Jan Mareš, Jan Hájek, Petra Urajová, Andreja Kust, Jouni Jokela, Kumar Saurav, Tomáš Galica, Kateřina Čapková, Antti Mattila, Esa Haapaniemi, Perttu Permi, Ivar Mysterud, Olav M. Skulberg, Jan Karlsen, David P. Fewer, Kaarina Sivonen, Hanne Hjorth Tønnesen, Pavel Hrouzek.

Applied and Environmental Microbiology Feb 2019, 85 (4) e02675-18; DOI: 10.1128/AEM.02675-18

Alternative biosynthetic starter units enhance the structural diversity of cyanobacterial lipopeptides

Jan Mareš,[1,2,3] Jan Hájek,[2,3] Petra Urajová,[2] Andreja Kust,[1,2,3] Jouni Jokela,[4] Kumar Saurav,[2] Tomáš Galica,[2,3] Kateřina Čapková,[1] Antti Mattila,[4] Esa Haapaniemi,[5,6] Perttu Permi,[5] Ivar Mysterud,[7] Olav M. Skulberg,[8] Jan Karlsen,[9] David P. Fewer,[4] Kaarina Sivonen,[4] Hanne Hjorth Tønnesen,[9] Pavel Hrouzek[2,3#]

[1]The Czech Academy of Sciences, Biology Centre, Institute of Hydrobiology, České Budějovice, Czech Republic

[2]The Czech Academy of Sciences, Institute of Microbiology, Center Algatech, Třeboň, Czech Republic

[3]University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic

[4]Department of Microbiology, Biocenter 1, Helsinki, Finland

[5]Department of Chemistry, University of Jyväskylä, Finland

[6]Department of Biological and Environmental Science, Nanoscience Center, University of Jyväskylä, Finland

[7]Department of Biosciences, University of Oslo, Oslo, Norway

[8]Norwegian Institute for Water Research (NIVA), Oslo, Norway

[9]School of Pharmacy, University of Oslo, Oslo, Norway

Running title: Cyanobacterial lipopeptide biosynthesis

#Address correspondence to Pavel Hrouzek, hrouzekp@gmail.com

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 biosynthetic gene clusters for 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 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 length and FA substitution. 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* UTEX B 1613. We detected PUW variants bearing an acetyl substitution in *Symplocastrum muelleri* 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 PUW/MIN interact preferentially with eukaryotic plasma membranes.

Importance: Herein, we aimed to decipher 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 identified variants 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.

Introduction

Bacterial lipopeptides are a prominent group of secondary metabolites with pharmaceutical potential as antibacterial, antifungal, anticancer, and antiviral agents (1). Compounds such as 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 *β*-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 length and substitution 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). MINs A–L exhibited antiproliferative effects when tested against human cancer cell lines over a concentration range similar 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/non-ribosomal 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 resemblance to the biosynthesis of iturin-family lipopeptides (16–19) as well as small lipopeptide-like cyanobacterial metabolites such as 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-CoA ligases (FACLs) following a specific insertion that hampered subsequent ligation to 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 identified lipopeptide variants, and compare the predicted functions of synthetase enzymes.

Results and Discussion

Structural variability vs. common biosynthetic origin of PUWs and MINs

In the present study, we collected all known PUW/MIN producers (except for *Anabaena* sp. UIC10035). The strains 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).

The MS/MS data acquired for crude extracts were used to create a molecular network (Fig. 2), analysis of which demonstrated that *Cylindrospermum* strains 1–3 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 the common peptide sequence FA^{1}-Val^{2}-Dhb^{3}-Asn^{4}-Dhb^{5}-Asn^{6}-Ala^{7}-Thr^{8}-*N*MeAsn^{9}-Pro^{10} (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). By 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^{1}-Val^{2}-Dhb^{3}-Thr^{4}-Thr/Val^{5}-Gln^{6}-Ala^{7}-*O*Me-Thr^{8}-*N*MeAsn^{9}-Pro^{10} (Fig. 3), identical to PUW A–D and MIN I, K, L isolated previously from *Anabaena* sp. (9, 11).

The peptide core 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*, *puwE–H*; Fig. 5) encoding a sequence of nine amino acid-incorporating modules (Fig. 6) was detected in all analyzed strains. 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 observed types of peptide cores (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→*O*Me-Thr), as shown in Fig. 3 and Table S1. This was reflected in the predicted substrates of the corresponding 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^{8} (Fig. 6, Table S2). In contrast to the variability observed at the previously noted amino acid positions, the two positions adjacent to both sides to the modified fatty acid (*N*MeAsn^{9}-Pro^{10}-(FA^{1})-Val^{2}-Dhb^{3}) 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.

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 of the pathways (Fig. 6), indicating 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-H (10, 11), and D-Gly was identified in MIN I-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^4-Thr^4 and Ala^7-Ser^7, respectively (Fig. 6). This degree of substrate promiscuity is relatively uncommon. 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 vs. Ser was previously reported from fungal class IV adenylate-forming reductases that contain A-domains homologous to 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 β-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 biosythesis 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 (13) (Fig. 4). We identified three alternative arrangements of the putative

175 FAAL starter units (Fig. 5 and 6), each corresponding to a different array of FA side chains
176 detected by HPLC-HRMS/MS, which presumably reflects the range of FA substrates
177 activated during their biosynthesis (Fig. 4). *Cylindrospermum* sp. strains 1–3 possess the Type
178 I putative starter unit consisting of a standalone FAAL enzyme PuwC and a separate ACP
179 PuwD (Fig. 5, Table 2). By contrast, the biosynthetic gene cluster of *S. muelleri* strain 6
180 contains the Type II putative starter unit (PuwI) consisting of a FAAL fused to an ACP (Fig.
181 5, Table 2). *Anabaena* spp. strains 4 and 5 combine both Type I and Type II putative starter
182 units in their biosynthetic gene clusters (Fig. 5, Table 2). Although the functions and substrate
183 ranges of these hypothetical starter units requires further confirmation by gene manipulation
184 experiments, they are supported by the patterns of lipopeptide variants detected by HPLC-
185 HRMS/MS (Fig. 4, Table S1). In *Cylindrospermum* strains 1–3 that exclusively contain the
186 Type I starter unit, the PUW/MIN products exhibited an almost continuous FA distribution
187 between C_{10}–C_{15} (up to C_{17} in negligible trace amounts; Fig. 4). In *S. muelleri* strain 6, the
188 presence of the Type II loading module resulted in production of PUW/MIN variants with
189 discrete FA lengths of C_{16} and C_{18}. Strains containing both Type I and Type II starter units
190 (*Anabaena* strains 4 and 5) produced two sets of PUW/MIN products with no overlap
191 (C_{12}–$C_{14\text{-}15}$ for the Type I pathway, and C_{16} for the Type II pathway), but exhibited a slightly
192 shifted length distribution (Fig. 4). Based on these results, it seems plausible that PuwC/D and
193 PuwI represent two alternative FAAL starter modules capable of initiating PUW/MIN
194 biosynthesis (Figs. 5, 6). An analogous situation was previously described for the alternative
195 NRPS starter modules in the anabaenopeptin synthetase (32).

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

Mycobacterium tuberculosis was 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_{12} substrate, but shorter chains (C_2 or C_{10}, 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* (28) (Fig. S2b). Experimental evidence such as *in vitro* activity assays and crystallization of protein-ligand complexes is required to explain the variable substrate specificity of PuwC vs. PuwI. 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 1–3, 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, 4).

FA tailoring reactions: oxidation, halogenation, and acetylation

Intriguingly, all products originating from biosynthesis initiated by the Type II starter unit (variants with a C_{16} and C_{18} FA tail in *Anabaena* strains 4–5 and *S. muelleri* strain 6) include 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_{14}), as described previously (10), and this position was confirmed by NMR in variants produced by *Anabaena sp.* strain 4 in our study (Table S3, Figs. S3–6). In agreement with this hydroxy- and oxo-substitution, the respective gene clusters each encode PuwJ, a putative cytochrome P450-like oxidase (Table 2), immediately downstream of the gene encoding the Type II starter module.

We therefore hypothesize that the PuwJ enzyme is responsible for hydroxylation of FA residues activated by PuwI (Fig. 6). However, the formation of the keto variant remains unexplained by our data.

Another gene, the putative halogenase *puwK*, 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 halogenation of cyanobacterial chlorinated acyl amides known as columbamides (33), and *N*-oxygenases similar to *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 *ω*-chlorinated product MIN B, originally described in strain *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* sp. strains 4 and 5 share identical organization across the entire gene cluster, and lack of the putative halogenase gene *puwK* is the only difference between these two clusters in terms of presence of genes (Fig. 5).

In *Cylindrospermum* sp. strains 1–3 that 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 knock-out 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) involved in acetylation of the aliphatic chain of microcystin in cyanobacteria (37). Additionally, this gene is similar to 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* 1265.7338, 1279.7496, 1277.7695, 1291.7870 and 1293.7654) yielding high-energy fragments, proving the presence of an acetyl group bonded to the FA moiety, have been detected. In the *m/z* 1279.8 and 1293.8 peaks, the high-energy fragment ion at *m/z* 312 corresponds to the FA immonium ion bearing an acetyl group, and 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* 1265.7 peak revealed analogous fragments at *m/z* 284/411 and 351/224 (Table 3).

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 MINs A, C, and D) did not exert antibacterial effects against either Gram-positive or Gram-negative bacteria using a panel of 13 selected strains (Table 4). PUW F was the only tested variant manifesting antagonistic activity against two yeast strains utilized in our experiment, namely *Candida albicans* HAMBI 261 and *Saccharomyces cerevisiae* HAMBI 1164, with inhibition zones of 14 and 18 mm, respectively, and minimum inhibitory concentration (MIC) values of 6.3 µg mL^{-1} (5.5 µM; 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 $-CH_2-CH_3$ 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- substitution also compromises 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 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 anabanenolysin A and hassalidins 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 non-planktic 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 FAsubstrates of variable length. 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.* UHCC-0399 (previously *Anabaena* sp. 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 1–5 were cultivated in BG-11 media (47) in glass columns (300 mL) bubbled with air enriched in 1.5% CO_2 at a temperature of 28°C and

constant illumination of 100 µmol photons m^{-2} s^{-1}. 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 pre-mix (according to SAG – Sammlung von Algenkulturen der Universität Göttingen, but without biotin). Cultivation was performed in 100–200 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^{-2} s^{-1} PHAR generated using RGB LED strips). Cells were harvested by centrifugation (3125 × *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 L 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 excluding 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 Mini Kit (Qiagen, Hilden, Germany) was followed by PCR and sequencing to monitor the cyanobacterial 16S rRNA gene using primers 16S387F and 16S1494R (49). Positive samples (7–10 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 a ~350 bp average insert length Pair-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 v. 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 cycles run. Genome data (1Gb for each strain) were first checked using SPAdes version 3.7.1 (51) 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; available from http://www.geneious.com) 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 $<10^{-20}$) were then analyzed using the Glimmer 3 (50) 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 investigated strains, either directly after pair-end read assembly, or based on mapping to the reference gene cluster from *C. alatosporum* CCALA 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% MeOH, v/v). 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 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 C18 column (150 × 4.6 mm, 2.6 µm) using H_2O (A)/acetonitrile (B) containing 0.1% HCOOH as a mobile phase, at a flow rate of 0.6 mL min^{-1}. 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 number of carbons in the FA moiety in PUW/MIN variants containing nonsubstituted and hydroxy-/chloro-substituted FA was determined using a method described previously by our team (13). Characteristic FA immonium fragments in oxo-substituted PUW/MIN variants were identified by employing 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 $C_{15}H_{30}NO^+$ was obtained for MIN D (Fig. S1), analogous fragments with general formula $C_xH_{2x}NO^+$ were used to identify oxo-substituted components in unknown PUW/MIN variants from other investigated strains.

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 a 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 >5%. The diluted extract was pre-purified using a Supelco C18 SPE cartridge (10 g, 60 mL) pre-equilibrated with 60 mL of MeOH and 120 mL of H_2O. After loading, retained components were eluted with 60 mL of pure MeOH, concentrated to dryness, and resuspended in 10 mL of pure MeOH. MINs 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 multi-wavelength detector and automatic fraction collector. A preparative Reprosil 100 C18 column (252 × 25 mm) was employed for

423 separation at a flow rate of 10 mL min^{-1} using the following gradient of MeOH containing
424 0.1% HCOOH (A) and 10% MeOH containing 0.1% HCOOH (B): 0 min (100% B), 6 min
425 (100% B), 15 min (43% B), 43 min (12% B), 45 min (0% B), 58 min (0% B), 60 min (100%
426 B), and 64 min (100% B). Fractions were collected using an automatic fraction collector at 1
427 min intervals, and fractions were analyzed for MIN A, C, and D using the method described
428 above. Fractions containing MIN A, C, and D were collected in separate vials and
429 concentrated using a rotary evaporator.The second purification step was performed on a semi-
430 preparative HPLC (Agilent 1100 Infinity series) using a Reprosil 100 Phenyl column (250 × 8
431 mm) with (A) acetonitrile containing 0.1% HCOOH and (B) water containing 0.1% HCOOH
432 using the following gradient: 0 min (60% B), 2 min (60% B), 6 min (50% B), 28 min (18%
433 B), 30 min (0% B), 30 min (0% B), 32 min (0% B), 31 min (60% B), and 36 min (60% B).
434 The flow rate was 1 mL min^{-1} throughout, fractions were collected manually, and the purity
435 was analyzed using the HPLC-HRMS method described above. NMR spectra of
436 minutissamides were measured in dimethyl sulfoxide (DMSO)-*d6* at 30°C. All NMR spectra
437 were collected using a Bruker Avance III 500 MHz NMR spectrometer, equipped with a 5
438 mm ∅ BBI probehead with actively shielded z-gradient.

439

440 ***Antibacterial and antifungal assays***

441 The antimicrobial activity of four major variants (PUW F, and MINs A, C, and D) was
442 tested against 13 bacterial and two yeast strains (Table 4) using disc diffusion assays (8) in
443 three independent experiments with kanamycin/nystatin and MeOH as positive and negative
444 controls, respectively. Antifungal activity of PUW F was further evaluated by determining the
445 MIC against *Candida albicans* (HAMBI 261) and *Saccharomyces cerevisiae* (HAMBI 1164)
446 as described previously (8). PUW F was isolated from *Cylindrospermum* strain 1 according to

447 a protocol described previously (5), and isolation of MIN A, C, and D was performed as
448 described above. The variants produced by *S. muelleri* strain 6 were impossible to isolate due
449 to the slow growth of the cyanobacterium, resulting in low biomass yields during the study
450 period.

451 **Accession numbers** for the newly sequenced complete putative biosynthetic gene
452 clusters uploaded to the NCBI GenBank database are MH325197-MH325201.

453

454 **Acknowledgements**

455 This work was supported by Czech Science Foundation grant no. 16-09381S (Bioactive
456 cyanobacterial lipopeptides: genome mining, detection, and structure-activity relationships),
457 and by the Ministry of Education, Youth and Sports of the Czech Republic, National
458 Programme of Sustainability I, ID: LO1416 project ALGAMIC (ID: CZ.1.05/2.1.00/19.0392)
459 and MSCA IF II project (CZ.02.2.69/0.0/0.0/18_070/0010493). Access to instruments and
460 other facilities was supported by the Czech Research Infrastructure for Systems Biology
461 (C4SYS; project no. LM2015055). This research was also supported by a grant from the
462 NordForsk NCoE programme "NordAqua" (project no. 82845). The Norwegian participation
463 was supported by grants from the Department of Agriculture and Forestry, the County
464 Governor of Møre og Romsdal, the County Governor of Sogn og Fjordane, the University of
465 Oslo, and the Norwegian Institute for Water Research. The authors declare no conflict of
466 interest. The funders had no role in study design, data collection and interpretation, or the
467 decision to submit the work for publication.

468

469 **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.
2. Taylor SD, Palmer M. 2016. The action mechanism of daptomycin. Bioorg Med Chem 24:6253–6268.
3. Velkov T, Roberts KD, Li J. 2017. Rediscovering the octapeptins. Nat Prod Rep 34:295–309.
4. Ines M, Dhouha G. 2015. Lipopeptide surfactants: Production, recovery and pore forming capacity. Peptides 71:100–112.
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.
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.
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.
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. P Natl Acad Sci USA 111:E1909–E1917.

9. Gregson JM, Chen JL, Patterson GML, Moore RE. 1992. Structures of puwainaphycins A-E. Tetrahedron 48:3727–3734.

10. Kang HS, Krunic 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.

11. Kang HS, Sturdy M, Krunic A, Kim H, Shen Q, Swanson SM, Orjala J. 2012. Minutissamides E-L, antiproliferative cyclic lipodecapeptides from the cultured freshwater cyanobacterium cf. *Anabaena* sp. Bioorg Med Chem 20:6134–6143.

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.

13. Urajová P, Hájek J, Wahlsten M, Jokela J, Galica T, Fewer DP, Kust A, Zapomělová-Kozlíková E, Delawská 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.

14. Cheel J, Urajová P, Hajek 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.

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.

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. P Nat Acad Sci USA 96:13294–13299.

17. Tsuge K, Akiyama T, Shoda M. 2001. Cloning, sequencing, and characterization of the iturin A operon. J Bacteriol 183:6265–6273.

18. Koumoutsi 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.

19. Sood S, Steinmetz H, Beims H, Mohr KI, Stadler M, Djukic M, von der Ohe W, Steinert M, Daniel R, Muller R. 2014. Paenilarvins: Iturin family lipopeptides from the honey bee pathogen *Paenibacillus larvae*. ChemBioChem 15:1947–1955.

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

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.

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 *Lyngbya majuscula*. Chem Biol 11:817–833.

544 23. Galica T, Hrouzek P, Mareš J. 2017. Genome mining reveals high incidence of
545 putative lipopeptide biosynthesis NRPS/PKS clusters containing fatty acyl-AMP
546 ligase genes in biofilm-forming cyanobacteria. J Phycol 53:985–998.
547 24. Arora P, Goyal A, Natarajan VT, Rajakumara E, Verma P, Gupta R, Yousuf M,
548 Trivedi OA, Mohanty D, Tyagi A, Sankaranarayanan R, Gokhale RS. 2009.
549 Mechanistic and functional insights into fatty acid activation in *Mycobacterium*
550 *tuberculosis*. Nat Chem Biol 5:166–173.
551 25. Liu Z, Ioerger TR, Wang F, Sacchettini JC. 2013. Structures of *Mycobacterium*
552 *tuberculosis* FadD10 protein reveal a new type of adenylate-forming enzyme. J
553 Biol Chem 288:18473–18483.
554 26. Coates RC, Podell S, Korobeynikov A, Lapidus A, Pevzner P, Sherman DH, Allen EE,
555 Gerwick L, Gerwick WH. 2014. Characterization of cyanobacterial hydrocarbon
556 composition and distribution of biosynthetic pathways. Plos One 9:e85140.
557 27. Mohanty D, Sankaranarayanan R, Gokhale RS. 2011. Fatty acyl-AMP ligases and
558 polyketide synthases are unique enzymes of lipid biosynthetic machinery in
559 *Mycobacterium tuberculosis*. Tuberculosis 91:448–455.
560 28. Goyal A, Verma P, Anandhakrishnan M, Gokhale RS, Sankaranarayanan R. 2012.
561 Molecular basis of the functional divergence of fatty acyl-AMP ligase biosynthetic
562 enzymes of *Mycobacterium tuberculosis*. J Mol Biol 416:221–238.
563 29. Villiers BRM, Hollfelder F. 2009. Mapping the limits of substrate specificity of the
564 adenylation domain of TycA. ChemBioChem 10:671–682.
565 30. Christiansen G, Philmus B, Hemscheidt T, Kurmayer R. 2011. Genetic variation of
566 adenylation domains of the anabaenopeptin synthesis operon and evolution of
567 substrate promiscuity. J Bacteriol 193:3822–3831.

568 31. Brandenburger E, Braga D, Kombrink A, Lackner G, Gressler J, Künzler M,
569 Hoffmeister D. 2018. Multi-genome analysis identifies functional and
570 phylogenetic diversity of basidiomycete adenylate-forming reductases. Fungal
571 Genet Biol 112:55–63.
572 32. Rouhiainen L, Jokela J, Fewer DP, Urmann M, Sivonen K. 2010 Two alternative
573 starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin
574 variants in *Anabaena* (Cyanobacteria). Chem Biol 17:265–273.
575 33. Kleigrewe K, Almaliti J, Tian IY, Kinnel RB, Korobeynikov A, Monroe EA, Duggan
576 BM, Di Marzo V, Sherman DH, Dorrestein PC, Gerwick L, Gerwick WH. 2015.
577 Combining mass spectrometric metabolic profiling with genomic analysis: A
578 powerful approach for discovering natural products from cyanobacteria. J Nat Prod
579 78:1671–1682.
580 34. Voráčová K, Hájek J, Mareš J, Urajová P, Kuzma M, Cheel J, Villunger A, Kapuscik
581 A, Bally M, Novák P, Kabeláč M, Krumschnabel G, Lukeš M, Voloshko L,
582 Kopecký J, Hrouzek P. 2017. The cyanobacterial metabolite nocuolin A is a
583 natural oxadiazine that triggers apoptosis in human cancer cells. Plos One
584 12:e0172850.
585 35. He J, Hertweck C. 2004. Biosynthetic origin of the rare nitroaryl moiety of the
586 polyketide antibiotic aureothin: Involvement of an unprecedented N-oxygenase. J
587 Am Chem Soc 126:3694–3695.
588 36. Choi YS, Zhang HJ, Brunzelle JS, Nair SK, Zhao HM. 2008. In vitro reconstitution
589 and crystal structure of p-aminobenzoate N-oxygenase (AurF) involved in
590 aureothin biosynthesis. P Natl Acad Sci USA 105:6858-6863.

591 37. Fewer DP, Wahlsten M, Osterholm J, Jokela J, Rouhiainen L, Kaasalainen U,
592 Rikkinen J, Sivonen K. 2013. The genetic basis for O-Acetylation of the
593 microcystin toxin in cyanobacteria. Chem Biol 20:861–869.
594 38. Murray IA, Shaw WV. 1997. O-acetyltransferases for chloramphenicol and other
595 natural products. Antimicrob Agents Chemother 41:1–6.
596 39. Zapomělová E, Jezberová J, Hrouzek P, Hisem D, Řeháková K, Komárková J. 2009.
597 Polyphasic characterization of three strains of *Anabaena reniformis* and
598 *Aphanizomenon aphanizomenoides* (cyanobacteria) and their reclassification to
599 *Sphaerospermum* gen. nov. (incl. *Anabaena kisseleviana*). J Phycol 45:1363–1373.
600 40. Skulberg OM, Mysterud I, Karlsen J, Tønnesen HH, Laane CMM, Schumacher T.
601 2012. Alveld research per annum 2012: Searchlight on cyanobacteria we have
602 minor knowledge of. (In Norwegian). Biolog 30: 32–41.
603 41. Tønnesen HH, Mysterud I, Karlsen J, Skulberg OM, Laane CMM., Schumacher T.
604 2013. Identification of singlet oxygen photosensitizers in lambs drinking water in
605 an alveld risk area in West Norway. J Photochem Photobiol B 119: 37–45.
606 42. Hegge AB, Mysterud I, Karlsen J, Skulberg OM, Laane CMM, Schumacher T,
607 Tønnesen HH. 2013. Impaired secondary oxidant deactivation capacity and
608 enhanced oxidative stress in serum from alveld affected lambs. J Photochem
609 Photobiol B 126: 126–134.
610 43. Heinze R. 1996. A biotest for hepatotoxins using primary rat hepatocytes. Phycologia
611 35 (Supplement), 89–93.
612 44. Skulberg, OM. 1996. Toxins produced by cyanophytes in Norwegian inland waters –
613 health and environment. – In Chemical data as a basis of geomedical
614 investigations. (J. Låg, J. Ed), pp. 197–216, The Norwegian Academy of Science
615 and Letters.

45. Jones CG, Firn RD. 1991. On the evolution of plant secondary chemical diversity. Phil Trans R Soc Lond B 333:273–280.

46. Firn RD, Jones CG. 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.

48. Skulberg R, Skulberg OM. 1990. Research with algal cultures – NIVA's Culture Collection of Algae. *Norwegian Institute for Water Research*, 1–32. ISBN: 82–577–1743–6.

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.

50. Delcher AL, Bratke KA, Powers EC, Salzberg SL. 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics 23:673–679.

51. 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.

52. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri 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.

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.

54. Wang MX, 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 WT, Criisemann M, Boudreau PD, Esquenazi E, Sandoval-Calderon M, Kersten RD, Pace LA, Quinn RA, Duncan KR, Hsu CC, 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 CC, Yang YL, Humpf HU, Maansson M, Keyzers RA, Sims AC, Johnson AR, Sidebottom AM, Sedio BE, *et al.* 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat Biotechnol 34:828–837.

55. Johansen JR, Bohunická M, Lukešová A, Hrčková 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.

56. Tamrakar A. 2016. Isolation of benthic cyanobacteria and screening of bioactivities and natural products from culture collection strains. Master Thesis. University of Helsinki, Faculty of Agriculture and Forestry, Department of Food and Environmental Sciences.

663 57. Kantz T., Bold HC. 1969. Phycological studies IX. Morphological and taxonomic
664 investigations of *Nostoc* and *Anabaena* in culture. Publication no. 6924. University
665 of Texas, Austin.
666 58. Skulberg OM, Mysterud I, Karlsen J, Třnnesen HH, Laane CMM, Schumacher T.
667 2012. Alveld research per annum 2012: Searchlight on cyanobacteria we have
668 minor knowledge of. (In Norwegian). Biolog 30: 32-41.
669

Table 1. Strains analyzed for PUW/MIN production

Strain No.	Strain	Isolated by	Date	Locality	Reference
1	*Cylindrospermum alatosporum* CCALA 988	A. Lukešová	1989	Canada, Manitoba, Riding Mountain National Park, soil	Johansen *et al.* 2014[55]
2	*Cylindrospermum moravicum* CCALA 993	A. Lukešová	2008	Czech Republic, South Moravia, Moravian Karst, Amaterska Cave, cave sediment	Johansen *et al.* 2014[55]
3	*Cylindrospermum alatosporum* CCALA 994	A. Lukešová	2011	Czech Republic, Moravian Karst, earthworms collected from soil above Amaterska Cave, earthworm casings	Johansen *et al.* 2014[55]
4	*Anabaena* sp. UHCC-0399	M. Wahlsten	N/A	Finland, Jurmo, Southwestern Archipelago National Park, copepods	Tamrakar 2016[56]
5	*Anabaena minutissima* UTEX B1613	T. Kantz	1967	South Texas, USA, soil	Kantz & Bold 1969[57]
6	*Symplocastrum muelleri* NIVA-CYA 644	O.M. Skulberg	2009	Norway; Møre og Romsdal county; Halsa municipality, western slope of Slettfjellet mountain in semiterrestrial alpine habitat, biofilm on turf in ombrotrophic blanket bog	Skulberg *et al.* 2012[58]

677 **Table 2. Deduced proteins encoded by the *puw* gene cluster in six cyanobacterial strains,**

678 **including length and functional annotation. ACP, acyl carrier protein; FAAL, fatty acyl-**

679 **AMP ligase; PKS, polyketide synthase; NRPS, non-ribosomal peptide synthetase.**

Protein		Strain No.						Predicted Function
		1	2	3	4	5	6	
ORF1		659	664	664	643	643	647	ABC transporter
PuwA		2870	2870	2870	2854	2854	2866	NRPS
ORF2		1116	1499	1875	643	670	376	patatin-like phospholipase
ORF3		-	-	-	696	696	-	dynamin family protein
PuwI		-	-	-	709	702	711	FAAL, ACP
PuwJ		-	-	-	427	427	529	cytochrome-like protein
PuwB		2534	2592	2592	2549	2537	2555	hybrid PKS/NRPS, aminotransferase, oxygenase
PuwC	**Length (aa)**	597	590	590	597	589	-	FAAL
PuwD		101	104	96	93	92	-	ACP
PuwK		-	-	-	-	465	-	halogenase
PuwE		3077	3121	3121	3099	3112	3113	NRPS
PuwF		2370	5851	5851	5877	5871	3310	NRPS
PuwG		3492					2620	NRPS
PuwH		1102	1081	1102	1121	1121	1408	NRPS
PuwL		-	-	-	-	-	217	*O*-acetyltransferase

Table 3. Fragmentation of PUW variants from *Symplocastrum muelleri* strain 6 bearing acetyl substitutions on the FA moiety revealed by high energy (100 eV) fragmentation, and amino acid composition deduced by fragmentation at 60 eV.

	X=Ala, Y=Thr, FA=C_{16}			X=Gly, Y=Thr, FA=C_{18}			X=Ala, Y=Thr, FA=C_{18}			X=Gly, Y=Val, FA=C_{18}			X=Ala, Y=Val, FA=C_{18}		
Low fragmentation energy (60eV)	*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
$[M]^+$	**1265.7338**	**+0.7**	**$C_{59}H_{101}N_{12}O_{18}$**	**1279.7496**	**+0.9**	**$C_{60}H_{103}N_{12}O_{18}$**	**1293.7654**	**+0.8**	**$C_{61}H_{105}N_{12}O_{18}$**	**1277.7695**	**+1.6**	**$C_{61}H_{105}N_{12}O_{17}$**	**1291.7870**	**+0.1**	**$C_{62}H_{107}N_{12}O_{17}$**
$[M\text{-}CH_3OH]^+$	1233.7170	-6.6	$C_{58}H_{97}N_{12}O_{17}$	1247.7194	+4.1	$C_{59}H_{99}N_{12}O_{17}$	1261.7494	-7.3	$C_{60}H_{101}N_{12}O_{17}$	low int.		$C_{60}H_{101}N_{12}O_{16}$	low int.		$C_{61}H_{103}N_{12}O_{16}$
$[M\text{-}CH_3OH\text{-}NMeAsn]^+$	1105.6558	-4.9	$C_{53}H_{89}N_{10}O_{15}$	1119.6619	+3.7	$C_{54}H_{91}N_{10}O_{15}$	1133.681	+0.6	$C_{55}H_{93}N_{10}O_{15}$	1117.6924	-5.0	$C_{55}H_{93}N_{10}O_{14}$	1131.7307	-25.0	$C_{56}H_{95}N_{10}O_{14}$
$[M\text{-}CH_3OH\text{-}NMeAsn\text{-}dhb]^+$	1022.6180	-4.7	$C_{49}H_{84}N_9O_{14}$	1036.6365	-7.3	$C_{50}H_{86}N_9O_{14}$	1050.6478	-3.1	$C_{51}H_{88}N_9O_{14}$	1134.6603	-10.3	$C_{51}H_{88}N_9O_{13}$	1048.6671	-1.7	$C_{52}H_{90}N_9O_{13}$
$[M\text{-}CH_3OH\text{-}NMeAsn\text{-}dhb\text{-}X]^+$	951.5785	-2.5	$C_{48}H_{83}N_8O_{13}$	979.589	+18.8	$C_{48}H_{83}N_8O_{13}$	979.6041	+3.4	$C_{48}H_{83}N_8O_{13}$	977.6481	-20.5	$C_{49}H_{85}N_8O_{12}$	977.6518	-24.1	$C_{49}H_{85}N_8O_{12}$
$[M\text{-}CH_3OH\text{-}NMeAsn\text{-}dhb\text{-}X\text{-}Gln]^+$	823.5253	-9.4	$C_{41}H_{71}N_6O_{11}$	851.5473	+1.8	$C_{43}H_{75}N_6O_{11}$	851.5478	+1.2	$C_{43}H_{75}N_6O_{11}$	849.5838	-16.7	$C_{44}H_{76}N_6O_{10}$	849.5589	+12.5	$C_{44}H_{77}N_6O_{10}$
$[M\text{-}CH_3OH\text{-}NMeAsn\text{-}dhb\text{-}X\text{-}Gln\text{-}Y]^+$	722.4729	-4.2	$C_{37}H_{64}N_5O_9$	750.5005	+0.9	$C_{39}H_{68}N_5O_9$	750.5147	-18.1	$C_{39}H_{68}N_5O_9$	low int.		$C_{40}H_{72}N_5O_8$	low int.		$C_{40}H_{72}N_5O_8$
$[M\text{-}CH_3OH\text{-}NMeAsn\text{-}dhb\text{-}X\text{-}Gln\text{-}Y\text{-}Thr]^+$	621.4223	-0.2	$C_{33}H_{57}N_4O_7$	649.4526	+1.4	$C_{35}H_{61}N_4O_7$	649.4539	-0.6	$C_{35}H_{61}N_4O_7$	649.465	-17.8	$C_{35}H_{61}N_4O_7$	649.4483	+8.0	$C_{35}H_{61}N_4O_7$
High fragmentation energy (100eV)															
Fragment 1	411.3208	+2.2	$C_{23}H_{43}N_2O_4$	439.3559	-6.5	$C_{25}H_{47}N_2O_4$	439.3556	-5.8	$C_{25}H_{47}N_2O_4$	439.3556	-5.8	$C_{25}H_{47}N_2O_4$	439.3508	+5.1	$C_{25}H_{47}N_2O_4$
Fragment 1 - $C_2H_4O_2$	351.3006	+0.0	$C_{21}H_{39}N_2O_2$	379.3334	-4.0	$C_{23}H_{43}N_2O_2$	379.3329	-2.6	$C_{23}H_{43}N_2O_2$	379.3328	-2.4	$C_{23}H_{43}N_2O_2$	379.3360	-10.8	$C_{23}H_{43}N_2O_2$
Fragment 2	284.2583	+0.4	$C_{17}H_{34}NO_2$	312.2919	-6.9	$C_{19}H_{38}NO_2$	312.2892	+1.6	$C_{19}H_{38}NO_2$	low int.		$C_{19}H_{38}NO_2$	low int.		$C_{19}H_{38}NO_2$
Fragment 2 - $C_2H_4O_2$	224.2367	+2.6	$C_{15}H_{30}N$	252.2686	0.0	$C_{17}H_{34}N$	252.2687	-0.5	$C_{17}H_{34}N$	252.2684	+0.7	$C_{17}H_{34}N$	252.2677	+3.5	$C_{17}H_{34}N$

684 **Table 4. Bacterial and yeast strains used for antimicrobial testing of PUW F and MIN A,**
685 **C, and D. HAMBI, culture collection of University of Helsinki, Faculty of Agriculture**
686 **and Forestry, Department of Microbiology.**

Test organisms (HAMBI nr.)	Media[a]	Incubation temp. (°C)	Incubation time (h)	Gram strain reaction (+\-)
Pseudomonas sp. (2796)	TGY	28	24	-
Micrococcus luteus (2688)	TGY	28	24	+
Bacillus subtilis (251)	TGY	28	24	+
Pseudomonas aeruginosa (25)	TGY	37	24	-
Escherichia coli (396)	TGY	37	24	-
Bacillus cereus (1881)	TSA	28	24	+
Burkholderia cepacia (2487)	TSA	37	24	-
Staphylococcus aureus (11)	TSA	37	24	+
Xanthomonas campestris (104)	NA	28	24	-
Burkholderia pseudomallei (33)	NA	37	24	-
Salmonella typhi (1306)	NA	37	24	-
Arthrobacter globiformis (1863)	NA	28	24	-
Kocuria varians (40)	NA	28	24	+
Candida albicans (261)	YM agar	37	24	yeast
Cryptococcus albidus (264)	YM agar	28	24	yeast
Saccharomyces cerevisiae (1164)	YM agar	28	24	yeast

687

688 [a]The composition of all media was obtained from the American Type Culture Collection (ATCC).

689 TGY, tryptone glucose yeast; TSA, tryptic soy agar; NA, Nutrient agar; YM agar, yeast malt agar.

690

691 **Figure legends**

692

693 **Fig. 1.** HPLC-HRMS/MS analysis of crude extracts from the investigated strains. Major
694 puwainaphycin (PUW) and minutissamide (MIN) variants are highlighted. Numbers in
695 brackets following the peak designation refer to the corresponding Supplementary Figure
696 containing full MS/MS data. For variants without complete structural information, only *m/z*
697 values are shown.

698 **Fig. 2.** Molecular network created using the Global Natural Products Social Molecular
699 Networking (GNPS) web platform. Two separate networks were obtained during GNPS
700 analysis; (a) a group containing *Cylindrospermum* strains 1–3 and *Anabaena* strains 4–5, and
701 (b) a group containing only variants detected in *Symplocastrum muelleri* strain 6. The separate
702 groups differ mainly in the peptide core of the molecule. Numbers in brackets following the
703 peak designation refer to the corresponding Supplementary Figure containing full MS/MS
704 data. For variants without complete structural information, only *m/z* values are shown. (*)
705 refers to compounds present in trace amounts and (#) refers to compounds for which MS/MS
706 data failed to resolve the structural information.

707 **Fig. 3.** Structural variability of the peptide core of PUW/MIN variants. Examples of
708 structural variants PUW F (a) and PUW A (b) with designated aminoacid positions
709 representing the two major peptide cores. (c) Table summarizing all types of the PUW/MIN
710 peptide core found in known compounds reported in literature and compounds (Comp.)
711 detected in studied strains. Columns shaded in grey highlight the conserved aminoacid
712 positions.

713 **Fig. 4.** Structural variability of the FA moiety of PUW/MIN variants. The relative proportion
714 of variants with differences in FA length and substitution (y-axis) is depicted using a color
715 scale (z-axis). For comparison, the peak area of a given variant was normalized against the

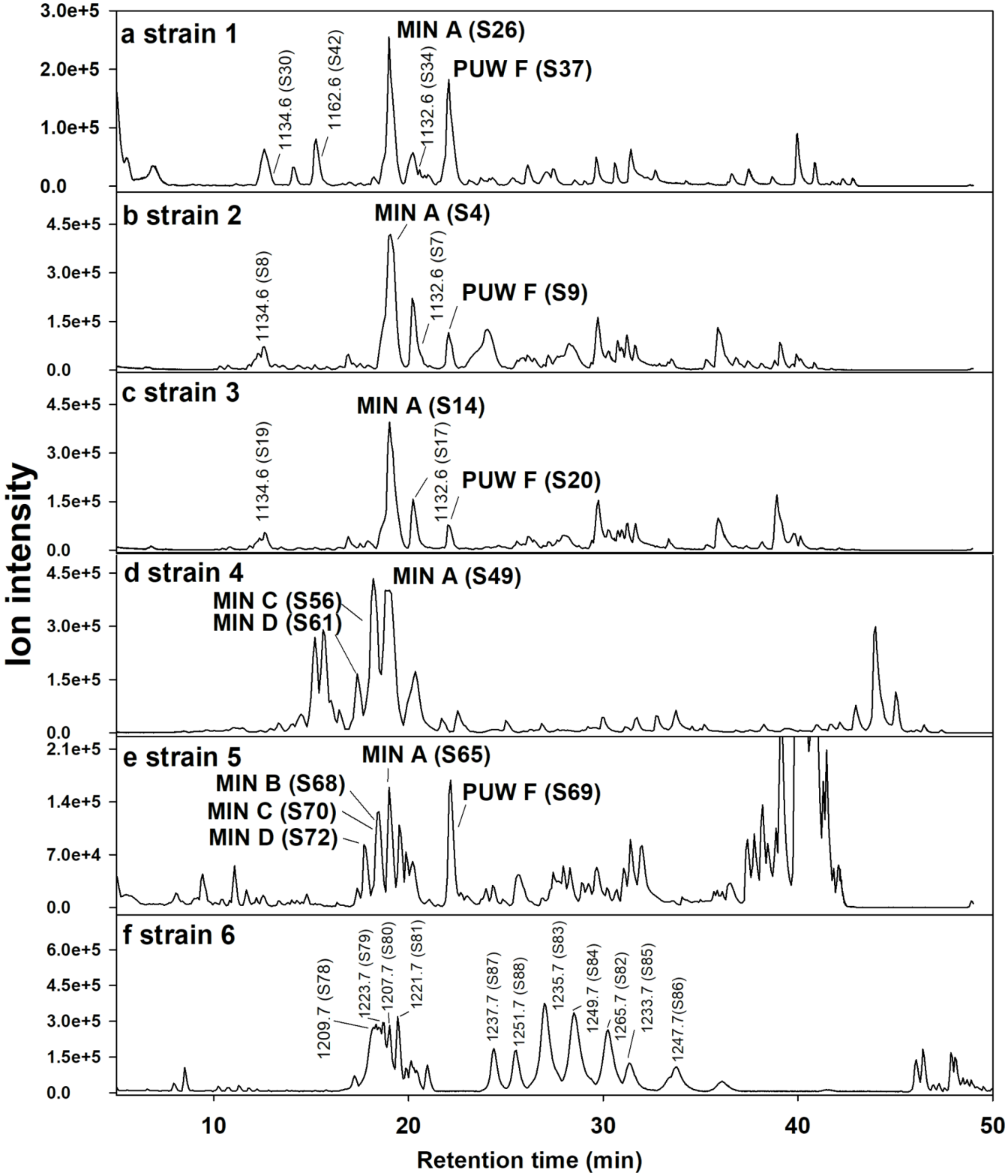
716 peak area of the major variant present in the strain (MIN A for strains 1–5, and *m/z* 1235.7 for
717 strain 6).

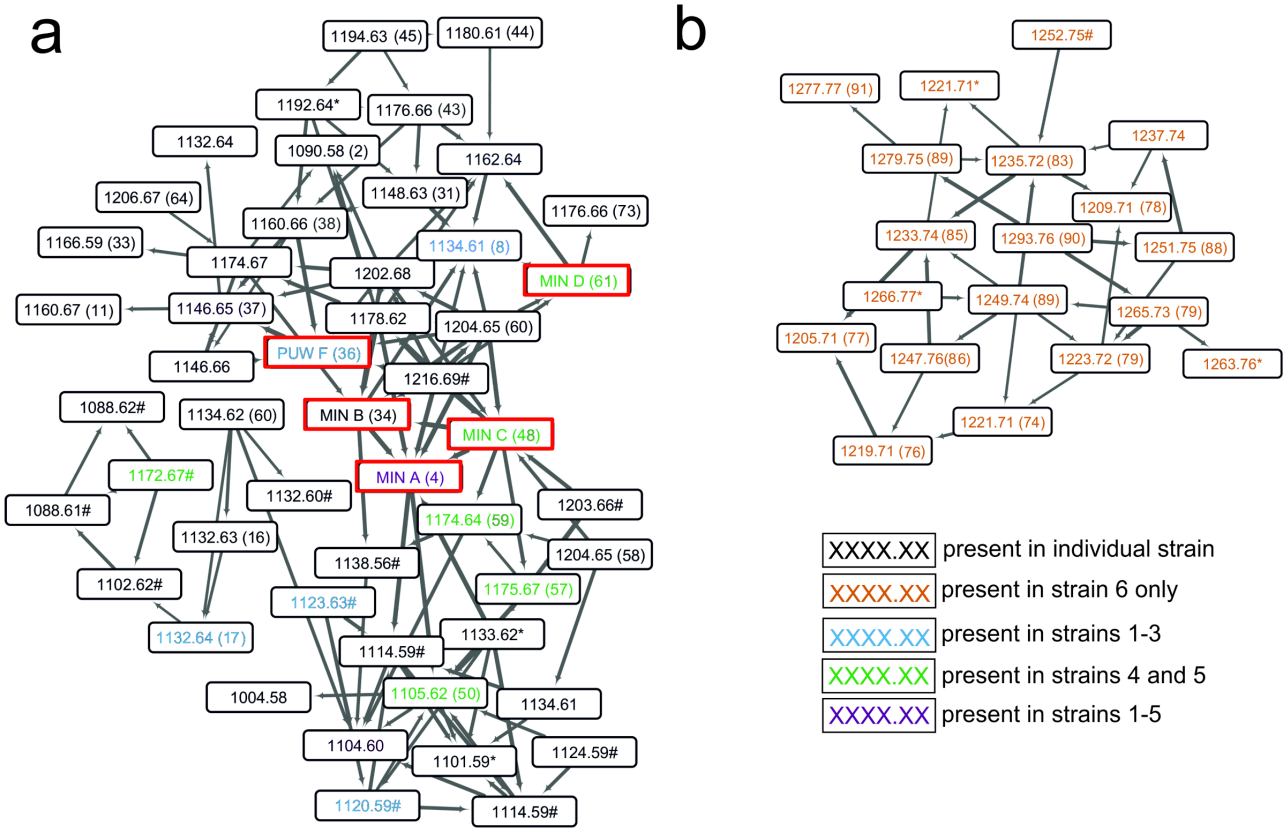
718 **Fig. 5.** Structure of the *puw* gene cluster in the six investigated cyanobacterial strains. Gene
719 arrangement and functional annotation of *puwA–L* genes and selected PKS/NRPS tailoring
720 domains is indicated by colored arrows. The distribution of the two observed types of putative
721 starter modules (shaded boxes) is indicated by bars.

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

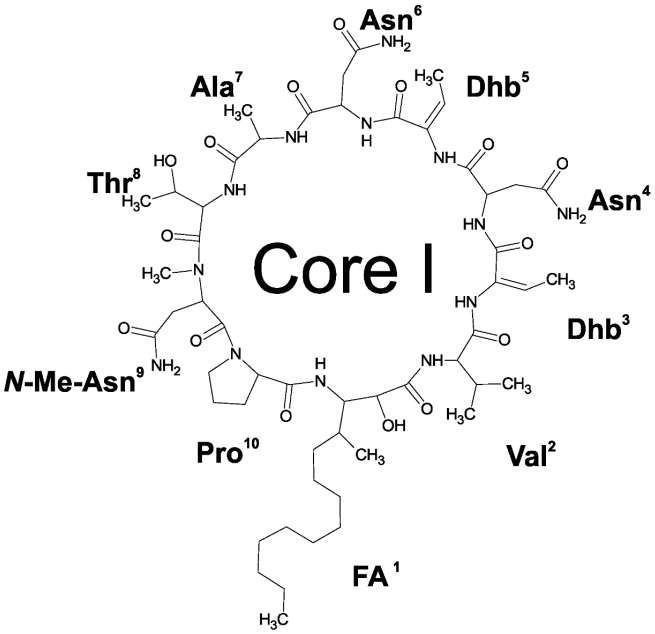
730 **Fig 7.** MS/MS fragmentation of MIN A (a, c, e) and the PUW variant at *m/z* 1279 bearing an
731 acetyl substitution of the fatty acid chain (b, d, f). (a–b) base peak chromatograms; (c–d)
732 fragmentation of the protonated molecule at low fragmentation energy, yielding *b* series of
733 ions corresponding to the losses of particular amino acid residues, (e–f) fragmentation of the
734 protonated molecule at high energy (100 eV) yielding fragments characteristic for the ß-
735 amino fatty acid.

736 **Fig. 8.** Antifungal activity of PUW F against yeast strains (a) *Saccharomyces cerevisiae*
737 HAMBI 1164 and (b) *Candida albicans* HAMBI 261. Discs were treated with a concentration
738 range from 25.2 µg mL^{-1} to 0.0394 µg/mL to determine the minimum inhibitory concentration
739 (MIC). Numbers represent concentrations: (1) = 25.2 µg mL^{-1}; (2) = 12.6 µg mL^{-1}; (3) = 6.3
740 µg mL^{-1}; (+) = positive control (10 µg of nystatin). (-) = negative control (10 µL of methanol).



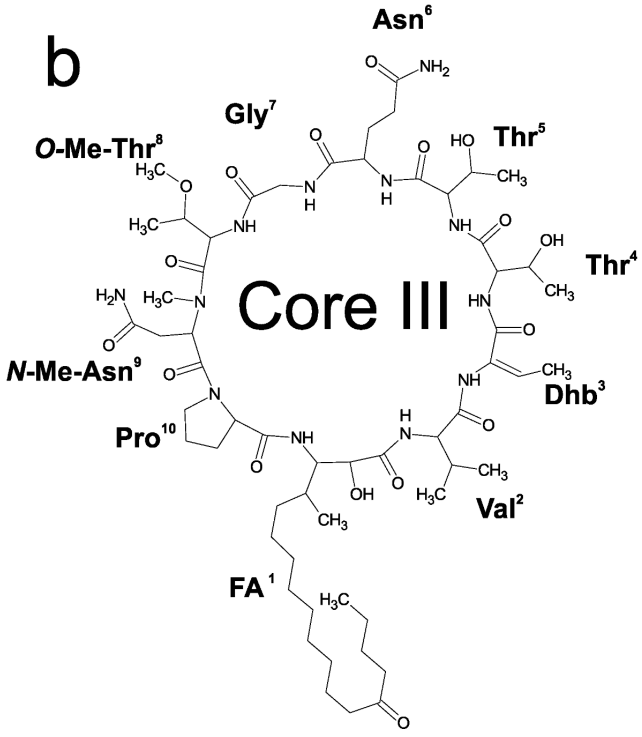


a



PUW F

b



PUW A

c

	AA_1	AA_2	AA_3	AA_4	AA_5	AA_6	AA_7	AA_8	AA_9	AA_{10}	known compounds	variants found in strains 1-6
Core I	**FA**	**Val**	**Dhb**	Asn	Dhb	Asn	Ala	Thr	***N*-Me-Asn**	**Pro**	MIN A, MIN B, MIN C, MIN D, PUW F	**MIN A, MIN B, MIN C, MIN D, PUW F,** Comp.1,3,10,11,12,14,16,19,20, 25,26,28,30,31,37,45,51
Core II	**FA**	**Val**	**Dhb**	Gln	Dhb	Asn	Ala	Thr	***N*-Me-Asn**	**Pro**	PUW G	**PUW G,** Comp.2,6,13,15,17,27,29,32
Core III	**FA**	**Val**	**Dhb**	Thr	Thr	Gln	Gly	*O*-Me-Thr	***N*-Me-Asn**	**Pro**	PUW A, PUW C, PUW E, MIN I, MIN K	Comp. 38,46,50,52,53,56,58
Core IV	**FA**	**Val**	**Dhb**	Thr	Thr	Gln	Ala	*O*-Me-Thr	***N*-Me-Asn**	**Pro**	MIN E, MIN G	Comp.39,47,53,57,59
Core V	**FA**	**Val**	**Dhb**	Thr	Val	Gln	Gly	*O*-Me-Thr	***N*-Me-Asn**	**Pro**	PUW B, PUW D, MIN L	Comp. 40,48,54,60
Core VI	**FA**	**Val**	**Dhb**	Thr	Val	Gln	Ala	*O*-Me-Thr	***N*-Me-Asn**	**Pro**	MIN H	Comp. 41,49,55,61
Core VII	**FA**	**Val**	**Dhb**	Asn	Dhb	Asn	Gly	Thr	***N*-Me-Asn**	**Pro**		Comp. 7,36,44
Core VIII	**FA**	**Val**	**Dhb**	Asn	Dhb	Asn	Ser	Thr	***N*-Me-Asn**	**Pro**		Comp. 8,23,35
Core IX	**FA**	**Val**	**Dhb**	Asn	Dhb	Asn	Ala	Ser	***N*-Me-Asn**	**Pro**		Comp. 9,18,24
Core X	**FA**	**Val**	**Dhb**	Thr	Dhb	Asn	Ala	Thr	***N*-Me-Asn**	**Pro**		Comp. 5,34,43