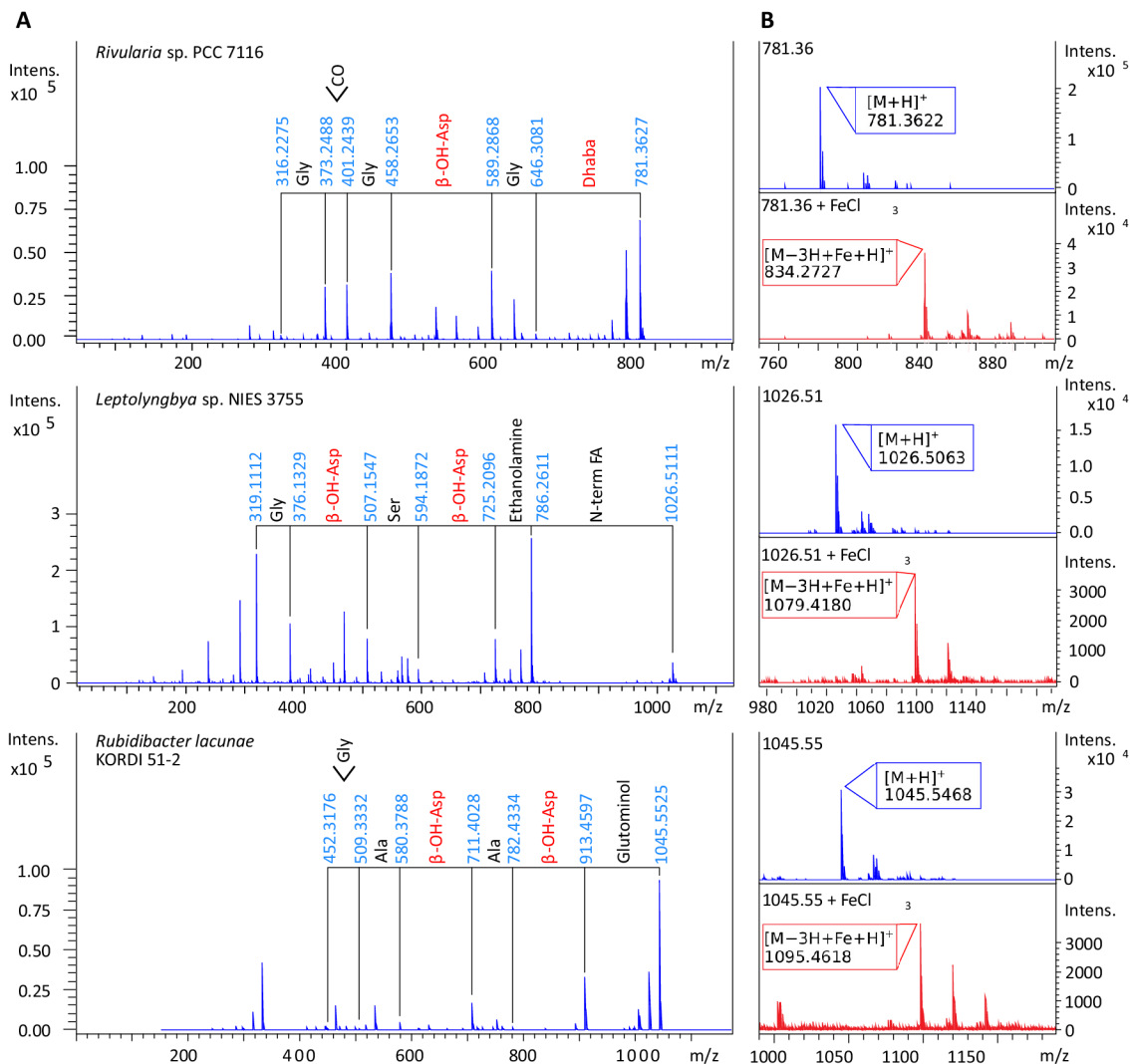
**FIG 1** (A) Gene clusters encoding hybrid PKS/NRPS biosynthesis of cyanochelin with proportions of genes colored according to the type of biosynthetic step (FAAL, red; PKS, green; NRPS, blue; dioxygenase involved in  $\beta$ -OH-Asp formation, purple; thioesterase/terminal reductase, orange; aldol-keto reductase, pink) and genes related to siderophore transport (depicted in yellow). (B) Selected NRPS modules and related dioxygenases responsible for incorporation of two  $\beta$ -OH-Asp into cyanochelins. C, condensation domain; A, adenylation domain; E, epimerase domain; Te, thioesterase domain; pink, thiolation domain; dark blue, dioxygenase. Subscript specifies the type of C domain or specificity of A-domain.

*Rubidibacter lacunae* KORDI 51-2 ([ASSJ01000034.1](https://doi.org/10.1093/aem/assj01000034.1), 54 kbp). Organization of the biosynthetic genes in one direction, with genes for tailoring enzymes scattered among the PKS/NRPS genes, suggests that the chronology of biosynthetic steps will reflect the cluster topology. All three gene clusters encode a fatty acyl-AMP ligase (FAAL) domain (in *Rivularia* as a standalone, in the other two fused to subsequent PKS modules) followed by several PKS genes (*ccsB* to *D*), suggesting the presence of a fatty acyl residue elongated by several (modified) malonyl units at the N terminus of the compounds. Analysis of the specificity of the individual NRPS A-domains contained in *ccsE* to *M* (Table S4) led to prediction of a peptide sequence composed of 6 to 7 amino acids.

The first NRPS module (*CcsE*) contains an A-domain with predicted specificity for cysteine preceded by a heterocyclization domain, suggesting the formation of a thiazole heterocycle attached to a nascent N-terminal chain. The following two amino acids activated by the downstream NRPS modules *CcsFG* vary among the gene clusters: Gly-Gly (*Rivularia*), Phe-Gly (*Leptolyngbya*), and Gly-Ala (*Rubidibacter*). Most importantly, in all three candidate clusters, the presence of two Asp-specific NRPS modules (*ccsI* and *ccsL*) with associated dioxygenases (*ccsH* and *ccsK*) suggests the incorporation of two  $\beta$ -OH-Asp in the putative product (Fig. 1B). The  $\beta$ -OH-Asp-specific modules are separated by NRPS modules (*ccsJ*) with predicted specificities to Gly (*Rivularia*) or Ser (*Leptolyngbya* and *Rubidibacter*). In *Rubidibacter* and *Leptolyngbya*, an additional NRPS module activating Glu and Gly, respectively, is encoded at the 3' end. While in *Rivularia* the gene cluster is terminated by a canonical thioesterase domain, the last module in *Leptolyngbya* and *Rubidibacter* instead contains a terminal reductase domain. An analogous organization of the terminal NRPS module was previously reported to result in the formation of a primary alcohol at the C terminus (14). Intriguingly, an enzyme of analogous function (aldol-keto reductase) is encoded within the *Rivularia* gene cluster by the *ccsN* gene, immediately following the terminal NRPS module.

Several genes encoding siderophore transporters and accessory processing proteins (colored yellow in Fig. 2, Tables S5 and S6) are localized in the immediate vicinity of the candidate clusters in *Rivularia* and *Leptolyngbya* and, thus, strengthen the evidence that the putative products of the investigated biosynthetic gene clusters (BGCs) are siderophores. In *Rubidibacter*, the biosynthetic genes span over the whole 54-kb genomic fragment and thus any further analysis of the surrounding genomic context is hindered.

In bacteria, iron-stress response is frequently controlled by a transcriptional repressor called ferric uptake regulator (*FurA*), which is bound to a DNA regulatory sequence called a furbox (3, 15). An *in silico* search yielded highly scoring putative furboxes located upstream of the first biosynthesis-related gene in both *Rivularia* and *Leptolyngbya* (Tables S7 and S8).



**FIG 2** HRMS experiments demonstrating key features of cyanochelins. (A) MS/MS fragmentation spectra of  $\beta$ -OH-Asp siderophores detected at  $m/z$  781.36, 1,026.51, and 1,045.55, demonstrating neutral loss characteristic for  $\beta$ -OH-Asp residues (see also Table S9 to S11 in the supplemental material). (B) Mass spectra obtained after injection of cyanochelin A1 and B1 with (red) or without (blue) supplemented  $\text{Fe}^{3+}$ .

In both aforementioned cases, the promoter region of the biosynthetic genes is shared with a cassette of genes coding for siderophore transporters in the opposite direction and an additional furbox, likely regulating this cassette. In *Rivularia*, the furbox for transporter cassette partly overlaps with the furbox for biosynthetic genes. Such findings further support the hypothesis that the DNA region upstream of *ccsA* is likely to be a regulatory region responsive to iron stress via FurA that governs the expression of the clusters described above.

**Detection of  $\beta$ -OH-Asp lipopeptides.** Upon iron deprivation, cultures of all the three strains accumulated compound(s) that showed tandem mass spectrometry (MS/MS) fragmentation patterns highly congruent with the bioinformatic prediction. Fragmentation analysis of compound with  $m/z$  at 781.36 found in *Rivularia* confirmed the partial amino acid sequence [Gly]-[Gly]-[ $\beta$ -OH-Asp]-[Gly] (Fig. 2, Fig. S1, Table S9). For candidate compounds from *Leptolyngbya* ( $m/z$  1,026.51) and *Rubidibacter* ( $m/z$  1,045.51), the analysis provided the AA sequences [Phe]-[Gly]-[ $\beta$ -OH-Asp]-[Ser]-[ $\beta$ -OH-Asp] and [Gly]-[Ala]-[ $\beta$ -OH-Asp]-[Ala]-[ $\beta$ -OH-Asp], respectively (Fig. 2A, Fig. S2 and S3, Table S10 and S11). Cleavage of a  $\beta$ -OH-Asp moiety likely involved in the Fe chelation was represented by a distinctive neutral loss of 131.02 within the peptide sequence.

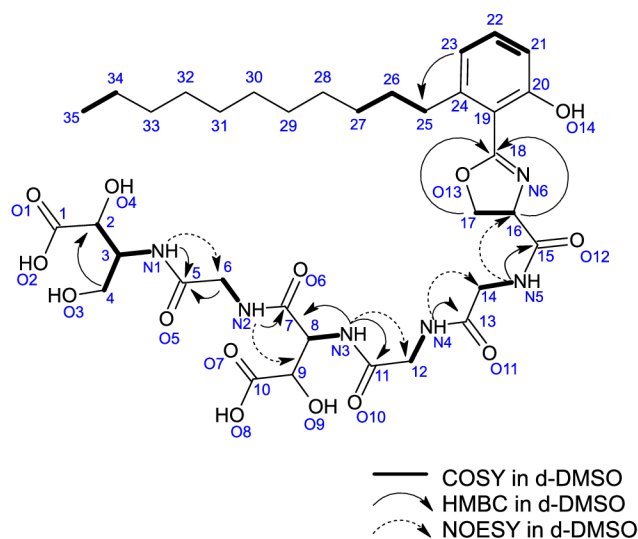
Furthermore, all the compounds provided neutral loss corresponding to the presence of an amino acid with a carboxylic functional group reduced to its primary alcohol: *Rivularia* (135.05, Asp reduced to dihydroxy-amino-butyric acid [Dhaba]); *Leptolyngbya* (61.05, Gly reduced to ethanolamine); and *Rubidibacter* (132.09, Gln reduced to glutaminol). Additionally, all the detected compounds provided various a and b types of N-terminal fragments, suggesting the linkage of the thiazole/oxazole heterocycle to an aliphatic chain (Fig. S1 to S3 and Table S9 to S11). *Rivularia* and *Leptolyngbya* strains produced additional compounds at  $m/z$  753.33 and 998.48 with the same partial AA sequence as 781.36 and 1,026.51, respectively, but with different N-terminal fragments (Fig. S4A and B). Based on their  $m/z$ , it was possible to conclude that these variants possess hydrocarbon chains shorter by two methylene groups ( $\Delta \sim 28.0307$  Da,  $C_2H_4$ ) (Fig. S4). Finally, the compound at  $m/z$  1,046.54 (a variant of 1,045.55 from *Rubidibacter*) was found to differ in the last AA containing hydroxy-amino-pentanoic acid instead of glutaminol (Fig. S4C) as a result of the relaxed A-domain specificity, which activates either Glu or Gln, which is subsequently reduced to a corresponding alcohol.

Interestingly, all candidate compounds were found in their putative apo- and iron-complexed ( $m/z$  greater by 52.9120 Da) form, with the latter eluting earlier during analysis (Fig. S5). To verify siderophore activity, minute amounts of investigated compounds were isolated and analyzed by direct high-resolution mass spectrometry (HRMS), either with or without roughly equimolar amounts of  $Fe^{3+}$  ions (Fig. 2B). While samples without  $Fe^{3+}$  showed a compound in  $[M+H]^+$  at the expected  $m/z$  along with common adducts ( $Na^+$ ,  $K^+$ ), the samples with supplemented iron showed  $m/z$  greater by 52.91 Da from the original  $m/z$ , i.e., 834.27 for 781.36, 1,079.42 for 1,026.51, and 1,098.46 for 1,045.55, corresponding to  $[M - 3H + Fe + H]^+$ .

The candidate compounds share the key structural features, in particular the two  $\beta$ -OH-Asp residues (alternatively  $\beta$ -OH-Asp/Dhaba) separated by a small amino acid residue close to the C terminus, the hydrocarbon chain on the N terminus, and having the C-terminal amino acid reduced to a primary alcohol. Furthermore, all the candidate compounds are capable of binding iron; hence, we collectively refer to them as cyanochelins and encourage the future use of the term for similar compounds.

**Photoreduction of iron by identified compounds.** A prominent feature of  $\beta$ -OH-Asp siderophores is their ability to reduce ferric iron upon exposure to light, accompanied by cleavage of the siderophore at its peptide backbone close to the  $\beta$ -OH-Asp residue. Minute amounts of cyanochelins ( $m/z$  781.36, 1,026.51, and 1,045.51) were isolated by high-performance liquid chromatography (HPLC) and subsequently exposed to direct sunlight in the presence of  $FeCl_3$  (0.1 mM) and bathophenanthroline disulfonic acid (0.1 mM), which develops colorful complexes with divalent iron. Light-exposed aliquots clearly developed BPDS- $Fe^{2+}$  complexes, manifested as pink coloration of the solution visible by eye and, therefore, confirmed the potential of the isolated compounds to reduce ferric iron. Moreover, in contrast to light-protected samples the high-performance liquid chromatography high-resolution tandem mass spectrometry (HPLC-HRMS/MS) analysis of sunlight-exposed aliquots revealed the expected products of photolysis accompanied by a decreased amount of the candidate compounds. Molecular ions at  $m/z$  475.29 and  $m/z$  476.28 matched the expected weight for photolytic products from  $m/z$  781.36 cleaved between glycine (AA3) and  $\beta$ -OH-Asp (AA4). Furthermore, their MS/MS product ions correspond to that of cyanochelin  $m/z$  781.36 (Fig. S6). Similarly, compounds with  $m/z$  633.40 and  $m/z$  634.39 originate from cyanochelin  $m/z$  1,026.51 and  $m/z$  598.39 from  $m/z$  1,045.55 (Fig. S6).

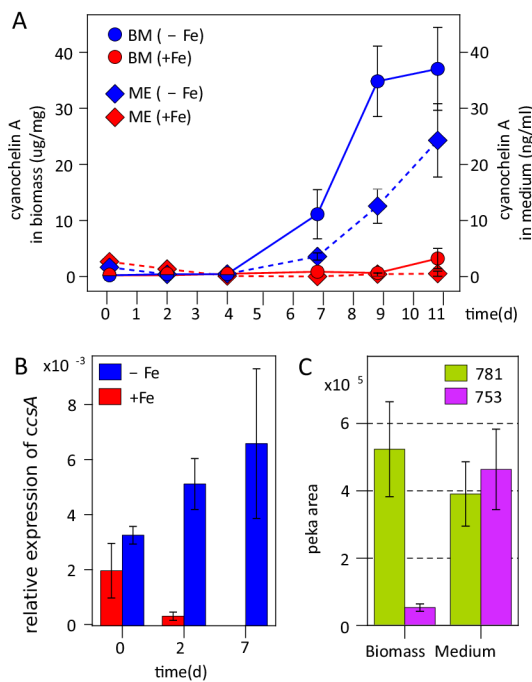
For further experiments, we isolated 2.2 mg of cyanochelin  $m/z$  781.36. An absorbance spectrum of the isolated compound was measured in the presence of equimolar amounts of  $Fe^{3+}$  before and after exposure to direct sunlight. An absorbance peak at 318 nm, corresponding to the  $Fe^{3+}$ - $\alpha$ -hydroxycarboxylate charge transfer band reported previously (9), was present before light exposure but was clearly diminished afterward (Fig. S6B).



**FIG 3** Structure of cyanochelin A. The most relevant 2D NMR correlations are depicted.

**Structural elucidation of cyanochelin A.** Cyanochelin A, the major  $\beta$ -OH-Asp side-phore product from the *Rivularia* strain, was isolated and its chemical structure was elucidated using a combination of HRMS and NMR techniques. We recorded full sets of 1D and 2D NMR spectra ( $^1\text{H}$ ,  $^{13}\text{C}$ , COSY, HMBC, HSQC [Fig. S7 to S11, respectively], TOCSY, NOESY, and ROESY) in  $\text{CD}_3\text{OD}$  (*D*-methanol) and  $(\text{CD}_3)_2\text{SO}$  (*D*-DMSO) deuterated solvents to take advantage of the spread of overlapping signals in different solvents. The analysis in *D*-DMSO also allowed us to trace the COSY connectivities involving NH protons of the amino acidic chain. The most relevant COSY, HMBC, and NOESY correlations and all  $^1\text{H}$  and  $^{13}\text{C}$  chemical shift assignments are summarized in Fig. 3, Table S12, and Fig. S7 to S11. The  $^1\text{H}$ -NMR spectrum of cyanochelin A was clearly suggestive of a lipopeptide molecule with distinctive signals of a saturated hydrocarbon chain, i.e., the presence of a methyl triplet typical for a methyl terminus at  $\delta$  0.85 ppm ( $\text{C}_{35}$ ) (Fig. 3, Table S12), coupled in the COSY spectrum with a methylene at  $\delta$  1.26 ppm ( $\text{C}_{34}$ ) that is part of a 10-methylene-long alkyl chain ( $\delta_{\text{H}}$  1.23 to 2.75 ppm,  $\delta_{\text{C}}$  22.1 to 34.5 ppm). The 11-carbon-long chain is directly linked to a phenol ring, as confirmed by the  $\text{H}_{23} \rightarrow \text{C}_{25}$  HMBC correlation shown in Fig. 3, which is in turn connected to  $\text{C}_{18}$  of the 2-oxazoline moiety, likely resulting from the cyclization of a serine residue. The oxazoline substructure was confirmed by the COSY and HMBC correlations depicted in Fig. 3. The connection of the saturated hydrocarbon chain with a phenolate moiety and the 2-oxazoline was further confirmed by formation of the HRMS/MS fragment at  $m/z$  316.2275 (Fig. 2, Fig. S1, Table S9). The peptide moiety was shown by resonances of five distinct amide-NH signals ( $\delta_{\text{H-DMSO}}$  8.42, 8.20, 7.97, 7.95, and 7.39), three  $\alpha$ -proton signals ( $\delta_{\text{H-DMSO}}$  4.14, 4.64, and 4.94), three methylene resonances attributable to three glycine residues ( $\delta_{\text{H-DMSO}}$  3.68, 3.77, and 3.74), and two oxygenated methines,  $\text{C}_2$  and  $\text{C}_9$ , resonating, respectively, at  $\delta_{\text{H-DMSO}}$  4.23 and 4.43. Finally, the presence of the 2,4-dihydroxy, 3-amino butyric acid (Dhaba) in the structure was confirmed by the full set of COSY and HMBC correlations ( $\text{C}_2$ - $\text{C}_4$ ) (Fig. 3, Table S12). The amino acid sequence was elucidated in combination with observed MS/MS fragmentation. Cyanochelin A monoisotopic  $[\text{M}+\text{H}]^+$  peak at  $m/z$  781.3619 (calc. for  $\text{C}_{35}\text{H}_{53}\text{N}_6\text{O}_{14}^+$ ) provided fragment ions at  $m/z$  646.31 (b5; loss of Dhaba) followed by a sequential loss of Gly (b4;  $m/z$  589.28),  $\beta$ -OH-Asp (b3;  $m/z$  458.26), and Gly (b2;  $m/z$  401.24). Finally, the formation of the a2 and a1 ions at  $m/z$  373.25 and 316.23 can be attributed to fragmentations involving the  $\text{C}_{13}$ - $\text{C}_{14}$  and  $\text{C}_{15}$ - $\text{C}_{16}$  bonds, respectively (Fig. S1, Table S9).

**Cyanochelin A production dynamics.** To further support the hypothesis that the detected  $\beta$ -OH-Asp lipopeptides are the final products of the candidate gene clusters



**FIG 4** Cyanochelin A production in response to iron starvation. (A) Cultivation of *Rivularia* in iron-depleted (–Fe) and control (+Fe) media showed accumulation of cyanochelin A in the biomass (BM) and in the spent medium (ME). (B) Expression of *ccsA*, the first gene of the cyanochelin gene cluster, normalized to *rnpB* during batch cultivation with or without provided iron. (C) Cyanochelin A (*m/z* 781.36) and its variant with a two-carbon-shorter acyl chain (*m/z* 753.33) were found in the biomass and the spent medium in different ratios.

expressed under iron deprivation, we monitored relative mRNA levels of the starter module gene *ccsA* in *Rivularia* (normalized to *rnpB*) along with accumulation of the cyanochelin A and its putative analogue compound *m/z* 753.33 (Fig. 4). We observed a 2-fold increase in the expression of *ccsA* at 7 days after transferring the *Rivularia* culture to iron-depleted medium (Fig. 4B). In contrast, culture inoculated into fresh Fe-full medium exhibited a gradual downregulation of *ccsA* (Fig. 4B) and finally accounted for a >1,000-fold difference in *ccsA* expression between treated and control cultures after 7 days of the experiment. Along with the changes in expression, we also detected the accumulation of cyanochelin A and its variant compound with *m/z* 753.33 in the starved culture. The maximal observed concentration of cyanochelin A at the end of the experiment reached 35 µg per mg of biomass and 25 ng per ml of culture medium. Interestingly, the relative amount of 753.33 to cyanochelin A, evaluated as ion peak areas in HRMS analysis at the end of the experiment, differed between biomass and culture medium. In the culture medium, amounts of both compounds were approximately equal, whereas the amount of 753.33 was proportionally 10 times lower in the biomass.

#### Phylum-wide distribution of cyanochelin-like gene clusters in cyanobacteria.

The cyanochelins reported here were found in strains from three divergent lineages and, hence, indicated a possible widespread distribution in cyanobacteria. Therefore, we investigated publicly available cyanobacterial genomes for the presence of PKS/NRPS gene clusters that would include FAAL and two or more Asp/Asn-specific A-domains with associated β-OH-Asp-specific C-domain in such an arrangement that would be compatible with the putative biosynthesis of cyanochelins. An additional 27 clusters from 29 strains (Table S13) matched the criteria and further interpretation of the clusters split the batch into two groups (Table S14). The first contained the eight clusters with the two β-OH-Asp residues positioned on the N terminus of the expected products, close to the FA, similar to as is found in cupriachelin (7). The clusters in the

second group (19 cases) were expected to encode biosynthesis of lipopeptides with  $\beta$ -OH-Asp residues on the C terminus, as in the isolated cyanochelin A. In 17 cases of the latter group, the  $\beta$ -OH-Asp-specific module is the last PKS/NRPS module; moreover, in 14 cases it is fused to a thioesterase domain, further supporting its C-terminal position. In a majority of the expected products, the  $\beta$ -OH-Asp are separated by Gly (18 cases), analogous to cyanochelin A. Other AA predicted to be found between the two  $\beta$ -OH-Asp are Ser (2), Thr (2), Orn (2), and Asn (1); in the remaining 2 cases, the prediction software could not resolve the specificity of the A-domain. The genetic surroundings of the clusters were investigated for the presence of genes encoding transport-related proteins in a similar manner as was done for *Rivularia* and *Leptolyngbya*. In particular, we looked for TonB-dependent siderophore receptors, as they are known to play a key role in siderophore retrieval in bacteria. TonB-dependent transporter genes were found in the vicinity of 25 out of 27 clusters, mostly (17/25) in close proximity (~1,000 bp) upstream of the starting FAAL-encoding gene (Table S15).

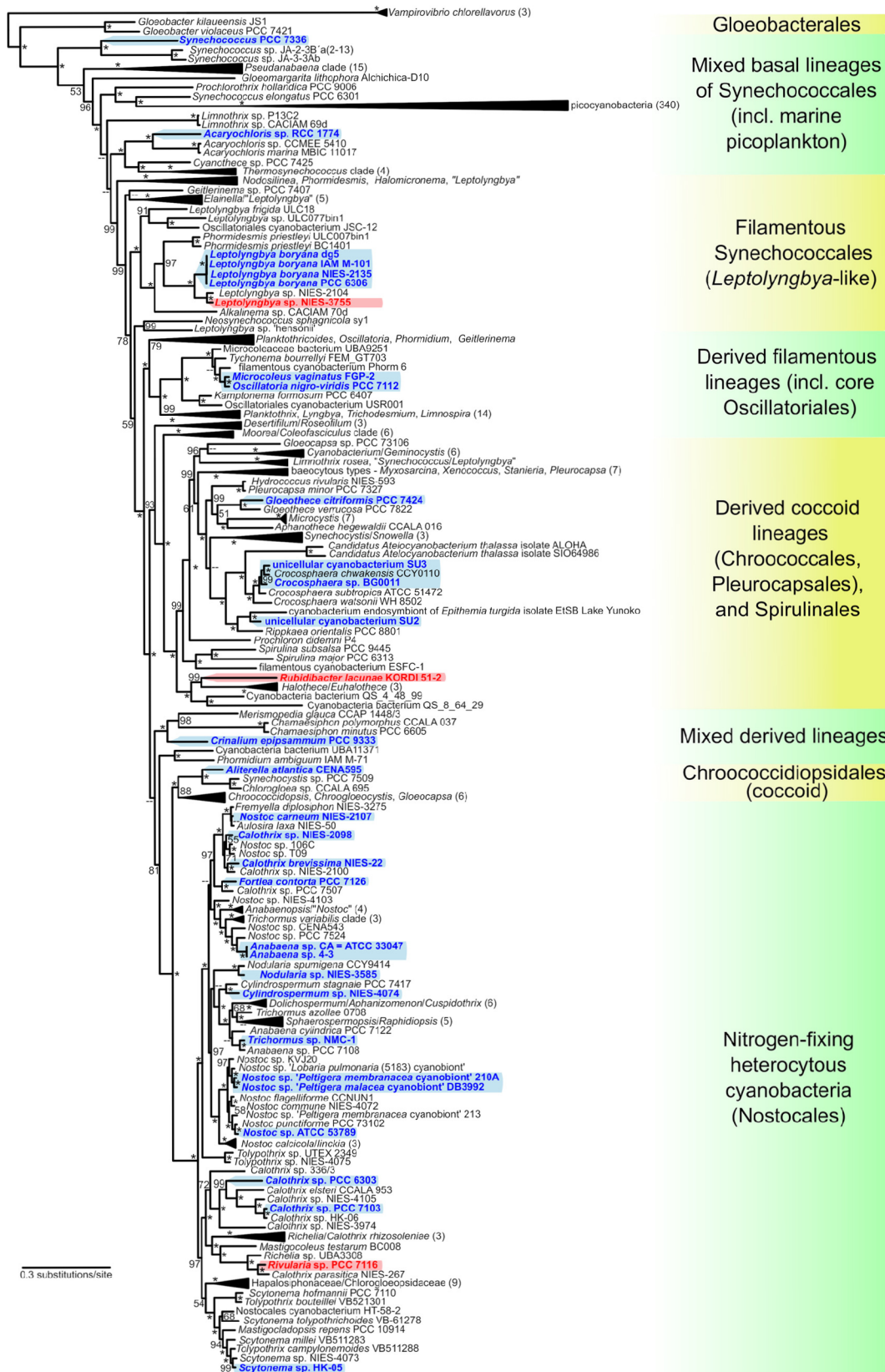
Phylogenomic analysis of the available cyanobacterial genomes (Fig. 5) has shown that the three confirmed cyanochelin-producing strains fall into three divergent phylogenetic clades: a deep-branching lineage of simple filamentous cyanobacteria of the genus *Leptolyngbya*, the derived lineage of halotolerant coccoid cyanobacteria (*Rubidibacter*), and the most derived group of heterocytous cyanobacteria (*Rivularia*). The remaining 29 strains containing cyanochelin-like biosynthetic gene clusters were widely distributed throughout the phylum, encompassing both basal and derived clades, with slightly higher abundance in the heterocytous clade.

## DISCUSSION

Cyanobacteria depend on considerable amounts of iron as an essential cofactor for basic metabolic processes such as photosynthesis, respiration, or nitrogen fixation (2, 3). Therefore, it is surprising that only three types of cyanobacterial siderophores (with limited structural variability) have been documented in the literature. In particular, schizokinen and synechobactin both employ hydroxamate, while anachelin uses catecholate residues for iron chelation (2). The recent accumulation of sequenced cyanobacterial genomes and silent gene clusters detected therein offers a great waypoint for the identification of novel siderophores, if iron-depleted medium is employed for the cultivation. In the present study, we report novel cyanobacterial lipopeptides that employ  $\beta$ -OH-Asp for chelation of iron and, hence, broaden the repertoire of chelating residues and siderophores known from cyanobacteria. The most striking finding was the actual distribution of cyanochelins. The compounds investigated in this study come from vastly separated phylogenetic clades across three cyanobacterial orders. The growth mode among the producing strains varies from simple trichal soil-inhabiting cyanobacteria (*Leptolyngbya*) through unicellular in marine sediment (*Rubidibacter*) to heterocytous filament-forming macroscopic colonies on rocks in the oceanic splash zone (*Rivularia*). An additional 27 gene clusters putatively encoding biosynthesis of cyanochelins or related double  $\beta$ -OH-Asp-containing lipopeptides were found across the major cyanobacterial lineages, spanning from basal clades to the most derived ones (Fig. 5). The observed pattern suggests the wide phylogenetic and ecological distribution of  $\beta$ -OH-Asp-containing siderophores, including cyanochelins. Since the specific cyanochelin genes do not show a close homology between the three confirmed producer strains (Table S16), the production of cyanochelins does not seem to be a result of recent horizontal gene transfer. The question of whether the characteristic structure of cyanochelins is maintained as an ancestral metabolic capacity in cyanobacteria or, rather, serially introduced by functional convergence remains to be addressed when more records of cyanochelin producers are available. Previous studies suggested that strong selective pressure is important in siderophore evolution, as illustrated by the divergent biosynthetic origin of hydroxamate siderophores that can be assembled both by NRPS and NRPS-independent biosynthetic machineries (16).

The two  $\beta$ -OH-Asp separated by a single amino acid residue are expected to





**FIG 5** Phylogenetic distribution of cyanochelin biosynthetic gene clusters. The cyanochelin producers in this study (highlighted in red) are found in three highly divergent lineages of cyanobacteria, whereas candidate cyanochelin biosynthetic (Continued on next page)

participate in iron chelation by alterobactin and have been experimentally proven to provide four coordination bonds for Fe chelation in the bacterial siderophore pacifibactin (9, 17). The participation of  $\beta$ -OH-Asp in iron chelation is manifested in the absorption spectra as a charge transfer band around 300 nm and also as an ability to reduce ferric iron accompanied by lysis of the siderophore upon exposure to light (9). The cyanochelins share these key structural features and properties with the above-mentioned siderophores, which implies that they also employ a double  $\beta$ -OH-Asp motif to provide four out of the six electron pairs required for chelation. As for the remaining two electron pairs, it was not possible to unambiguously determine the donors from the collected data. In the case of cyanochelin A, it is most probable that the phenolate moiety and nitrogen atom form the oxazole ring as reported for mycobactin (1, 18).

The incorporation of  $\beta$ -OH-Asp into cyanochelins is likely implemented in the same manner as in the biosynthesis of cupriachelin and pacifibactin (7, 9) and requires the same enzymatic elements, i.e., dioxygenase participating in conversion of Asp to  $\beta$ -OH-Asp and a C domain with selective affinity for  $\beta$ -OH-Asp. Cupriachelin and pacifibactin have the dioxygenase encoded either as part of a long PKS or NRPS gene or as a standalone gene in a cassette of accessory genes apart from the main body of NRPS genes (7, 9). In the aforementioned cyanochelin gene clusters, the whole machinery that converts Asp to  $\beta$ -OH-Asp and incorporates it into the peptide is integrated among NRPS genes, either fused or standalone, but the order of the genes is fully colinear with the biosynthesis. Such topology is observed in all the other putative cyanochelin gene clusters. Therefore, the observed topology facilitates the identification of further  $\beta$ -OH-Asp-containing NRPS peptides, at least in cyanobacteria, and may improve bioinformatic NRPS prediction tools that still lack accuracy when it comes to modified amino acids.

To limit diffusibility, siderophores from aquatic environments frequently feature a fatty acyl residue linked by a peptide bond (6). In cyanochelin A, the peptide bond is subsequently included into the oxazole ring during the heterocyclization of the first amino acid (serine), further strengthening the binding of the hydrophobic moiety to the peptide core. In an analogous manner, cysteine forms a thiazole ring in cyanochelins from *Rubidibacter* and *Leptolyngbya*. Interestingly, siderophores are often found in multiple variant forms differing only in the size of the hydrophobic moiety and resulting polarity. The N-terminal hydrophobic residue of cyanobacterial siderophores consists of 8 to 14 carbons (synechobactin) (2), whereas in the majority of other bacterial siderophores it consists of 12 to 18 carbons (e.g., marinobactin or amphibactin) (19). Cyanochelins produced by *Rivularia* possess a linear hydrophobic moiety containing 9 to 11 carbons, and in *Rubidibacter* and *Leptolyngbya* this is expected to account for C22 and C18–C20 alkyl chains, respectively. The variability of hydrophobic chains is likely achieved by the extended specificity of FAAL, which recognizes several FAs and introduces them into the biosynthetic assembly line, as reported previously from other cyanobacterial lipopeptides (12). Notably, in *Rivularia*, the ratio between cyanochelin variants possessing different hydrophobic moieties varies between biomass and the culture medium, suggesting they may also exhibit distinct or complementary functions during the iron acquisition. Keeping the siderophore close to the producer prevents losses of the metabolite and may be especially important considering its photolability. Ferrous iron released during photolytic cleavage of the siderophore is likely to be consumed by other organisms unless the reaction happens in close proximity to the producer. On the other hand, siderophores diffusing further away may present a service to the whole community (20).

#### FIG 5 Legend (Continued)

gene clusters (highlighted in blue) are distributed throughout the cyanobacterial tree of life. The phylogenetic tree was inferred using approximately maximum-likelihood in FastTree2 based on a concatenated alignment of 120 universally conserved bacterial proteins employing the Genome Taxonomy Database toolkit *de novo* workflow. The phylogeny includes 607 publicly available and quality-checked cyanobacterial genome assemblies. The tree is rooted by *Vamprvivibrio chlorellavorus* from a lineage sister to photosynthetic cyanobacteria; bootstrap supports are shown at nodes (100% support is marked with asterisks).

The discovery of cyanochelins substantially broadens the collection of cyanobacterial siderophores and provides valuable waypoints for understanding of the orphan gene clusters in this specific phylum. We show that genetic elements encoding synthesis of acylated peptides with two  $\beta$ -OH-Asp functioning as siderophores are widespread in cyanobacteria. In detail, however, the presence of photolabile siderophores in cyanobacteria as phototrophic organisms is a bit surprising and raises intriguing questions about the possible recipients of the benefits implied by the siderophore presence in the environment. The nonchelating residues can provide specific recognition patterns to ensure exclusive uptake of the Fe-siderophore by the producer. In contrast, the ferrous iron generated by photolytic cleavage of the compound is much harder to monopolize in complex microbial communities. Identifying the recipient of ferrous iron fluxes generated by the photolysis of cyanobacterial photoreactive siderophores will require innovative approaches to cyanobacterial cultivation methods, most likely along with other members of their natural microbial communities.

## MATERIALS AND METHODS

**Bioinformatic analysis and genome mining.** Gene clusters from *Rivularia*, *Leptolyngbya*, and *Rubidibacter* (CP003549, AP017310, and ASSJ01000034.1, respectively) (for nucleotide positions see Table S1 to S3) were selected from the previous study (13) and interpreted by identification of the closest known functional homologue in the MIBiG repository 2.0 (21) in combination with antiSMASH 5.0.0 (22) and functional analysis of individual domains of multidomain proteins by CDD (23). Promoter regions responsive to FUR (ferric uptake regulator) were identified by sequence comparison to a weight matrix (15).

FAAL domains from proteins AFY58527, BAU16016, and ERN42091, found to be encoded at the beginning of the candidate gene clusters, were aligned by MAFFT (24) v. 7.308 (as provided via Geneious 10.0.9 MAFFT plugin 1.3.6) and used as a template for PSSM creation and subsequent PSI-BLAST (25), a search taxonomically limited to cyanobacteria. The search from 27 August 2019 with bitscore cutoff set to 250 yielded 943 hits. Regions spanning 40 kbp upstream and 80 kbp downstream from the start of the hit-encoding gene were obtained and inspected for the presence of two  $\beta$ -OH-Asp-specific NRPS modules. A module was evaluated as  $\beta$ -OH-Asp specific if it included an A-domain specific for Asp/Asn and a C-domain aligning to PSSM of  $\beta$ -OH-Asp-specific C-domains constructed from *Rivularia*, *Leptolyngbya*, and *Rubidibacter* with a bitscore of  $>300$ . Clusters encoding two such modules separated by a single NRPS module, along with a FAAL domain, from the first PSI-BLAST search were considered cyanochelin-like clusters. Cluster surroundings were inspected for the presence of siderophore transport-related genes by antiSMASH and by identification of the closest known functional homologue.

**Phylogenomic analysis.** To assess the phylogenetic distribution of cyanobacterial strains containing cyanochelin-like gene clusters, a phylogenomic tree was constructed utilizing the *de novo* workflow available in the Genome Taxonomy Database toolkit (26, 27). The GTDB-Tk v0.3.2 release from July 2019, offering 600 quality-checked cyanobacterial genomes, was used to produce an approximately maximum-likelihood tree (FastTree2-v2.1.9) (28) based on 120 concatenated conserved bacterial markers, and rooted by the closely related outgroup *Vampirovibrio chlorellavorus*.

**Strains and cultivation.** *Rivularia* sp. strain PCC 7116 and *Leptolyngbya* sp. strain NIES-3755 were obtained from respective culture collections. *Rubidibacter lacunae* strain KORDI 51-2 was kindly provided by Dong Han Choi (Korea Institute of Ocean Science and Technology, Republic of Korea). Strains were grown in BG-11 medium, with *Rivularia* and *Rubidibacter* requiring an addition of Turks Islands Salts (TIS) solution (NaCl 112 g/liter, KCl 2.68 g/liter, MgSO<sub>4</sub>·7 H<sub>2</sub>O 27.7 g, CaCl<sub>2</sub>·2 H<sub>2</sub>O 5.8 g/liter) at a ratio of 3:1. All strains were grown under ambient temperature (21  $\pm$  2°C) and continuous dispersed light.

**Isolation of cyanochelins.** Freeze-dried iron-starved biomass of *Rivularia* sp. PCC 7116 was extracted using 60% acetonitrile (ACN) three times in final biomass-to-solvent ratio of 10 g/liter. The crude extract was subjected to liquid-liquid partitioning using an equal volume of ethyl acetate in a separatory funnel. The retained water phase containing the compound was acidified using formic acid to pH 3.5 and subjected to a second liquid-liquid partitioning step using four volumes of ethyl acetate/ACN (in ratio 3:5). The drop in pH caused the compound to prefer the organic phase, which was subsequently dried, dissolved in dimethyl sulfoxide (DMSO), and subjected to preparative reverse phase HPLC on a phenyl column ( $\mu$ Bondapak Phenyl, 10  $\mu$ m, 125 Å, 7.8  $\times$  300 mm). Water (A) and ACN (B), both acidified using formic acid to pH 2, were used as solvents in the following gradient: 0 min 15% B, 1 min 15% B, 8 min 37% B, 31 min 54% B, 35 min 100% B, 40 min 100% B at flow rate 2.5 ml/min. The diode array detector (DAD) was set to 254 nm and used for detection of the target compounds. Cyanochelin A eluted at 30 min and its variant 753.33 eluted at 25 min. The isolation procedure yielded 2.2 mg of cyanochelin A from *Rivularia* sp. PCC 7116; the variant 753.33 was not collected in a sufficient amount that it could be weighed. Analogously, cyanochelins from *Leptolyngbya* and *Rubidibacter* were isolated for iron chelation experiments using the identical semipreparative column, flow rate, and detection settings as described above. For *Leptolyngbya*, the following gradient was applied. (A) water/(B) ACN (both acidified using 0.1% formic acid): 0 min 15% B, 8 min 48% B, 21 min 56% B, 26 min 100% B, 31 min 100%, and 35 min 15%. The compound of interest (*m/z* 1,026.51) eluted at 24.2 min. For *Rubidibacter*, the following gradient was used. (A) 25% ACN/(B) 96% ACN (both acidified using trifluoroacetic acid): 0 min 0% B, 1 min

0% B, 6 min 30% B, 46 min 100% B, 51 min 100% B, 52 min 0%. The compound of interest ( $m/z$  1,045.55) eluted at 24.5 min. The compounds from *Leptolyngbya* and *Rubidibacter* were collected in insufficient amounts to be weighed. The purity of the collected fractions was confirmed by HPLC-HRMS/MS analysis by the method described below.

**Structural characterization of cyanochelin A.** NMR experiments were performed at 25°C on a Bruker AvanceNeo 700 MHz spectrometer (Billerica, MA, US) equipped with a triple resonance CHN cryoprobe, using CD<sub>3</sub>OD or *D*-DMSO (Sigma-Aldrich, Milan, Italy) as solvents and the 1D and 2D standard pulse sequences provided by the manufacturer. <sup>1</sup>H chemical shifts were referenced to the residual solvents' protons resonating at 3.31 (CHD<sub>2</sub>OD) or 2.50 (CHD<sub>2</sub>SOCD<sub>2</sub>) ppm. <sup>13</sup>C-NMR chemical shifts were referenced to the solvents' methyl carbons resonating at 49.01 (CD<sub>3</sub>OD) or 39.51 ppm (*D*-DMSO). Abbreviations for signal couplings are as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet, b = broad.

**Iron depletion experiment.** *Rivularia* biomass homogenized by pipetting was distributed equally into three flasks of each standard or iron-deprived medium supplemented with 10 μM desferrioxamine B. Expression of the candidate gene cluster along with content of cyanochelins in the medium and biomass was monitored at six time points of 0, 2, 4, 7, 9, and 11 days. Solid phase extractions (SPE) of supernatants were performed on a Supelco Discovery DSC-18 cartridge (500 mg, 3 ml). The spent media samples (5 ml) were loaded on preconditioned SPE cartridges (two volumes of methanol and one volume of water) and eluted to 1 ml of 100% methanol. Samples of biomass were spun at 1,600 × *g* for 5 min to remove residual supernatant and either supplied with 1 ml of PGTX and frozen at -80°C for RNA extraction, or frozen, freeze-dried, weighted, and extracted using 60% ACN in a biomass-to-solvent ratio of 4 mg/ml for HPLC-HRMS/MS analysis (see below). RNA was isolated using Direct-zol RNA Miniprep (Zymo Research) and genomic DNA was degraded with the TURBO DNA-free kit (Life Technologies) according to the manufacturer's protocol. cDNA was generated from 500 ng of total RNA using Transcriptor reverse transcriptase (Roche). A specific primer pair was designed to amplify *ccsA* (q\_AFY58527-R) (CCAAGCTA TAGTCGCTCTAA) and -F (ATACGGTTGGATTCAGTGAT), while a primer pair for *rnpB*-F (CTGCTGGTGC GCTCTTACC) and *rnpB*-R (GTGAGGATAGTGCCACAGAA) was used to amplify a housekeeping reference gene (*rnpB*). Reverse transcriptase quantitative PCR (RT-qPCR) was performed using LightCycler 96 (Roche) and FastStart Essential DNA Green Master (Roche). A 20-μl reaction contained 10 μl FastStart Essential DNA Green Master (Roche), 500 nM primers, and 1 μl of cDNA. qPCR cycling parameters included preincubation at 1 min for 95°C, 40 times 95°C for 10 s, 60°C for 10 s, and 72°C for 20 s, melting at 95°C for 10 s, 65°C for 60 s, 97°C for 1 s, with a final cooling step to 37°C. The relative gene expression of *ccsA* was normalized to *rnpB* and expressed as an average of three replicates.

**HPLC-HRMS/MS.** The biomass samples in batch cultures, as well as those of media and biomass obtained from iron-deprivation experiments, were analyzed using the HPLC system (Dionex Ultimate 3000) connected to HRMS with electrospray ionization (Bruker Impact HD II). The separation was performed on the reverse-phase phenyl column (μBondapak Phenyl, 10 μm, 125 Å, 7.8 × 300 mm) using an ACN/water gradient (A/B) of 0 min 15/85%, 1 min 15/85%, 25 min 100/0%, 30 min 100/0%, and 33 min 15/85% at a flow rate of 0.6 ml/min. To ensure that the vast majority of the compounds were present in the nonchelated form, the pH of the mobile phase was kept below 3.0 by the addition of trifluoroacetic acid into the mobile phases (0.1% of TFA in water and 0.002% of TFA in ACN). The following settings for the mass spectrometer were used: dry temperature 200°C; drying gas flow 12 liters/min; nebulizer 3 bar; capillary voltage 4500 V; endplate offset 500 V. The spectra were collected in the range of 100 to 1,500  $m/z$  with a spectral rate of 0.5 Hz. Collision-induced dissociation (CID) was set as a ramp from 20 to 80 eV on masses of 200 to 1,500, respectively, using the automated precursor ion selection. Calibration of the instrument was performed using CH<sub>3</sub>COONa clusters at the beginning of each analysis. For analysis where elemental composition analysis was required, lock mass ( $m/z$  622) was added to the ion source. The individual spectra were analyzed using Bruker Data Analysis software (version 5.2) to obtain elemental formulas. Manual direct injections were applied to obtain high-quality fragmentation spectra, as well as to test the cyanochelin capability to bind iron. Cyanochelin A (25 μg/ml) in 50% MeOH/water (vol/vol) with 0.5% formic acid and lockmass (0.2 μg/ml) was injected to the ionization source using an automated syringe at a flow rate of 1,200 μl/h. To test whether it would bind Fe<sup>3+</sup>, the solution was supplied with approximately 1.5 to 2 molar equivalents of FeCl<sub>3</sub>. The following settings of the mass spectrometer were used: dry temperature 200°C; drying gas flow 3 liters/min; nebulizer 0.5 bar; capillary voltage 4500 V; endplate offset 500 V. The compound of interest was isolated in quadrupole using an isolation window of 10 Da and fragmented at different fragmentation energies (20 to 80 eV). Analogously, cyanochelins from *Leptolyngbya* and *Rubidibacter* were analyzed with the compound concentration roughly estimated based on the comparison to the HRMS signal of the cyanochelin A.

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1**, PDF file, 1.9 MB.

## ACKNOWLEDGMENTS

The study was supported by the Czech Science Foundation projects "Bioactive cyanobacterial lipopeptides: genome mining, detection, and structure-activity relationships" (number 16-09381S) and "Co-existence of quorum sensing and quorum sensing inhibition in cyanobacteria": Chemical warfare or cooperation? (number 19-17868Y), the Ministry of

Education, Youth and Sports of the Czech Republic, National Program of Sustainability I, number LO1416 and RVO 67985939, and by the Ministry of Regional Development of the Czech Republic—Cross-Border cooperation Czech-Bavaria project no. 41. N.B., A.C., G.E., and V.C. were supported by Regione Campania, PO FESR2014 to 2020, O.S. 1.2, Project “Campania Oncoterapie” no. B61G18000470007.

We declare no competing interests.

## REFERENCES

1. Sandy M, Butler A. 2009. Microbial iron acquisition: marine and terrestrial siderophores. *Chem Rev* 109:4580–4595. <https://doi.org/10.1021/cr9002787>.
2. Årstøl E, Hohmann-Marriott MF. 2019. Cyanobacterial siderophores—physiology, structure, biosynthesis, and applications. *Mar Drugs* 17:281. <https://doi.org/10.3390/md17050281>.
3. Morrissey J, Bowler C. 2012. Iron utilization in marine cyanobacteria and eukaryotic algae. *Front Microbiol* 3:43. <https://doi.org/10.3389/fmicb.2012.00043>.
4. Shih P, Wu D, Latifi A, Axen S, Fewer D, Talla E, Calteau A, Cai F, de Marsac N, Rippka R, Herdman M, Sivonen K, Coursin T, Laurent T, Goodwin L, Nolan M, Davenport K, Han C, Rubin E, Eisen J, Woyke T, Gugger M, Kerfeld C. 2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci U S A* 110:1053–1058. <https://doi.org/10.1073/pnas.1217107110>.
5. Gross H. 2007. Strategies to unravel the function of orphan biosynthesis pathways: recent examples and future prospects. *Appl Microbiol Biotechnol* 75:267–277. <https://doi.org/10.1007/s00253-007-0900-5>.
6. Butler A, Theisen RM. 2010. Iron(III)-siderophore coordination chemistry: reactivity of marine siderophores. *Coord Chem Rev* 254:288–296. <https://doi.org/10.1016/j.ccr.2009.09.010>.
7. Kreuzer MF, Kage H, Nett M. 2012. Structure and biosynthetic assembly of cupriachelin, a photoreactive siderophore from the bioplastic producer *Cupriavidus necator* H16. *J Am Chem Soc* 134:5415–5422. <https://doi.org/10.1021/ja300620z>.
8. Rosconi F, Davyt D, Martínez V, Martínez M, Abin-Carriquiry JA, Zane H, Butler A, de Souza EM, Fabiano E. 2013. Identification and structural characterization of serobactins, a suite of lipopeptide siderophores produced by the grass endophyte *Herbaspirillum seropedicae*. *Environ Microbiol* 15:916–927. <https://doi.org/10.1111/1462-2920.12075>.
9. Hardy CD, Butler A. 2019. Ambiguity of NRPS structure predictions: four bidentate chelating groups in the siderophore pacifibactin. *J Nat Prod* 82:990–997. <https://doi.org/10.1021/acs.jnatprod.8b01073>.
10. Singh GM, Fortin PD, Koglin A, Walsh CT. 2008. beta-Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: characterization of two candidate hydroxylases AspH and SyrP in *Pseudomonas syringae*. *Biochemistry* 47:11310–11320. <https://doi.org/10.1021/bi801322z>.
11. Boiteau RM, Mende DR, Hawco NJ, McIlvin MR, Fitzsimmons JN, Saito MA, Sedwick PN, DeLong EF, Repeta DJ. 2016. Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean. *Proc Natl Acad Sci U S A* 113:14237–14242. <https://doi.org/10.1073/pnas.1608594113>.
12. Mareš J, Hájek J, Urajová P, Kust A, Jokela J, Saurav K, Galica T, Čapková K, Mattila A, Haapaniemi E, Permi P, Mylsterud I, Skulberg OM, Karlsen J, Fewer DP, Sivonen K, Tonnesen HH, Hrouzek P. 2019. Alternative biosynthetic starter units enhance the structural diversity of cyanobacterial lipopeptides. *Appl Environ Microbiol* 85:e02675-18. <https://doi.org/10.1128/AEM.02675-18>.
13. Galica T, Hrouzek P, Mareš J. 2017. Genome mining reveals high incidence of putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP ligase genes in biofilm-forming cyanobacteria. *J Phycol* 53:985–998. <https://doi.org/10.1111/jpy.12555>.
14. Manavalan B, Murugapiran SK, Lee G, Choi S. 2010. Molecular modeling of the reductase domain to elucidate the reaction mechanism of reduction of peptidyl thioester into its corresponding alcohol in non-ribosomal peptide synthetases. *BMC Struct Biol* 10:1. <https://doi.org/10.1186/1472-6807-10-1>.
15. González A, Angarica VE, Sancho J, Fillat MF. 2014. The FurA regulon in *Anabaena* sp. PCC 7120: in silico prediction and experimental validation of novel target genes. *Nucleic Acids Res* 42:4833–4846. <https://doi.org/10.1093/nar/gku123>.
16. Carroll CS, Moore MM. 2018. Ironing out siderophore biosynthesis: a review of non-ribosomal peptide synthetase (NRPS)-independent siderophore synthetases. *Crit Rev Biochem Mol Biol* 53:356–381. <https://doi.org/10.1080/10409238.2018.1476449>.
17. Reid RT, Live DH, Faulkner DJ, Butler A. 1993. A siderophore from a marine bacterium with an exceptional ferric ion affinity constant. *Nature* 366:455–458. <https://doi.org/10.1038/366455a0>.
18. Sriharan M. 2016. Iron homeostasis in *Mycobacterium tuberculosis*: mechanistic insights into siderophore-mediated iron uptake. *J Bacteriol* 198:2399–2409. <https://doi.org/10.1128/JB.00359-16>.
19. Kem MP, Butler A. 2015. Acyl peptidic siderophores: structures, biosyntheses and post-assembly modifications. *Biometals* 28:445–459. <https://doi.org/10.1007/s10534-015-9827-y>.
20. Barbeau K, Rue EL, Bruland KW, Butler A. 2001. Photochemical cycling of iron in the surface ocean mediated by microbial iron(III)-binding ligands. *Nature* 413:409–413. <https://doi.org/10.1038/35096545>.
21. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Maschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, et al. 2015. Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 11:625–631. <https://doi.org/10.1038/nchembio.1890>.
22. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>.
23. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43:D222–D226. <https://doi.org/10.1093/nar/gku1221>.
24. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
25. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.
26. Parks DH, Chuvpochina M, Waite DW, Rinke C, Scarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>.
27. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
28. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. <https://doi.org/10.1371/journal.pone.0009490>.

## 4.5. Paper V. (manuscript)

# A newly discovered NRPS/PKS pathway of the cyanobacterial lipopeptide muscotoxin utilizes 4-methylproline generated by the nostopeptolide biosynthetic machinery

Dr. Jan Mareš<sup>[a,b,c]</sup>, Dominik Chmelík<sup>[b,d,e]</sup>, Tomáš Galica<sup>[b,c]</sup>, Dr. Petra Urajová<sup>[b]</sup>, Jan Hájek<sup>[b,c]</sup>, Dr. Klára Řeháková<sup>[a]</sup> and Dr. Pavel Hrouzek<sup>\*[b,c]</sup>

- [a] Dr. J. Mareš, Dr. K. Řeháková  
Institute of Hydrobiology  
Biology Centre of the Czech Academy of Sciences  
Na Sádkách 702/7, 37005 České Budějovice, Czech Republic
- [b] Dr. J. Mareš, D. Chmelík, T. Galica, Dr. P. Urajová, J. Hájek, Dr. P. Hrouzek  
Center Algatex  
Institute of Microbiology, Czech Academy of Sciences  
Novohradská 237 – Opatovický mlýn, 379 01 Třeboň, Czech Republic  
E-mail: [hrouzekp@gmail.com](mailto:hrouzekp@gmail.com)
- [c] Dr. J. Mareš, T. Galica, J. Hájek, Dr. Pavel Hrouzek  
Faculty of Science  
University of South Bohemia  
Branišovská 1760, 37005 České Budějovice, Czech Republic
- [d] D. Chmelík  
Department of Adaptive Biotechnologies  
Global Change Research Institute, Czech Academy of Sciences  
Bélidla 986/4a, 603 00 Brno, Czech Republic
- [e] D. Chmelík  
Department of Plant Physiology  
Faculty of Science, Masaryk University  
Kotlářská 267/2, 611 37 Brno, Czech Republic

Supporting information for this article is given via a link at the end of the document.

**Abstract:** An enormous pool of bioactive cyanobacterial peptides are encoded in modular biosynthetic gene clusters (BGCs). Here we report the identification of BGC for cytotoxic and antifungal lipopeptide muscotoxin along with a nostopeptolide BGC in *Desmonostoc muscorum* CCALA 125 genome. Muscotoxin BGC consists of three NRPS/PKS genes (*mcxA–C*) homologous to nostophycin BGC but surprisingly lacking a loading adenylation domain. Detected products of both encoded BGCs, nostopeptolide 1111 and muscotoxin B, contain the rare amino-acid 4-MePro in their structures. Feeding by isotopically substituted Leu demonstrated that 4-MePro is provided to M<sub>cx</sub>B–C as a product of nostopeptolide biosynthetic enzymes (NosE–F) responsible for transforming Leu into 4-MePro. Despite being an intermediate of the nostopeptolide pathway, 4-MePro was incorporated into muscotoxin B at a comparative rate. After pseudoaeruginosin-spumigin, the biosynthetic coupling of muscotoxins and nostopeptolide is another example of 4-MePro-mediated metabolic crosstalk, suggesting a more widespread adaptive mechanism to extend the repertoire of bioactive compounds.

## Introduction

Cyanobacterial lipopeptides have recently been in the scope of experimental studies due to their substantial bioactive properties<sup>[1–5]</sup>. Genome mining has proven the deduced

lipopeptide biosynthetic gene clusters to be widely distributed in the cyanobacterial tree of life<sup>[6,7]</sup>, offering a new source of antibiotics leads but also potential toxins. Detailed analyses of the biosynthetic gene clusters and partial reconstruction of the biosynthetic steps have been provided for puwainaphycins/ minutissamides, hassallidins, and anabaenolysins<sup>[7–10]</sup>. In general, the peptidic cores of these molecules are assembled via non-ribosomal peptide synthetases (NRPS) following the collinearity rule<sup>[11]</sup>, and modified by accompanying tailoring enzymes. However the first step of the biosynthesis usually involves activation of a fatty acid (FA) and its subsequent loading onto the biosynthetic assembly line<sup>[10]</sup>. In cyanobacteria, the activation of FA can be performed by fatty acyl-AMP ligases (FAALs) as documented in puwainaphycins/minutissamides<sup>[10]</sup> or fatty acyl-CoA ligases (FACLs) as predicted in hassallidins<sup>[7]</sup>, similar to other lipopeptide-producing groups of bacteria<sup>[12–13]</sup>. Interestingly, no FA-activating enzyme was found within the anabaenolysin biosynthetic gene cluster<sup>[9]</sup>. In cyclic lipopeptides containing the  $\beta$ -amino FA residue, the cyclization step involves formation of a peptide bond between the carboxyl of the last incorporated amino acid and the  $\beta$ -amino group of the FA residue, catalyzed by a terminal NRPS thioesterase domain<sup>[8]</sup>. The pre-requisite for this condensation reaction is previous elongation and amination of the FA intermediate by a specific polyketide synthase (PKS) module containing an aminotransferase domain<sup>[8]</sup>.

A particular group of cyanobacterial lipopeptides, investigated in this study, comprises cyclic  $\beta$ -amino undeca-lipopeptides muscotoxins<sup>[3]</sup>. Muscotoxins exhibit a notable cytotoxic and antifungal bioactivity based on reduction of the membrane fluidity and subsequent membrane permeabilization of the target cells<sup>[3,14]</sup>. While the biosynthetic gene cluster for muscotoxins is unknown, the general structural similarity of muscotoxins to cyclic  $\beta$ -amino FA lipopeptides puwainaphycins and minutissamides suggests that a similar biosynthetic mechanism may be involved. Intriguingly, all three previously reported variants of muscotoxin (A-C) contain an unsubstituted FA side chain of constant length and are distinguished by substitution of amino acids at position 4 (L-Pro/4-MePro) and 6 (D-*allo*-Ile/Val)<sup>[14]</sup>. In contrast, cyanobacterial strains producing puwainaphycins/minutissamides accumulate a broad variety of chemical analogues spanning over a range of FA chain lengths and substitutions by hydroxy-, oxo-, chloro-, and acetyl moieties<sup>[10]</sup>. A recently proposed explanation of the large variability observed in puwainaphycin/minutissamide products emphasized the role of two promiscuous FAAL starter units<sup>[10]</sup>. Based on these findings one could hypothesize that the biosynthetic assembly line of muscotoxin involves a different FA-activating mechanism exhibiting a strict substrate specificity. Another intriguing trait of one of the muscotoxin variants (muscotoxin B) is the presence of 4-MePro. This rare non-proteinogenic amino acid serves as a structural signature for several classes of bioactive cyanobacterial metabolites such as nostopeptolides, nostoweipeptins, spumigins, and pseudoaeruginosins<sup>[15,16]</sup>. The biosynthesis of L-4-MePro from L-Ile requires two subsequent enzymatic reactions catalyzed by NosE (alcohol dehydrogenase) and NosF (3-methyl- $\Delta$ 1-pyrroline-5-carboxylic acid reductase) as evidenced experimentally in nostopeptolides<sup>[17]</sup>. Interestingly, 4-MePro-containing protease inhibitors pseudoaeruginosins seem to be produced as a result of biosynthetic coupling of the aeruginosin pathway with spumigin pathway through *nosE-F* homologs *spuD-E*<sup>[16]</sup>. Part of the 4-MePro synthesized within the spumigin pathway is supplied as an alternative substrate to the adenylation domain of the aeruginosin NRPS protein AerG, otherwise specifically activating Pro.

Here we focus on identification and bioinformatic interpretation of the biosynthetic gene cluster of muscotoxins in the originally reported strain *Desmonostoc muscorum* CCALA 125. We show its similarity to a previously reported nostophycin pathway<sup>[18]</sup> except for an unexpected absence of the starter adenylation domain. We further experimentally demonstrate a biosynthetic crosstalk of the muscotoxin pathway with nostopeptolide pathway encoded in the same genome, resulting in the formation of two muscotoxin variants containing 4-MePro.

## Results and Discussion

### Main structural traits of muscotoxins

In our previous studies we have elucidated the structure of three muscotoxin variants – muscotoxins A, B, and C<sup>[3,14]</sup> produced by the strain *D. muscorum* CCALA 125. These lipopeptides involve

a cycle composed of ten amino acid residues enclosed by the 2,5-dihydroxy-3-amino-dodecanoic acid (5-OH Ahdoa). The resulting general sequence 5 OH-Ahdoa<sup>1</sup>-Gln<sup>2</sup>-Gly<sup>3</sup>-X<sup>4</sup>-Phe<sup>5</sup>-X<sup>6</sup>-Ser<sup>7</sup>-Dhb<sup>8</sup>-Ser<sup>9</sup>-Ile<sup>10</sup>-Pro<sup>11</sup> is variable at positions 4 and 6. The major variant muscotoxin A possesses a Pro residue at position 4 and Ile at position 6<sup>[3]</sup>. Muscotoxin B share the same sequence, however, the Pro residue at position 4 is replaced by 4-MePro<sup>[3]</sup>. The minor variant, muscotoxin C contains Pro at position 4 but possesses a Val instead of Ile residue at position 6<sup>[14]</sup>. After detailed HPLC-HRMS/MS analysis of the *D. muscorum* CCALA 125 extracts in the current study we have detected an additional minor muscotoxin variant, an isobaric compound to muscotoxin B (*m/z* 1225), with a switched position of the 4-Me-Pro residue, possessing Pro at position 4 and 4-Me-Pro at position 11 (Tables S1, S2). As these, together with the previously reported, were the only muscotoxin variants detected in the *D. muscorum* CCALA 125 extracts, the modified fatty acyl moiety of 5-OH Ahdoa was consistently present in all the detected structures.

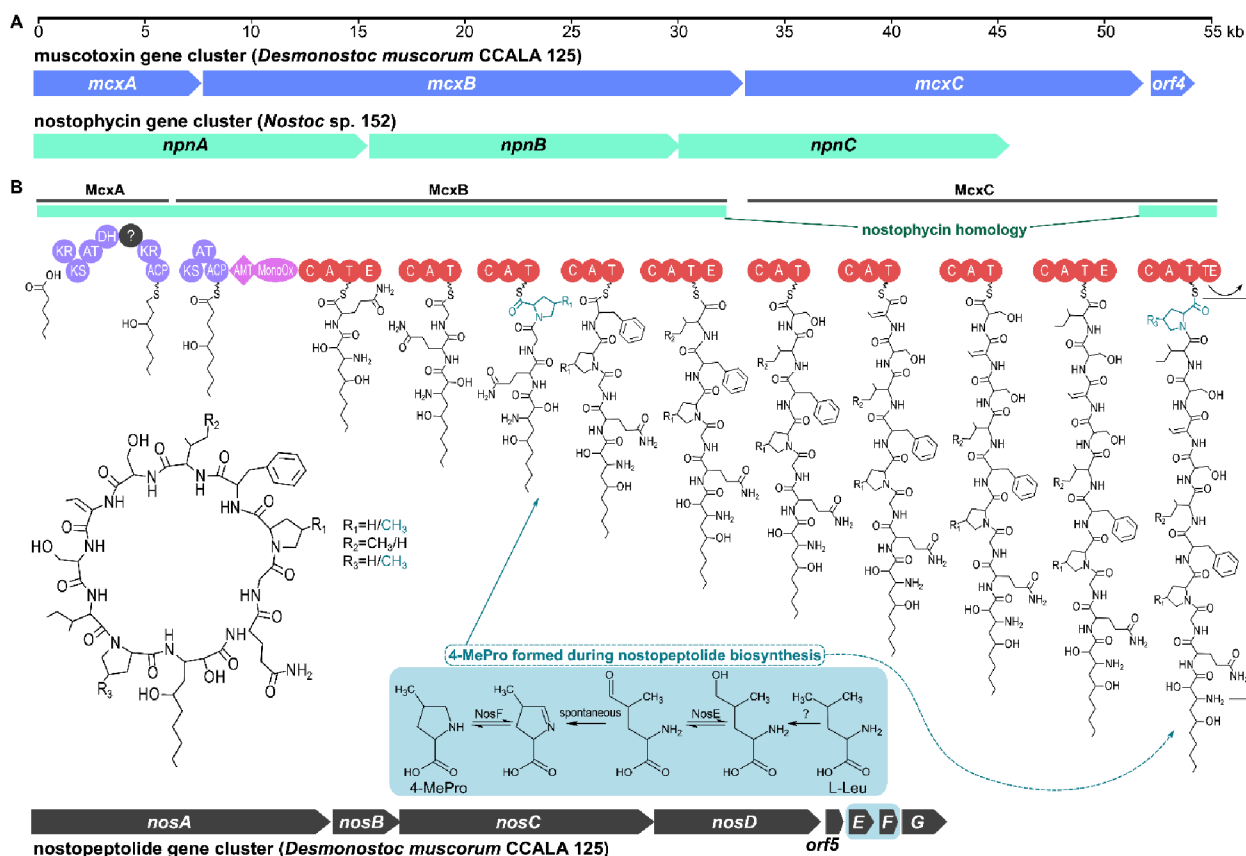
The occurrence of 4-MePro in muscotoxins represents a unique feature among cyanobacterial  $\beta$ -amino FA lipopeptides. Furthermore the uniformity of the FA side chain is a structural trait strikingly different from puwainaphycin/minutissamide lipopeptides, which show extensive variability in this part of the molecule<sup>[10,25]</sup>. Therefore our next goal was to identify the biosynthetic gene cluster for muscotoxin and to compare the encoded enzymes to the known ones involved in lipopeptide and 4-MePro biosynthesis.

### Muscotoxin biosynthetic gene cluster

The whole genome assembly of *D. muscorum* CCALA 125 (NCBI accession JACERF000000000) resulted in 151 genomic scaffolds of a 8.98 Mbp total length (41.7% GC content, 99.3% estimated completeness). A putative muscotoxin biosynthetic gene cluster (Figure 1, NCBI accession MT783953) was identified in the middle of a genomic scaffold obtained by combining the Illumina and Sanger sequencing data. The gene cluster consisted of four predicted genes and spanned approximately 53 kbp (Table 1). The core biosynthetic machinery was encoded in first three of them, *mcxA* – a PKS gene, *mcxB* – a hybrid PKS/NRPS gene, and *mcxC* – a NRPS gene, followed by a predicted ABC-transporter gene (*orf4*). *McxA-B* and the terminal part of *mcxC* exhibited sequence homology (Table 1) to an already known biosynthetic gene cluster of the cyanobacterial  $\beta$ -amino cyclic heptapeptide nostophycin<sup>[18]</sup> (Figure 1). Annotation of the PKS/NRPS modules further confirmed functional homology and synteny of the encoded enzymatic domains between muscotoxin and nostophycin, which results in the structural homology of muscotoxin A and nostophycin that share D-Gln<sup>2</sup>, L-Gly<sup>3</sup>, L-Pro<sup>4</sup>, L-Phe<sup>5</sup>, D-*allo*-Ile<sup>6</sup>, and the terminal proline residue (L-Pro<sup>11</sup> in muscotoxin). The amino acid residues at positions 7-10 are unique for muscotoxin and are presumably incorporated by a sequence of four additional NRPS modules present in *mcxC* (Figure 1, Table 2).

In the nostophycin gene cluster, the biosynthetic loading module of NpnA starts with an adenylation domain<sup>[18]</sup>. At the beginning of *McxA* we detected only an approximately 70 aa long A-domain rudiment (a C-terminal subdomain fragment) based on CDD search. While it is highly probable that both pathways have shared a common ancestor with a functional starter A-domain, the short fragment in *McxA* is very likely no longer active. The subsequent ketoreductase domain (Figure 1) could theoretically secure recognition of a correct substrate<sup>[26]</sup> however the mechanism of FA activation and loading is obscure.





**Figure 1.** Biosynthetic gene clusters of muscotoxin and nostopeptolide in *Desmonostoc muscorum* CCALA 125. A) Size and organization of the muscotoxin gene cluster in comparison to the homologous nostophycin gene cluster from *Nostoc* sp. 152. B) Proposed biosynthetic scheme of muscotoxins, and their biosynthetic crosstalk with the nostopeptolide gene cluster through the 4-MePro intermediate. C) Size and organization of the nostopeptolide gene cluster in *D. muscorum* CCALA 125. The biosynthesis of 4-MePro is presented as described previously<sup>17</sup>. A – adenylation domain; ACP – acyl-carrier protein; AmT – aminotransferase; AT – acyltransferase; C – condensation domain; DH – dehydratase; E – epimerase; KR – ketoreductase; KS – ketosynthase; MonoOx – monooxygenase; T – thiolation domain; TE – thioesterase.

**Table 1.** Summary of deduced proteins in the muscotoxin biosynthetic pathway and their closest homologues (top BLAST hits with known function – Feb 2020).

Muscotoxin Pathway					Closest Homologue	
Protein	Size [aa]	Proposed Function	Accession	Organism	Identity [%]	Function
MusA	2579	PKS	AEU11005.1	<i>Nostoc</i> sp. 152	66	PKS - NpnA (nostophycin)
MusB	8275	hybrid PKS/NRPS	AEU11006.1	<i>Nostoc</i> sp. 152	82	hybrid PKS/NRPS - NpnB (nostophycin)
MusC	6106	NRPS	AEU11003.1	<i>Nostoc</i> sp. 152	64	NRPS - NpnC (nostophycin)
ORF4	676	ABC transporter	AVK43371.1	<i>Nostoc</i> sp. HIID-D1B	78	ABC transporter - AptE (anabaenopeptin)

The situation is analogous to that previously described for anabaenolysin producers, in which the *abl* gene clusters did not contain FA-activating enzymes although a FAAL gene was identified at distant genome locations<sup>9</sup>. Also in the genome of *D. muscorum* CCALA 125 a FAAL gene associated with an unknown NRPS gene cluster is located on a different genomic scaffold. Utilization of activated FAs provided in trans have been previously described in surfactin biosynthesis in *Bacillus subtilis*, in this case supplied by FAALs from primary metabolism<sup>13</sup>.

Further steps catalyzed by the McxA PKS are non-canonical and require future experimental work. In our biosynthesis prediction scheme (Figure 1), we speculate that hexanoic acid could serve as the initial biosynthetic substrate. We hypothesize that its

carbonyl group is reduced to a hydroxyl group by the first ketoreductase domain prior to elongation by the subsequent ketoacyl synthase module. The PKS module of McxA contains an acyltransferase, a dehydratase, a ketoreductase, and an acyl-carrier domain. Inbetween the dehydratase and ketoreductase domains, there is a 366 aa long region, which does not show significant similarity to any known catalytic domain. This arrangement of domains again bears resemblance to the one found in the nostophycin gene cluster<sup>19</sup>, which however contains an enoyl-reductase domain at the position corresponding to the 366 aa long unannotated region of McxA. The following protein McxB contains a canonical PKS module, an aminotransferase domain, and a monooxygenase domain. These are altogether responsible for elongation and modification

of the nascent chain into a 3-amino-2-hydroxy intermediate. This particular modification allows beta-cyclization of the final product

as previously described in detail for nostophycin and puwainaphycins/minutissamides<sup>[8,10,18]</sup>.

**Table 2.** Summary of adenylation domains and their substrates involved in muscotoxin biosynthesis (BLAST performed in Feb 2020).

Protein	Domain (aa range)	Closest annotated BLAST hit (aa range): activated substrate	Identity [%]	Predicted substrate (antiSMASH)	Incorporated aa residue in muscotoxin
MusB	A1 (2497-2896)	NpnB1, <i>Nostoc</i> sp. 152, AEU11006.1 (2644-3043): D-Gln	87	D-Gln	D-Gln
	A2 (4023-4421)	NpnB2, <i>Nostoc</i> sp. 152, AEU11006.1 (4177-4570): L-Gly	87	L-Gly	L-Gly
	A3 (5083-5498)	NpnC4, <i>Nostoc</i> sp. 152, AEU11003.1 (4196-4601): L-Pro	81	L-Pro	L-Pro/ 4-MePro
	A4 (6160-6556)	NpnC2, <i>Nostoc</i> sp. 152, AEU11003.1 (1611-1999): L-Phe	86	L-Phe	L-Phe
	A5 (7218-7622)	NosA1, <i>Nostoc</i> sp. GSV224, AAF15891.2 (485-898): L-Ile/Leu/Val	83	D-Ile	D- <i>allo</i> -Ile/ D-Val
MusC	A1 (476-880)	NosA2, <i>Nostoc</i> sp. GSV224, AAF15891.2 (1569-1987): L-Ser	77	L-Ser	L-Ser
	A2 (1542-1969)	AdpB3, <i>Anabaena</i> sp. 90, CAC01604.1 (2629-3035): L-Thr	79	L-Thr	L-Thr
	A3 (2637-3039)	NosA2, <i>Nostoc</i> sp. GSV224, AAF15891.2 (1569-1987): L-Ser	75	L-Ser	L-Ser
	A4 (3701-4105)	NosA1, <i>Nostoc</i> sp. GSV224, AAF15891.2 (485-898): L-Ile/Leu/Val	82	D-Ile	D- <i>allo</i> -Ile
	A5 (5237-5652)	NpnC4, <i>Nostoc</i> sp. 152, AEU11003.1 (4196-4601): L-Pro	80	L-Pro	L-Pro/ 4-MePro

#### 4-MePro genes and detection of nostopeptolide 1111

While the NRPS part of the muscotoxin pathway was generally straightforward, the gene cluster did not contain homologues of alcohol dehydrogenase and 3-methyl- $\Delta$ 1-pyrroline-5-carboxylic acid reductase genes required for the biosynthesis of 4-MePro<sup>[17]</sup>. However bioinformatic search in the *D. muscorum* CCALA 125 genome revealed their presence as part of a complete nostopeptolide biosynthetic gene cluster (NCBI accession MT783959). This gene cluster was approximately 41 kbp long and consisted of eight ORFs encoding proteins homologous to NosA-G and ORF5 (Figure 1, Table S3) matching those reported previously<sup>[27]</sup>.

HPLC-HRMS/MS analysis of the *D. muscorum* CCALA 125 extract has revealed that the strain is capable to produce a variant of a cyclic peptide nostopeptolide. The compound was detected at *m/z* 1111.6 (Figure 2A). HRMS/MS experiments demonstrated formation of the fragment at *m/z* 928.47 corresponding to loss of butyrate and isoleucine residue (Figure 2B), an exocyclic moiety typical for nostopeptolides<sup>[28,29]</sup>. A series of b-y fragment ions led subsequently to partial characterization of amino acid sequence of the cycle to be Ser-Pro-Tyr-Asn-Ser-Leu/Ile (Figure 2B). The second fragmentation pathway is initiated by the loss of water molecule and followed by losses of dehydrated serine residue (dSer), 4-Me-Pro and acetyl leucine residue (LeuAc). Both types of fragments were previously reported for nostopeptolide A as well as its analog nostopeptolide 1052<sup>[29]</sup>. The presence of fragments [But+Ile+dSer+MePro+LeuAc]<sup>+</sup> and [Pro+Ser+MePro+LeuAc]<sup>+</sup> encloses the nostopeptolide cycle. Peptide sequence of nostopeptolide 1111 (But<sup>1</sup>-Ile<sup>2</sup>-Ser<sup>3</sup>-MePro<sup>4</sup>-LeuAc<sup>5</sup>-Leu<sup>6</sup>-Ser<sup>7</sup>-Asn<sup>8</sup>-Tyr<sup>9</sup>-Pro<sup>10</sup>) differed from the nostopeptolide A only by the substitution of Gly residue by Ser at position 7, which corresponded to predicted substrate specificity of the respective adenylation domain (antiSMASH).

Given the example of biosynthetic coupling between aeruginosin and spumigin pathways through *nosE-F* homologs, previously

assumed in cyanobacteria<sup>[16]</sup>, we speculated that a similar coupling could occur between the muscotoxin and nostopeptolide pathways in *D. muscorum* CCALA 125. Accordingly, we performed an experiment to test whether the 4-MePro facultatively incorporated in muscotoxins (positions 4 and 11) originates from the biosynthetic pathway of nostopeptolide, which constitutively contains 4-MePro<sup>3</sup>.

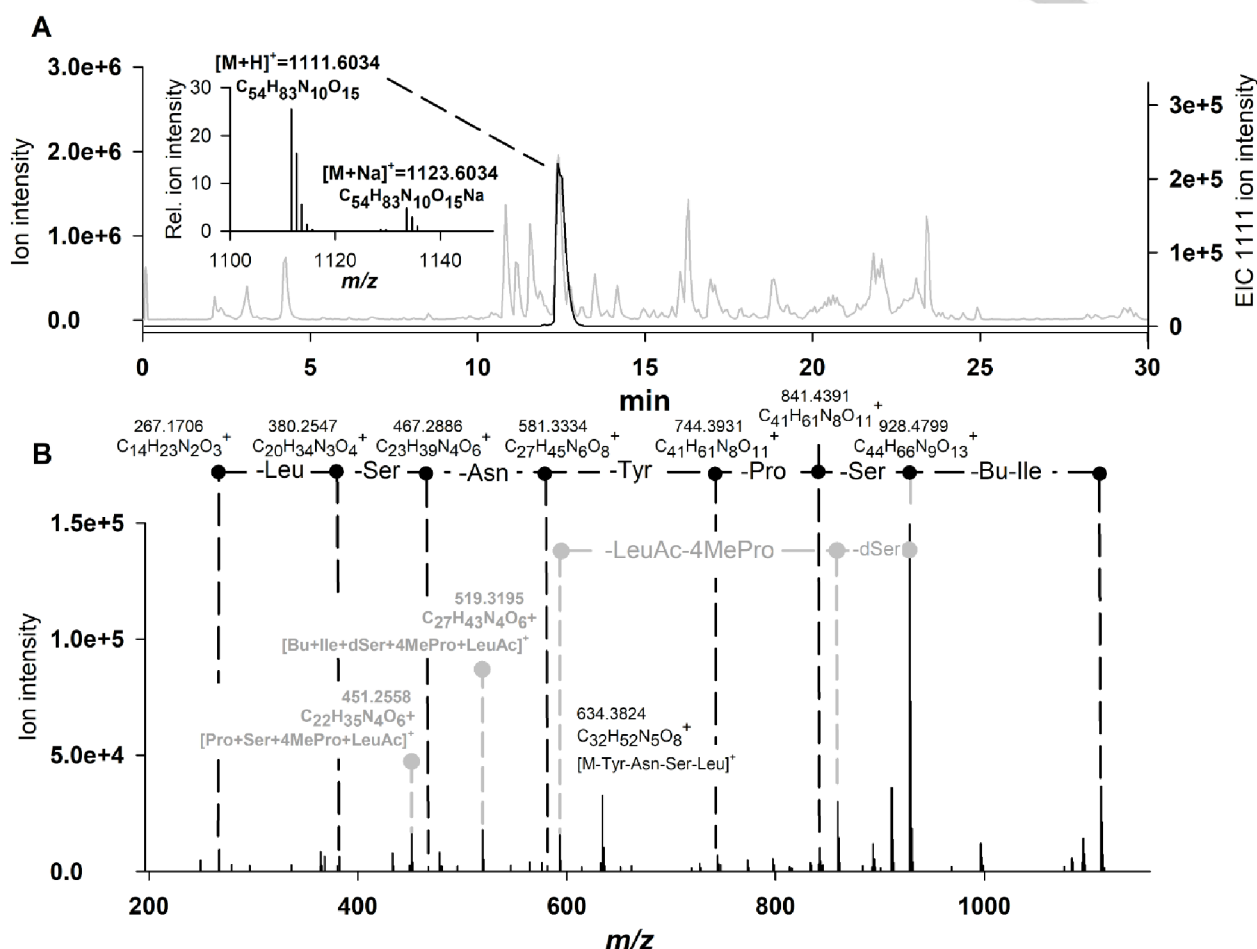
#### Biosynthetic coupling of muscotoxin and nostopeptolide

The 4-MePro residue in nostopeptolide was previously demonstrated to be formed by modification and cyclization of leucine by the action of enzymes encoded within the pathway<sup>[17]</sup>. We have monitored the incorporation of the isotopically substituted Leu into nostopeptolide 1111 as well as muscotoxin A and B molecules in order to reveal if muscotoxin B biosynthesis is dependent on the 4-MePro supply from the nostopeptolide biosynthetic machinery or is incorporated by an alternative mechanism.

The *D. muscorum* CCALA 125 culture was supplemented with Leu substituted with six <sup>13</sup>C atoms and its incorporation was detected as the increase of the +6/+12/+18 Da isotopologues. In the cultures treated with <sup>13</sup>C<sub>6</sub> leucine a clear formation of nostopeptolide-1111 isotopologues at *m/z* 1117 (+6 Da), 1123 (+12 Da) and 1129 (+18 Da) has been detected (Figure 3A). This corresponds well to the presence of two Leu and one 4-MePro residue in its molecule. In the case of muscotoxin B and its structural analog 1225, both possessing one 4-MePro residue, a single isotopologue at *m/z* 1231 (+6 Da) was detected (Figure 3B, Tables 3 and 4). We further confirmed the presence of <sup>13</sup>C<sub>6</sub> 4-MePro in its molecules by the MS/MS experiment. In muscotoxin B the <sup>13</sup>C<sub>6</sub> 4-MePro+5-OH Ahdooa fragment has been formed (Table S1) showing the position of the 4-MePro next to the 5-OH Ahdooa residue. In the structural analog 1225 the <sup>13</sup>C<sub>6</sub> 4-MePro was placed at the position 4, next to the Phe residue as expected (Table S2). We further supplemented the culture with isotopically substituted <sup>13</sup>C<sub>5</sub><sup>15</sup>N-Pro in order to test whether an alternative mechanism utilizing Pro molecule could be employed

in the 4-MePro formation. We did not detect any isotopologues suggesting the employment of Pro in 4-MePro synthesis in the molecules of interest. The incorporation of Pro residue resulting in the formation of +6 Da isotopologues was found only as expected in accordance to the presence of non-methylated Pro

in all the compounds (Figure S1). This clearly demonstrates that 4-MePro in muscotoxin B originates from the Leu molecule and its biosynthesis therefore seems to be fully dependent on the genes encoded in the nostopeptolide pathway.



**Figure 2.** Identification of nostopeptolide 1111 in *Desmonostoc muscorum* CCALA 125. Nostopeptolide 1111 was detected as prominent peak using the HPLC-HRMS/MS analysis (A-base peak chromatogram in grey, nostopeptolide 1111 extracted ion chromatogram in black). Its MS/MS analysis (B) provided series of b-y ions partly uncovering the peptide sequence (in black) and additional ions enclosing the peptide cycle (in grey). LeuAc - Acetylated leucine residue, dSer - dehydrated serine. For the ppm values and details on fragmentation see Table S3.

We further compared the rate of incorporation of 4-MePro into the nostopeptolide and muscotoxin B molecules. The peak areas of the non-substituted and substituted isotopologues in the extracts obtained at fixed biomass to solvent ratio were integrated and recalculated to the actual amount of biomass in the culture (and expressed as total peak area per culture - Figure 4). In control cultures the pool of the non-labelled nostopeptolide 1111 and muscotoxin A-B molecules increased linearly with time, showing the progress of their biosynthesis during the culture growth. In cultures treated with isotopically substituted Leu the pool of non-labelled nostopeptolide 1111, muscotoxin A and B increased only slightly at the beginning of the experiment and then remained stable across the time span 24-72 h (Figure 4).

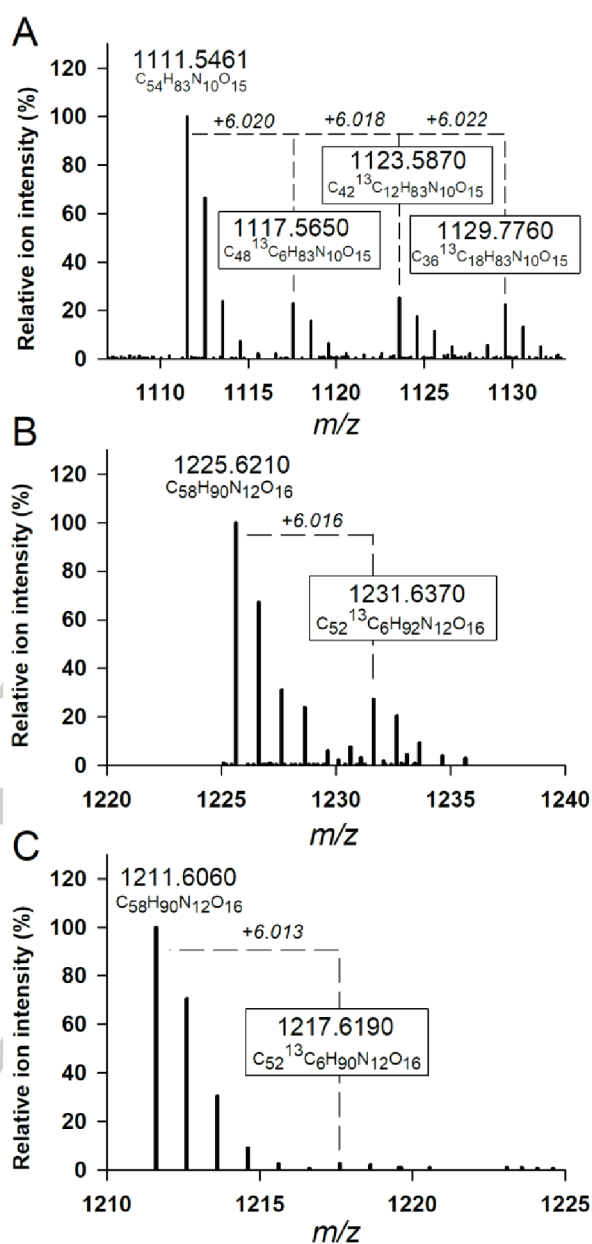
During that period the onset of nostopeptolide 1111 ( $m/z$  1117/1123/1131) as well as muscotoxin B ( $m/z$  1231) isotopologues was recorded, clearly showing that all newly synthesized molecules exhibited incorporation of one to three  $^{13}\text{C}_6$  Leu/4-MePro molecules (Figure 3A). This allowed us to compare the approximate rate by which 4-MePro was

incorporated into nostopeptolide vs muscotoxin B. In the case of nostopeptolide 1111 the minimum possible proportion of the molecules containing 4-MePro can be calculated from the isotopologue  $m/z$  1131 that contains both Leu residues as well as 4-MePro in the isotopically substituted form. Area of this ion contributed by ~13% to the total amount of nostopeptolide 1111 at the end of the experiment. On the other hand, the maximum possible rate of incorporation (assuming that all the nostopeptolide 1111 isotopologues contained 4-MePro) was calculated as 37%. Thus, although we were not able to precisely infer from our experimental data, which proportion of the newly synthesized nostopeptolide 1111 contained 4-MePro, it was certainly in the range 13-37%. In muscotoxin B the proportion of the isotopologue 1231 directly showed the incorporation of 4-MePro to be 19.5% at the end of the experiment. Interestingly, a very similar proportion of incorporated 4-MePro (20.1%) is obtained for nostopeptolide 1111 if we assume that in the isotopologues 1117 and 1123 the two Leu and one 4-MePro have the same probability to be incorporated (knowing that all 1131 molecules contain 4-MePro). From all above mentioned it

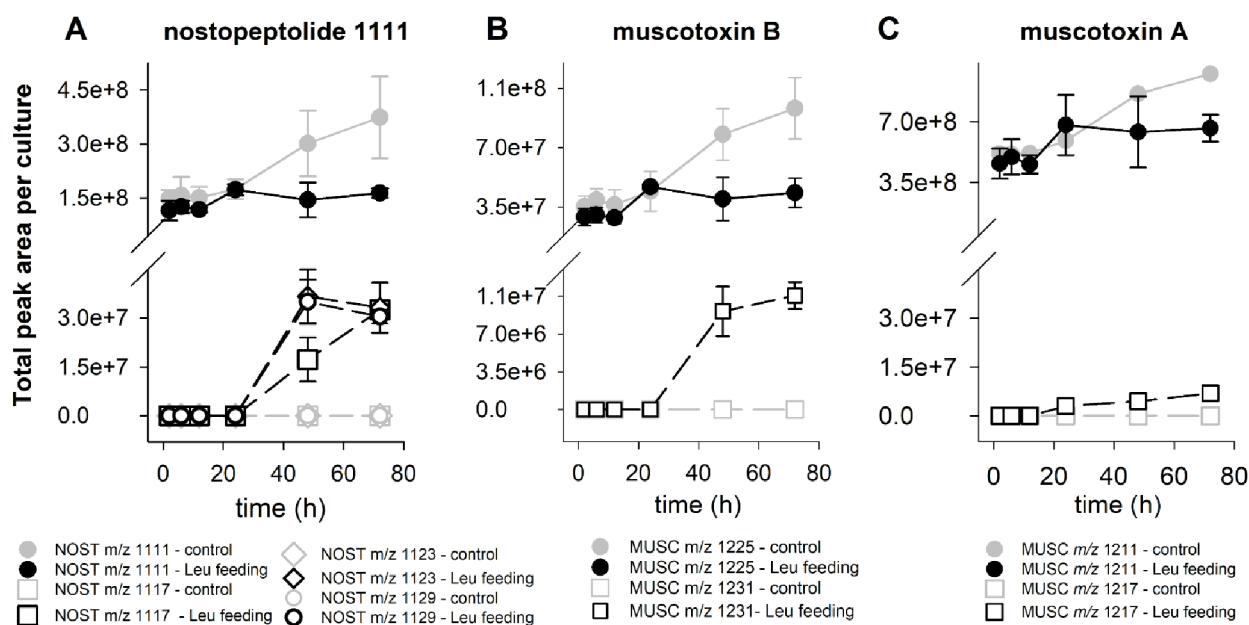
is clear that incorporation of 4-MePro into nostopeptolide 1111 and muscotoxin B molecules happens at a comparable rate. This further suggests that the alcohol dehydrogenase and 3-methyl- $\Delta^1$ -pyrroline-5-carboxylic acid reductase<sup>[17]</sup> provide an equal cytoplasmic pool of 4-MePro for both nostopeptolide and muscotoxin pathways rather than locally provide the substrate for the nostopeptolide, leaving the muscotoxin B synthesis dependent on a leak from the nostopeptolide 1111 biosynthetic machinery. The observed ratio of the major muscotoxin A variant (containing only Pro) and minor muscotoxin B (containing one 4-MePro<sup>[3]</sup>) is therefore most probably a result of the Pro A-domain specificity rather than the insufficient supply of the 4-MePro substrate.

## Conclusion

The muscotoxin gene cluster predicted in this study exhibits high sequence similarity and synteny to a known nostophycin gene cluster except for the presence of three additional NRPS modules. Intriguingly, the muscotoxin gene cluster does not encode a functional starter adenylation (or FAAL) domain typically employed in lipopeptide biosynthesis. The mechanism of FA activation and loading therefore needs experimental study in future. Further experiments proved biosynthetic coupling between a pair of bioactive compounds produced by the cyanobacterium. The genes involved in nostopeptolide assembly that are responsible for the formation of 4-MePro, provide an alternative substrate to a couple of Pro-specific NRPS adenylation domains involved in muscotoxin biosynthesis. Analogous metabolic coupling was previously assumed between spumigin and pseudoaeruginosin based on their distribution patterns<sup>[16]</sup>, however the current study for the first time provides an experimental proof using isotopically substituted amino acid feeding. Surprisingly the incorporation of 4-MePro into both peptides happened at a similar rate, suggesting that the pool of free 4-MePro was equally accessible to both pathways. The biosynthetic crosstalk between these two pathways increases the number of produced variants from 2 to 4. Moreover, one of these 4-MePro substituted variants (muscotoxin B) has recently been shown to exhibit an increased antifungal activity *in vitro*<sup>[14]</sup>. These findings altogether suggest that the biosynthetic crosstalk is not a coincidence and may provide a selective advantage by extending the metabolic potential in ecological interactions among co-existing microbes.



**Figure 3.** Incorporation of  $^{13}\text{C}_6$  leucine into the nostopeptolide 1111, muscotoxin B and A molecules as obtained after 48 h of the feeding experiment. In nostopeptolides three  $^{13}\text{C}_6$  leucine molecules were incorporated leading to formation of isotopologues at  $m/z$  1117, 1123 and 1129 (A). In muscotoxin B only isotopologue at  $m/z$  1231 was recorded (B). In muscotoxin A only a negligible incorporation was found (C).



**Figure 4.** The dynamics of the  $^{13}\text{C}_6$  leucine incorporation into the nostopeptolide 1111, muscotoxin A, and muscotoxin B molecules. In control cultures (grey full lines, full circles) continuous increase of the non-labeled compounds of interest was recorded over time and no increase of the corresponding isotopologues was recorded (grey dashed line, empty symbols). In cultures treated with  $^{13}\text{C}_6$  leucine the amount of non-labeled variants remained stable within the time range 24–72 h (black full line, full circles). During this time span an increase in nostopeptolide labeled isotopologues 1117, 1123 and 1129 (empty squares, circle a diamond, respectively – subfigure A) together with the increase of the isotopologue imuscotoxin B 1231 (empty squares, subfigure B) was recorded. Muscotoxin A (subfigure C) showed only minor incorporation of  $^{13}\text{C}_6$  leucine.

## Experimental Section

### Genome sequencing and bioinformatic analysis

Single filaments of strain *D. muscorum* CCALA 125 were isolated for whole-genome amplification (WGA) and subsequent preparation of a whole-genome sequencing (WGS) library, as described previously<sup>[9]</sup>. Briefly, the glass capillary technique was used to isolate filaments excluding minor bacterial contaminants. A set of 15 single filaments was then used as a template for WGA. Multiple displacement amplification (MDA) using a Repli-g Mini Kit (Qiagen, Hilden, Germany) was followed by PCR and sequencing to monitor the cyanobacterial 16S rRNA gene using primers 16S387F and 16S1494R<sup>[19]</sup>. Positive samples (4 MDA products yielding a clear 16S rRNA gene sequence of the respective strain) were then pooled to create a template for WGS. The quality of DNA in the resulting mixture was checked using gel electrophoresis, the concentration of dsDNA was measured using Qubit 2.0 (Thermo Fisher Scientific, Waltham, MA, USA) broad-range protocol, and the sample was sent for commercial de novo genome sequencing (EMBL Genomics Core Facility, Heidelberg, Germany) using the Illumina MiSeq platform (Illumina, San Diego, CA, USA) with a ~400 bp average insert length Pair-End library and 250 bp reads (~1.4 Gbp data yield). Raw data from de novo WGS were assembled using CLC Bio Genomics Workbench v. 7.5 (CLC Bio, Aarhus, Denmark) to generate genomic scaffolds. Protein coding genes were predicted using Prodigal<sup>[20]</sup> in the assembled

scaffolds longer than 999 bp. All predicted proteins were compared to the NCBI-nr database using MMSeqs<sup>[21]</sup> and only contigs which had most hits to the Cyanobacteria were kept for the rest of analysis. Completeness of the cyanobacterial bin was estimated using CheckM<sup>[22]</sup>. Genomic scaffolds were loaded into Geneious Pro R10 (Biomatters; available from <http://www.geneious.com>) and investigated for NRPS genes using custom BLASTp searches. Functional annotation of ORFs was conducted by applying a combination of BLASTp/CDD searches against the NCBI database, and using the antiSMASH 4.0 secondary metabolite gene cluster annotation pipeline<sup>[23,24]</sup>. Minor assembly gaps identified in the genomic scaffolds after pair-end read assembly were closed by PCR, and subsequent Sanger sequencing of PCR products using custom primers annealing to regions adjacent to the assembly gaps.

### High performance liquid chromatography – high resolution tandem mass spectrometry analysis (HPLC-HRMS/MS)

All *D. muscorum* CCALA 125 samples were analyzed using HPLC-HRMS/MS - Thermo Scientific Dionex UltiMate 3000 UHPLC (Thermo Scientific) equipped with a diode array detector (DAD) and high resolution mass spectrometer with electrospray ionization source (HRMS/MS; Impact HD II Mass Spectrometer, Bruker). HPLC separation was performed on reversed phase Kinetex Phenomenex C18 column (150 x 4.6 mm, 2.6  $\mu\text{m}$ ) with  $\text{H}_2\text{O}$ /acetonitrile acidified with 0.1 %

HCOOH as a mobile phase at a flow rate of 0.6 mL/min. The gradient was as follows: H<sub>2</sub>O/acetonitrile 85/15 (0 min), 85/15 (in 1 min), 0/100 (in 20min), 0/100 (in 25 min) and 85/15 (in 30 min). Following settings of the mass spectrometer were used: dry temperature 200 °C; drying gas flow 12 L/min; nebulizer 3 bar; capillary voltage 4500 V; endplate offset 500 V. The spectra were collected in the range 20-2000 *m/z* with spectral rate of 2 Hz. CID was set as a ramp from 20 to 80 eV on masses 200-2000, respectively, using the automated precursor ion selection. The fragmentation of the minor isotopically substituted muscotoxin variants was achieved via the precursor ion selection to the particular mass with the isolation window ± 5 Da. Calibration of the instrument was performed using CH<sub>3</sub>COONa at the beginning of each analysis. The individual spectra were analyzed using Bruker Data Analysis software (version 4.1) to obtain molecular formula.

### Feeding with isotopically substituted amino acids, extract preparation and data evaluation

The inoculum of *D. muscorum* CCALA 125 was homogeneously transferred into 9 cultivation vessels (100 mL Erlenmeyer flasks) and cultured in 50 mL of BG-11 medium (pH adjusted to 8.0). The cultures were placed on orbital shaker and kept at stable temperature (25°C) and illumination of 200 µE (warm white, AlgaeTron 230, Photon Systems Instruments). Three experimental conditions were used: 1) supplementation with isotopically substituted leucine (L-Leucine-<sup>13</sup>C<sub>6</sub> - Sigma Aldrich - 605239), 2) supplementation with isotopically substituted proline (L-Proline-<sup>13</sup>C<sub>5</sub><sup>15</sup>N, Sigma-Aldrich - 604801) and 3) control cultures. Cultures were sampled 2, 6, 12, 24, 48 and 72 hours after inoculation. The biomass was freeze-dried and the exact weight was recorded for each sample. Subsequently, 7-20 mg of the dry biomass was extracted into 70% MeOH (v/v) in biomass to solvent ratio 0.2 mg/mL. Extracts were dried under stream of nitrogen gas and dissolved in 100 µL of 70% MeOH (v/v). The samples were analyzed using HPLC-HRMS/MS (as described earlier). The peak area of the compounds under study (muscotoxin A, B, nostopeptilide 1111 and their isotopically substituted analogues) were calculated and expressed in relation to the total biomass content in the culture.

### Acknowledgements

The study was funded by the Czech Science Foundation (GAČR) project no. 16-09381S („Bioactive cyanobacterial lipopeptides: genome mining, detection, and structure-activity relationships“). DC was additionally supported by Czech Science Foundation (GAČR) project no. 18-24397S.

**Keywords:** lipopeptides • biosynthesis • cyanobacteria • 4-methylproline • metabolic crosstalk

- [1] S. Sumimoto, M. Kobayashi, R. Sato, S. Shinomiya, A. Iwasaki, S. Suda, T. Teruya, T. Inuzuka, O. Ohno, K. Suenaga, *Organic Letters* **2019**, *21*, 1187-1190.
- [2] A. Humisto, J. Jokela, K. Teigen, M. Wahlsten, P. Permi, K. Sivonen, L. Herfindal, *Biochimica et Biophysica Acta-Biomembranes* **2019**, *1861*, 1510-1521.
- [3] P. Tomek, P. Hrouzek, M. Kuzma, J. Sýkora, R. Fišer, J. Černý, P. Novák, S. Bártová, P. Šimek, M. Hof, D. Kavan, J. Kopecký, *Chemical Research in Toxicology* **2015**, *28*, 216-224.
- [4] P. Hrouzek, M. Kuzma, J. Černý, P. Novák, R. Fišer, P. Šimek, A. Lukešová, J. Kopecký, *Chemical Research in Toxicology* **2012**, *25*, 1203-1211.
- [5] L. Oftedal, L. Myhren, J. Jokela, G. Gausdal, K. Sivonen, S. O. Doskeland, L. Herfindal, *Biochimica et Biophysica Acta-Biomembranes* **2012**, *1818*, 3000-3009.
- [6] T. Galica, P. Hrouzek, J. Mareš, *Journal of Phycolgy* **2017**, *53*, 985-998.
- [7] J. Vestola, T. K. Shishido, J. Jokela, D. P. Fewer, O. Aitto, P. Permi, M. Wahlsten, H. Wang, L. Rouhiainen, K. Sivonen, *Proceedings of the National Academy of Sciences of the United States of America* **2014**, *111*, E1909-E1917.
- [8] J. Mareš, J. Hájek, P. Urajová, J. Kopecký, P. Hrouzek, *PLoS One* **2014**, *9*, e111904.
- [9] T. K. Shishido, J. Jokela, C.-T. Kolehmainen, D. P. Fewer, M. Wahlsten, H. Wang, L. Rouhiainen, E. Rizzi, G. De Bellis, P. Permi, K. Sivonen, *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, 13669-13674.
- [10] J. Mareš, J. Hájek, P. Urajová, A. Kust, J. Jokela, K. Saurav, T. Galica, K. Čápková, A. Mattila, E. Haapaniemi, P. Permi, I. Mysterud, O. M. Skulberg, J. Karlsen, D. P. Fewer, K. Sivonen, H. H. Tonnesen, P. Hrouzek, *Applied and Environmental Microbiology* **2019**, *85*, e02675-18.
- [11] M. A. Marahiel, T. Stachelhaus, H. D. Mootz, *Chemical Reviews* **1997**, *97*, 2651-2673.
- [12] D. B. Hansen, S. B. Bumpus, Z. D. Aron, N. L. Kelleher, C. T. Walsh, *Journal of the American Chemical Society* **2007**, *129*, 6366-6367.
- [13] F. I. Kraas, V. Helmetag, M. Wittmann, M. Strieker, M. A. Marahiel, *Chemistry & Biology* **2010**, *17*, 872-880.
- [14] J. Cheel, J. Hájek, M. Kuzma, K. Saurav, I. Smýkalová, E. Ondráčková, P. Urajová, V. Dai Long, K. Faure, J. Kopecký, P. Hrouzek, *Molecules* **2018**, *23*, 2653.
- [15] L. Liu, J. Jokela, L. Herfindal, M. Wahlsten, J. Sinkkonen, P. Permi, D. P. Fewer, S. O. Doskeland, K. Sivonen, *ACS Chemical Biology* **2014**, *9*, 2646-2655.
- [16] L. Liu, A. Budnjo, J. Jokela, B. E. Haug, D. P. Fewer, M. Wahlsten, L. Rouhiainen, P. Permi, T. Fossen, K. Sivonen, *ACS Chemical Biology* **2015**, *10*, 725-733.
- [17] H. Luesch, D. Hoffmann, J. M. Hevel, J. E. Becker, T. Golakoti, R. E. Moore, *Journal of Organic Chemistry* **2003**, *68*, 83-91.
- [18] D. P. Fewer, J. Osterholm, L. Rouhiainen, J. Jokela, M. Wahlsten, K. Sivonen, *Applied and Environmental Microbiology* **2011**, *77*, 8034-8040.
- [19] A. Taton, S. Grubisic, E. Brambilla, R. De Wit, A. Wilmotte, *Applied and Environmental Microbiology* **2003**, *69*, 5157-5169.
- [20] D. Hyatt, G.-L. Chen, P. F. LoCasio, M. L. Land, F. W. Larimer, L. J. Hauser, *BMC Bioinformatics* **2010**, *11*, 119.
- [21] M. Hauser, M. Steinegger, J. Soeding, *Bioinformatics* **2016**, *32*, 1323-1330.
- [22] D. H. Parks, M. Imelfort, C. T. Skennerton, P. Hugenholtz, G. W. Tyson, *Genome Research* **2015**, *25*, 1043-1055.
- [23] T. Weber, K. Blin, S. Duddela, D. Krug, H. U. Kim, R. Brucoleri, S. Y. Lee, M. A. Fischbach, R. Muller, W. Wohlleben, R. Breitling, E. Takano, M. H. Medema, *Nucleic Acids Research* **2015**, *43*, W237-W243.
- [24] K. Blin, T. Wolf, M. G. Chevrette, X. W. Lu, C. J. Schwalen, S. A. Kautsar, H. G. S. Duran, E. Santos, H. U. Kim, M. Nave, J. S. Dickschat, D. A. Mitchell, E. Shelest, R. Breitling, E. Takano, S. Y. Lee, T. Weber, M. H. Medema, *Nucleic Acids Research* **2017**, *45*, W36-W41.
- [25] P. Urajová, J. Hájek, M. Wahlsten, J. Jokela, T. Galica, D. P. Fewer, A. Kust, E. Zapomělová-Kozlíková, K. Delawska, K. Sivonen, J. Kopecký, P. Hrouzek, *Journal of Chromatography A* **2016**, *1438*, 76-83.
- [26] T. P. Korman, J. A. Hill, T. N. Vu, S. C. Tsai, *Biochemistry* **2004**, *43*, 14529-14538.
- [27] D. Hoffmann, J. M. Hevel, R. E. Moore, B. S. Moore, *Gene* **2003**, *311*, 171-180.
- [28] T. Golakoti, W. Y. Yoshida, S. Chaganty, R. E. Moore, *Tetrahedron* **2000**, *56*, 9093-9102.
- [29] A. Guljamov, M. Kreische, K. Ishida, A. Liaimer, B. Altermark, L. Baehr, C. Hertweck, R. Ewald, E. Dittmann, *Applied and Environmental Microbiology* **2017**, *83*, e01510-17.

## 5. Curriculum Vitae

### Personal information:

Name and surname: Tomáš Galica

E-mail: tomasgalica81@gmail.com

### Education:

Bc. (B.Sc.) - Program: Molecular Biology and Biochemistry of Organisms. (2007-2010), *Faculty of Sciences of Charles University in Prague, Prague* (Czech Republic)

Mgr. (M.Sc.) - Program: Genetics, Molecular Biology and Virology. (2010-2012), *Faculty of Sciences of Charles University in Prague, Prague* (Czech Republic)

Ph.D. - Program: Ecosystem biology. (2013-present), *Faculty of Science, University of South Bohemia in České Budějovice, České Budějovice* (Czech Republic)

### Employment:

Laboratory technician (part-time 2009-2012), Laboratory of Cell Immunology (prof. Jan Černý, Ph.D.), Department of Cell Biology, *Faculty of Sciences of Charles University in Prague, Prague* (Czech Republic)

Research assistant/Ph.D. student (part-time 2014, 2016), Laboratory of Phytoplakton Ecology (RNDr. Jan Mareš, Ph.D.), *Institute of Hydrobiology, Biology Centre, Czech Academy of Sciences, České Budějovice* (Czech Republic)

Research assistant/Ph.D. Student (2013-present), Laboratory of algal Biotechnology (RNDr. Pavel Hrouzek, Ph.D.), Centre AlgaTech, *Institute of Microbiology, Třeboň* (Czech Republic)

### Internships:

Erasmus study internship. (18. 1. – 7. 6. 2012), *University of Gotheburg, Gothenburg* (Sweden); Courses: Molecular prokaryotic microbiology (19. 1. - 26. 3. 2012) and Bioinformatics and functional genomics (26. 3. – 7. 6. 2012)

Research internship in research group of Kaarina Sivonen. (5. 9. – 3.12. 2014), *University of Helsinki, Helsinki* (Finland), supervisor David P. Fewer.

## Conferences:

*ICTC9 – 9th International Conference on Toxic Cyanobacteria.* (2013), Pilanesberg, Republic of South Africa, Oral presentation – Response of *Microcystis aeruginosa* cells to presence of potentially competitive cyanobacteria includes increase of intracellular microcystin concentration.

*EGAMEE - Ecology and Geochemical Activity of Microorganisms in Extreme Environments.* (2013), Ulan-Ude (Russian Federation), Oral presentation - The possible role of cyanobacterial lipopeptides in induction of Haff disease.

*ISPP2018 - 16th International Symposium on Phototrophic Prokaryotes.* (2018), Vancouver (Canada), Poster - Novel  $\beta$ -hydroxyaspartate Containing Lipopeptide with Iron Scavenging Potential Identified in *Cyanobacterium Rivularia* sp. PCC 7116

*Copenhagen Bioscience Conferences, 16th Conference - Natural Products: Discovery Biosynthesis and Application.* (2019), Poster -  $\beta$ -hydroxyaspartate-containing lipopeptide produced by *Leptolyngbya* sp. NISE-3755 is a representative of a novel class of cyanobacterial siderophores.

*3rd ISSNP - International Summer School of Natural Products.* (2019), Flash talk - Iron Throne:  $\beta$ -hydroxyaspartate Ligand System for Iron Chelation Is Found in Lipopeptides Produced by Cyanobacteria of Various Lineages.

## Publications:

Urajová, P., Hájek, J., Wahlsten, M., Jokela, J., **Galica, T.**, Fewer, D., Kust, A., Zapomělová-Kozlíková, E., Delawska, K., Sivonen, K., Kopecký, J. & Hrouzek, P. (2016). A liquid chromatography-mass spectrometric method for the detection of cyclic beta-amino fatty acid lipopeptides. *Journal of Chromatography A*, 1438, 76-83.

<https://doi.org/10.1016/j.chroma.2016.02.013>

Parajuli, A., Kwak, D. H., Dalponte, L., Leikoski, N., **Galica, T.**, Umeobika, U., Trembleau, L., Bent, A., Sivonen, K., Wahlsten, M., Wang, H., Rizzi, E., De Bellis, G., Naismith, J., Jaspars, M., Liu, X., Houssen, W., & Fewer, D. P. (2016). A Unique Tryptophan C-Prenyltransferase from the



Kawaguchi-peptin Biosynthetic Pathway. *Angewandte Chemie International Edition*, 55(11), 3596-3599. <https://doi.org/10.1002/anie.201509920>

**Galica, T.**, Hrouzek, P. & Mareš, J. (2017). Genome mining reveals high incidence of putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP ligase genes in biofilm-forming cyanobacteria. *Journal of Phycology*, 53(5), 985-998. <https://doi.org/10.1111/jpy.12555>

Mareš, J., Hájek, J., Urajová, P., Kust, A., Jokela, J., Saurav, K., **Galica, T.**, Čapková, K., Mattila, A., Haapaniemi, E., Permi, P., Myrnes, I., Skulberg, O., Karlsen, J., Fewer, D., Sivonen, K., Tonnesen, H. & Hrouzek, P. (2019).

Alternative Biosynthetic Starter Units Enhance the Structural Diversity of Cyanobacterial Lipopeptides. *Applied and Environmental Microbiology*, 85(4), Article ARTN e02675-18. <https://doi.org/10.1128/AEM.02675-18>

**Galica T.**, Borbone N., Mareš J., Kust A., Caso A., Esposito G., Saurav K., Hájek J., Řeháková K., Urajová P., Costantino V. & Hrouzek P. (2021) Cyanochelins, an overlooked class of widely distributed cyanobacterial siderophores, discovered by silent gene cluster awakening. *Journal of Applied Environmental Microbiology*, in press

Fewer, D. P., Jokela, J., Heinilä, L. M. P., Asoy, R., Sivonen, K., **Galica, T.**, Hrouzek, P., & Herfindal, L. (2021). Chemical diversity and cellular effects of antifungal cyclic lipopeptides from cyanobacteria. *Physiologia Plantarum*, in press

© for non-published parts:

Tomáš Galica

tomasgalica81@gmail.com

Diversity and Ecological Role of Cyanobacterial Lipopeptides

Ph.D. Thesis Series, 2021, No. 13

All rights reserved

For non-commercial use only

Printed in the Czech Republic by Typodesign

Edition of 15 copies

University of South Bohemia in České Budějovice

Faculty of Science

Branišovská 1760

CZ-37005 České Budějovice, Czech Republic

Phone: +420 387 776 201

[www.prf.jcu.cz](http://www.prf.jcu.cz), e-mail: [sekret-fpr@prf.jcu.cz](mailto:sekret-fpr@prf.jcu.cz)