

Article

# Discovery of unusual cyanobacterial tryptophan-containing anabaenopeptins by MS/MS based molecular networking

Subhasish Saha<sup>1#</sup>, Germana Esposito<sup>2#</sup>, Petra Urajová<sup>1</sup>, Jan Mareš<sup>1,3</sup>, Daniela Ewe<sup>1</sup>, Alessia Caso<sup>2</sup>, Markéta Macho<sup>1,4</sup>, Kateřina Delawská<sup>1,4</sup>, Andreja Kust<sup>1,3</sup>, Pavel Hrouzek<sup>1</sup>, Josef Juráň<sup>4,5</sup>, Valeria Costantino<sup>3\*</sup> and Kumar Saurav<sup>1\*</sup>

<sup>1</sup>Laboratory of Algal Biotechnology-Centre Algatech, Institute of Microbiology of the Czech Academy of Sciences, Třeboň, Czech Republic, [saha@alga.cz](mailto:saha@alga.cz) (S.S.); [urajova@alga.cz](mailto:urajova@alga.cz) (P.E.); [ewe@alga.cz](mailto:ewe@alga.cz) (D.E.); [macho@alga.cz](mailto:macho@alga.cz) (M.M.); [delawska@alga.cz](mailto:delawska@alga.cz) (K.D.); [kust@alga.cz](mailto:kust@alga.cz) (A.K.); [sauravverma17@gmail.com](mailto:sauravverma17@gmail.com) (K.S.)

<sup>2</sup>TheBlue Chemistry lab, Università degli Studi di Napoli, task Force "BigFed2, Napoli, Italy, [germana.esposito@unina.it](mailto:germana.esposito@unina.it) (G.E.); [alessia.caso@unina.it](mailto:alessia.caso@unina.it) (A.C.); [valeria.costantino@unina.it](mailto:valeria.costantino@unina.it) (V.C.).

<sup>3</sup>Biology Centre of the Czech Academy of Sciences, Institute of Hydrobiology, České Budějovice, Czech Republic. [andreja.kust@gmail.com](mailto:andreja.kust@gmail.com) (A.K.), [jan.mares@centrum.cz](mailto:jan.mares@centrum.cz) (J.M.)

<sup>4</sup>University of South Bohemia, Faculty of Science, Branišovská 1760, České Budějovice, Czech Republic. [delawska@alga.cz](mailto:delawska@alga.cz) (K.D.); [macho@alga.cz](mailto:macho@alga.cz) (M.M.)

<sup>5</sup>Institute of Botany of the Czech Academy of Sciences, Třeboň, Czech Republic, [juran.josef@seznam.cz](mailto:juran.josef@seznam.cz) (J.J.).

#Equal contribution

\*Correspondence: [valeria.costantino@unina.it](mailto:valeria.costantino@unina.it) (V.C.), [sauravverma17@gmail.com](mailto:sauravverma17@gmail.com) (K.S.); Tel.: (+420-737-003-359 (K.S.).

## Abstract:

Heterocytous cyanobacteria are among the most prolific source of bioactive secondary metabolites, including anabaenopeptins (APTs). A terrestrial filamentous *Brasilonema* sp. CT11 collected in Costa Rica bamboo forest, as black mat was studied using a multidisciplinary approach: genome mining and HPLC-HRMS/MS coupled with bionformatic analyses. Herein, we report the nearly complete genome consisting 8.79 Mbp with a GC content of 42.4%. Moreover, we report on three novel tryptophane-containing APTs; anabaenopeptin 788 (**1**), anabaenopeptin 802 (**2**) and anabaenopeptin 816 (**3**). Further, the structure of two homologues, i.e., anabaenopeptin 802 (**2a**) and anabaenopeptin 802 (**2b**) was determined by spectroscopic analysis (NMR and MS). Both compounds were shown to exert weak to moderate antiproliferative activity against HeLa cell lines. This study also provides the unique and diverse potential of biosynthetic gene clusters and an assessment of the predicted chemical space yet to be discovered from this genus.

**Keywords:** *Brasilonema*; Anabaenopeptins; hexapeptides; tryptophan-containing peptides; molecular networking; antiproliferative activity.

## 1. Introduction

For 3.5 billion years, cyanobacteria have kept developing new adaptation skills and defense mechanisms which allowed them to colonize the Earth, shape its atmosphere from anoxic to oxygenic, and ever since to survive in harsh and strongly competitive environments under extreme temperatures, salt stress, high UV-radiation, and pathogen attack [1,2]. One of these survival strategies is the production of a vast variety of secondary metabolites, which exhibit a broad spectrum of biological activities and properties, including peptides, lipopeptides, polyketides, alkaloids, lipids, and terpenoids [3-7]. When growth conditions are advantageous, cyanobacteria proliferate and form harmful blooms, known as cyanobacterial blooms (CBs), that can be risky for the aquatic life as well as for human health, due to the toxins that they can produce [8,9]. Cyanobacteria comprise a wide range of phylogenetic lineages, of which the monophyletic order Nostocales, (heterocytous cyanobacteria) are extremely competent group in terms of general secondary metabolite production [10]. Filamentous cyanobacteria, in particular the genera *Anabaena*, *Nostoc*, *Oscillatoria*, and *Lyngbya*, produce large variety of secondary metabolites, estimated to comprise around 25% of the total number of known cyanobacterial metabolites [11-14]. These compounds are usually produced by nonribosomal peptide synthetases (NRPS), polyketide synthases (PKS), or NRPS/PKS hybrid enzymes. The last advancement in the field of genome sequencing and the rise of fast and extremely accurate analytical and bioinformatics methods, such as high-performance liquid chromatography connected to high-resolution mass spectrometry (HPLC-HRMS/MS), has accelerated the discovery of novel secondary metabolites [e.g., 5,6,15]. The use of molecular networking as a dereplication strategy allow to identify the complexity of variants produced by respective biosynthetic genes, which in turn produce valuable data for pharmaceutical development, since structural variants often have different biological activities [16-18].

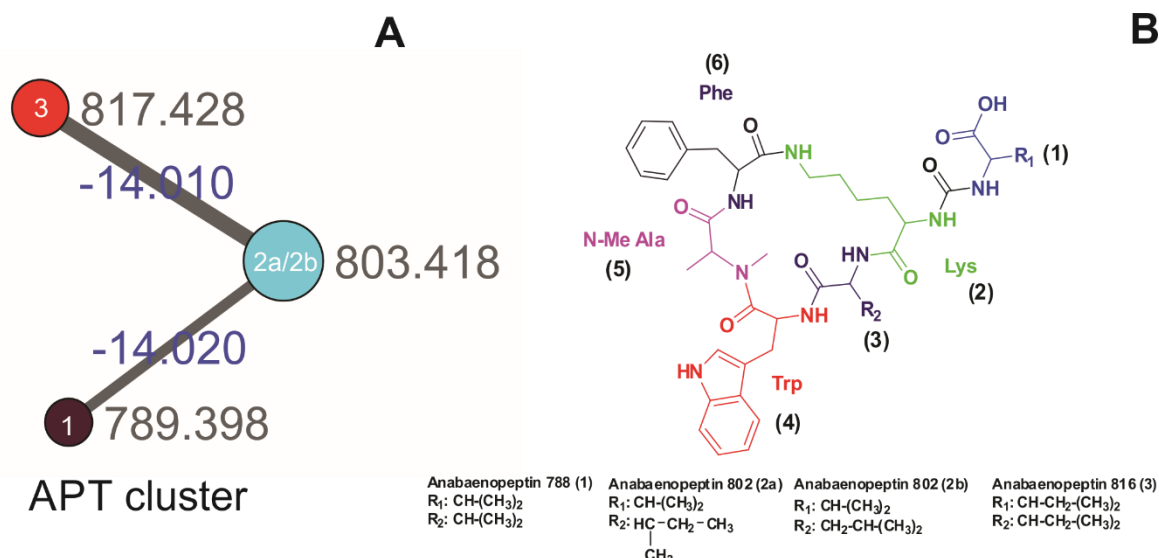
Anabaenopeptins (APTs) are a family of cyclic hexapeptides produced by diverse cyanobacterial taxa, firstly described from planktic representative *Anabaena flos-aqae* NRC 525-17 [19]. APTs manifest general structure  $X^1\text{-CO-[Lys}^2\text{-X}^3\text{-X}^4\text{-MeX}^5\text{-X}^6]$ , wherein the X1 and X3-6 are non-proteinogenic or proteinogenic amino acids. The cyclization bond is formed between the C-terminal carboxyl and the N-terminal amine of the conserved D-Lys at position 2 [20]. The exocyclic amino-acid X1 is connected to the  $\alpha$ -amino group of the conserved Lys residue via the ureido bond. APTs encoded in the biosynthetic gene cluster firstly identified from *Planktothrix rubescens* NIVA CYA 98 [21] are produced by NRPS following a sequential synthesis pathway using (non) proteinogenic amino acids as substrate [22]. These clusters have been characterised in many strains belonging to various genera such as *Anabaena*, *Nostoc*, *Nodularia* and *Planktothrix* [21-24]. Large structural variability is observed among APTs with members exhibiting masses between 679 and 956 Da with various biological activities, including inhibition of protein phosphatase, carboxypeptidases A and U, and other protease inhibitory activity [20]. Since the first study, more than 115 isoforms of APTs have been reported, of which 104 are produced by cyanobacteria, belonging to the genera *Anabaena*, *Aphanizomenon*, *Microcystis*, *Nodularia*, *Nostoc*, *Planktothrix*, *Oscillatoria*, *Schizothrix* and *Lyngbya* [25].

In the present study, we report the draft genome sequence of a new *Brasilonema* strain collected in Costa Rica. Further, the combined approach, genome mining and HPLC-HRMS/MS coupled

with bioinformatic analyses led us to identify four unusual novel tryptophan-containing APTs showing moderate antiproliferative activity.

## 2. Results and Discussion

Strains belonging to the genus *Brasilonema* is a good source for diverse secondary metabolites [26]. Genus *Brasilonema*, belonging to the family Scytonemataceae, are terrestrial filamentous, heterocytous false-branching cyanobacteria, that have been previously isolated from subtropical and tropical climate zones [27-29]. To the best of our knowledge, the only tryptophan-containing APTs were detected in 2015 from the genus *Brasilonema* sp. (CENA360 and CENA382) isolated from Brazilian Atlantic Coastal Forest [30], and a hydroxyl-tryptophan containing APT mozamide A, isolated from *Thenellide* sponge from Mozambique [31]. In this study, we analyzed a terrestrial filamentous *Brasilonema* sp. CT11 collected in Costa Rica, as black mat on a bamboo trunk. The crude extract (~16mg) was obtained upon one week of cultivation in 300ml glass column under continuous light condition and its subsequent solvent extraction. Crude extract (1mg/ml) was subjected to C-18 reversed-phase column chromatography and analyzed by LC-HRMS and LC-HRMS/MS on Impact HD instrument. The raw analytical data were elaborated and studied using molecular networking (Global natural product social networking, GNPS). The raw data was converted to mzXML format and directly imported to GNPS platform to analyze and dereplicated against publicly available data sets accessed via the GNPS-MassIVE database (<https://massive.ucsd.edu/>). To date, >3000 million MS/MS spectra from various instruments (e.g., Orbitrap and qTOF) have been searched at GNPS, yielding putative dereplication matches of 7.7 million spectra to more than 35,477 compounds. GNPS analyses yielded thirty clusters and one hundred and eight single nodes without grouping (Figure S1). The network was rearranged using Cytoscape v3.8.0 for better visualization. A schizopeptins related cluster was present in the network and is shown in Figure 1, forming a cluster representing three nodes with  $m/z$  817.428,  $m/z$  803.418 and  $m/z$  789.398. Further the manual curation and detailed MS/MS spectra study led us to interpret the putative presence of three novel tryptophane-containing APTs; anabaenopeptin 788 ( $m/z$  789.4425 [M+H]<sup>+</sup>), anabaenopeptin 802 ( $m/z$  803.4417 [M+H]<sup>+</sup>) and anabaenopeptin 816 ( $m/z$  817.4425 [M+H]<sup>+</sup>). The structures of these compounds were manually curated and postulated in accordance with the fragment ions listed in Table 1.



**Figure 1. A.** The molecular network with APT cluster obtained from the organic extract of *Brasilonema* sp. CT11. Nodes are labeled with parent  $m/z$  values and edges with the mass difference. Node size is indicative of the ion count; edge thickness is relative to cosine score. **B.** Four APTs detected using LC-HRMS/MS following general structure  $\text{X}^1\text{-CO-[Lys}^2\text{-X}^3\text{-X}^4\text{-MeX}^5\text{-X}^6]$ , wherein the  $\text{X}^1$  and  $\text{X}^3\text{-X}^6$  are non-proteinogenic or proteinogenic amino acids.

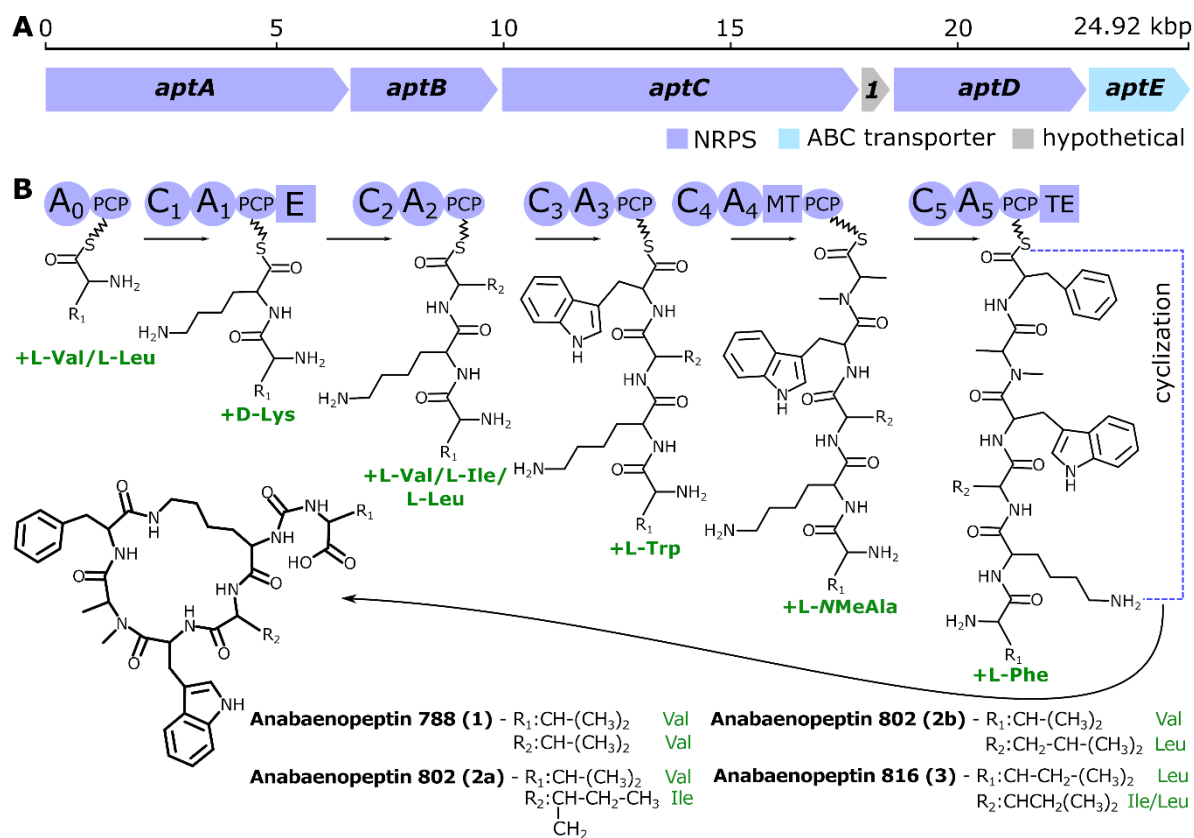
**Table 1:** Product ion spectra data for compounds **1** ( $m/z$  789.4425  $[\text{M}+\text{H}]^+$ ), **2a/2b** ( $m/z$  803.4417  $[\text{M}+\text{H}]^+$ ) and **3** ( $m/z$  817.4425  $[\text{M}+\text{H}]^+$ ).

Product Ion Assignment	1 ( $m/z$ )	error, ppm	2a/2b ( $m/z$ )	error, ppm	3 ( $m/z$ )	error, ppm
Lys fragment	84.0810	2.6	84.0808	0.7	84.0808	0.2
MeAla + CO + H <sup>+</sup>	114.0549	0.1	114.0551	1.1	114.0550	0.0
Trp fragment	130.0650	0.7	130.0653	1.0	130.0651	0.1
Trp-MeAla + H <sup>+</sup>	272.196	0.6	272.1399	2.1	272.1394	0.6
Trp-MeAla-Val + H <sup>+</sup>	371.2087	2.4	-	-	-	-
Trp-MeAla-Ile/Leu + H <sup>+</sup>	-	-	385.2249	3.9	385.2234	0.0
CO-Lys-Phe-MeAla + H <sup>+</sup>	405.2141	2.1	405.2143	2.5	405.2138	1.4
Val-CO-Lys-Phe-MeAla + H <sup>+</sup>	504.2811	1.0	504.2830	2.7	-	-
Ile/Leu-CO-Lys-Phe-MeAla + H <sup>+</sup>	-	-	-	-	518.2973	3.0
Val-CO-Lys-(Val)-(Phe-MeAla) + H <sup>+</sup>	603.3497	0.6	-	-	-	-
Val-CO-Lys-(Ile/Leu)-(Phe-MeAla) + H <sup>+</sup>	-	-	617.3677	4.0	-	-
Ile/Leu-CO-Lys-(Ile/Leu)-(Phe-MeAla) + H <sup>+</sup>	-	-	-	-	631.3797	2.7
Val-CO-[Lys-Val-Trp-MeAla-Phe] + H <sup>+</sup>	789.4304	1.3	-	-	-	-
Val-CO-[Lys-Ile/Leu-Trp-MeAla-Phe] + H <sup>+</sup>	-	-	803.4469	2.3	-	-
Ile/Leu-CO-[Lys-Ile/Leu-Trp-MeAla-Phe] + H <sup>+</sup>	-	-	-	-	817.4612	0.6

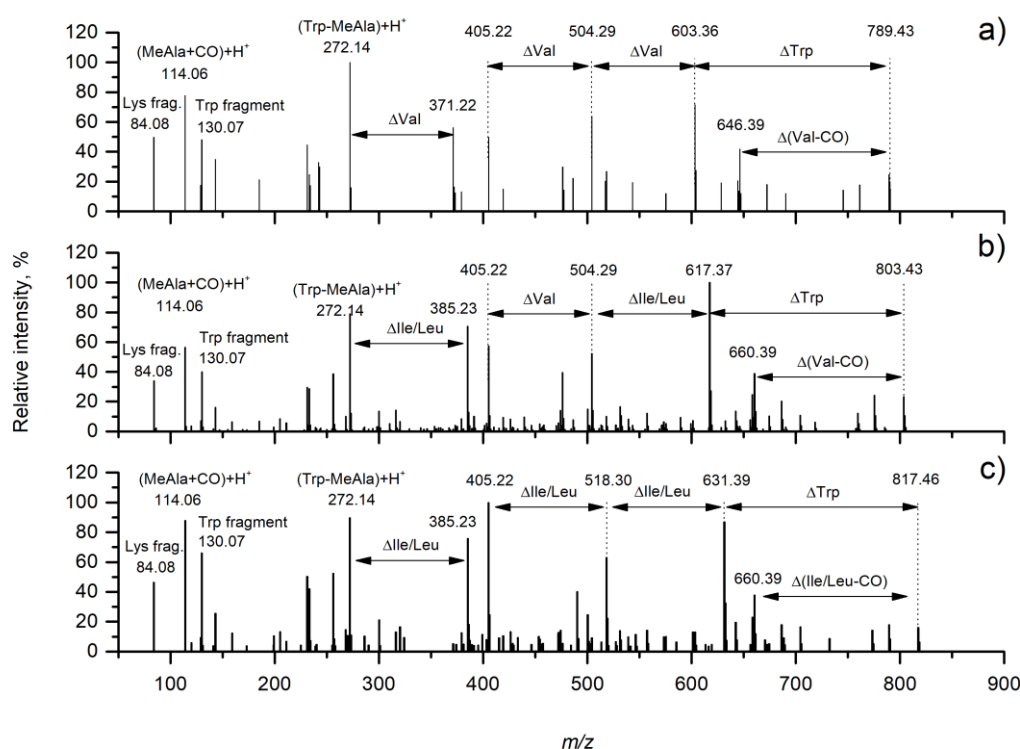
As a complementary approach to GNPS, we applied whole genome sequencing and subsequent genome mining to investigate the genetic potential of *Brasilonema* sp. CT11 for production of secondary metabolites. The newly sequenced draft genome of *Brasilonema* CT11 (NCBI accession number JABXYX010000000) was retrieved in 411 genomic scaffolds of

approximately 8.79 Mbp total length (99.48% estimated completeness) and 42.4% average GC content. The assembly was analyzed for the presence of secondary metabolite biosynthetic gene clusters (BGCs) using antiSMASH 5.1.1. AntiSMASH uses Hidden Markov Models (HMM) and rules-based detection to identify a broad array of BGCs, including those encoding polyketide, non-ribosomal peptides, terpenes, aminoglycosides and ribosomally synthesized and post-translationally modified peptides (RiPPs) from bacterial genomes. AntiSMASH predicted 36 putative secondary metabolite clusters from *Brasilonema* CT11. Out of 36 gene clusters, four clusters possess 100% similarity with other known BGCs whereas only two clusters were detected with more than 75% similarity with known BGCs (Table S1) [32]. The majority of these gene clusters were in the NRPS category with 12 recovered BGCs. These putative BGCs yielded hits among various known secondary metabolites showing variable similarity scores, ranging from maximum similarity with known natural product Anabaenopeptin 908/915 (100%) [22], to minimum similarity with nostocyclopeptide A2 (28%) [33], suggesting that the pathways may encode new natural products or natural products with no characterized BGCs. Next to NRPS the second most abundant BGCs were NRPS-PKS hybrid (6 clusters), off which one cluster showed 100% similarity to nostopeptolide A2 [34]. Another interesting hybrid BGC with 60% similarity to hapalosin BGC [35] was detected, which may lead to the discovery of a new cyclic peptide. Besides these, the other major BGCs detected were; one cluster of PKS type I, three terpenes, two RiPPs, four bacteriocin, and one indole. Among RiPPs class, one lassopeptide and one lantipeptide were detected with no similarity with any reported gene clusters. The PKS type I gene cluster, Merocyclophane C/D [36] was detected with highest similarity of 22%, which is particularly interesting due to its low similarity.

The discovery of a complete putative APT BGC (NCBI accession number MT670293) along with the detection of unusual Trp-containing ATP variants using MS/MS networking coupled with manual curation encouraged us to investigate this gene cluster in detail. The putative APT BGC was nearly 25 kbp long and consisted of five genes (*aptA-E*), clearly homologous to previously reported *apt* genes (Table S2., [23,25]) and a single short hypothetical open reading frame (Figure 3). Adenylation domains (A-domains) present in the six encoded NRPS modules (Table S3) showed predicted substrate specificities that were in agreement with the amino-acid residues observed in the APT variants detected herein (Figure 4, Table 1), including an unequivocally predicted Trp residue at position 4 of the peptide cycle. Intriguingly, *in silico* substrate prediction for the starter A-domain ( $A_0$ ) and the third A-domain ( $A_2$ ) reported only a broad specificity to hydrophobic aliphatic amino acids, with Val, Leu, and Ile among the most probable candidates based on algorithms implemented in antiSMASH. These results are in line with the observed substrate promiscuity, yielding APT variants having either Val or Leu at position 1 and Ile, Leu or Val at position 3. Previous reports have indicated both substrate promiscuity and occurrence of alternative starter modules to be responsible for variability in amino acids incorporated into APTs [22,23], however we have not detected any additional alternative starter module in our genomic scaffold.



**Figure 2.** Anabaenopeptin synthetase gene cluster arrangement of *Brasilonema* CT11 yielding four anabaenopeptins 1, 2a, 2b and 3. **A)** Gene map of the apt biosynthetic gene cluster. **B)** Proposed biosynthesis of anabaenopeptins in *Brasilonema* sp. CT11. A-adenylation domain; C-condensation domain; E-epimerization domain; MT-methyltransferase domain; PCP-peptidyl-carrier protein; TE-thioesterase domain.



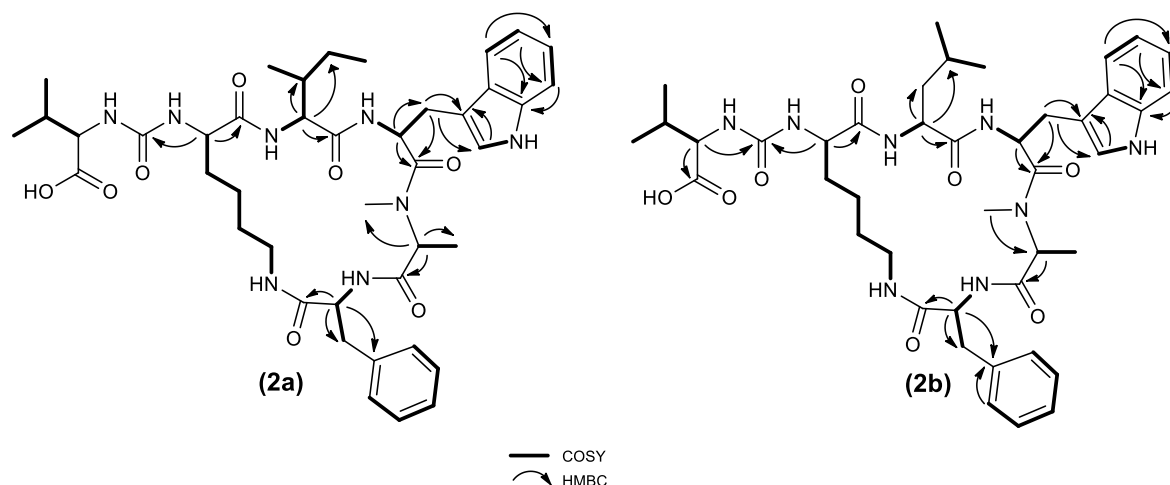
**Figure 3.** HR-MS/MS product ion spectra of protonated anabaenopeptins from *Brasilonema* CT11 a) compound 1, b) compound 2a/2b, and c) compound 3.

Further fractionation and manual curation of the spectra suggested us the presence of four unusual tryptophan containing anabaenopeptins (two homologues with having  $m/z$  803.4417  $[M + H]^+$ , and 803.4425  $[M + H]^+$ , one having  $m/z$  789.4425  $[M + H]^+$  and one with  $m/z$  817.4425  $[M + H]^+$ ) using LC-HRMS/MS. The structures of these compounds were postulated in accordance with the fragment ions listed in Table 1 (Figure 3). Further, subsequent chromatographical separation led us to isolate two homologues with  $m/z$  803.44  $[M + H]^+$ , **2a** (0.6 mg) and **2b** (2.6 mg).

The HR-ESI-MS mass spectrum of compound **2a** and **2b** showed  $[M+H]^+$  pseudomolecular ion peaks at  $m/z$  803.4417 and 803.4425, respectively, which defined its molecular formula as  $C_{42}H_{58}N_8O_8$ . Their MS/MS spectra were identical. The fragmentation pattern was indicative of two cyclic peptide compounds, with fragments  $m/z$  504.28 that indicated presence of Val at position 1, Lys at position 2, MeAla at position 5 and Phe at position 6; and  $m/z$  385.22 which indicated presence of Ile/Leu at position 3 and Trp at position 4. Binding of leucine or isoleucine cannot be distinguished by conventional mass spectrometry, so then the HPLC-purification was made to get the two pure compounds, to be analyzed by NMR. The structures of both the homologues were confirmed by extensive NMR analysis. A full set of homonuclear and heteronuclear two-dimensional NMR spectra (COSY, NOESY, HSQC, and HMBC) was recorded (Table S4 and S5, Figure S3-14). In particular, the proton spectra showed seven amide NH signals and six different  $\alpha$ -proton signals, as expected for an hexapeptide with a *N*-Methyl Alanine, a lysine and a tryptophan residue. Analysis of the TOCSY spectrum (Figure S8 and S14) showed correlation of all the protons of the six side chains with the corresponding  $\alpha$ -proton. This information combined with that derived from the analysis of COSY, HMBC and

NOESY spectra confirmed the structures of compounds **2a** and **2b**, the analysis of COSY spectra let to distinguish the presence of an isoleucine and a leucine residue in compound **2a** and **2b** respectively (Figure 4).

Taken together, MS evidence, <sup>1</sup>H-NMR and 2D-NMR data, allowed the identification of compounds **2a** and **2b**, and designated as anabaenopeptin 802a and anabaenopeptin 802b, respectively (Figure 4).

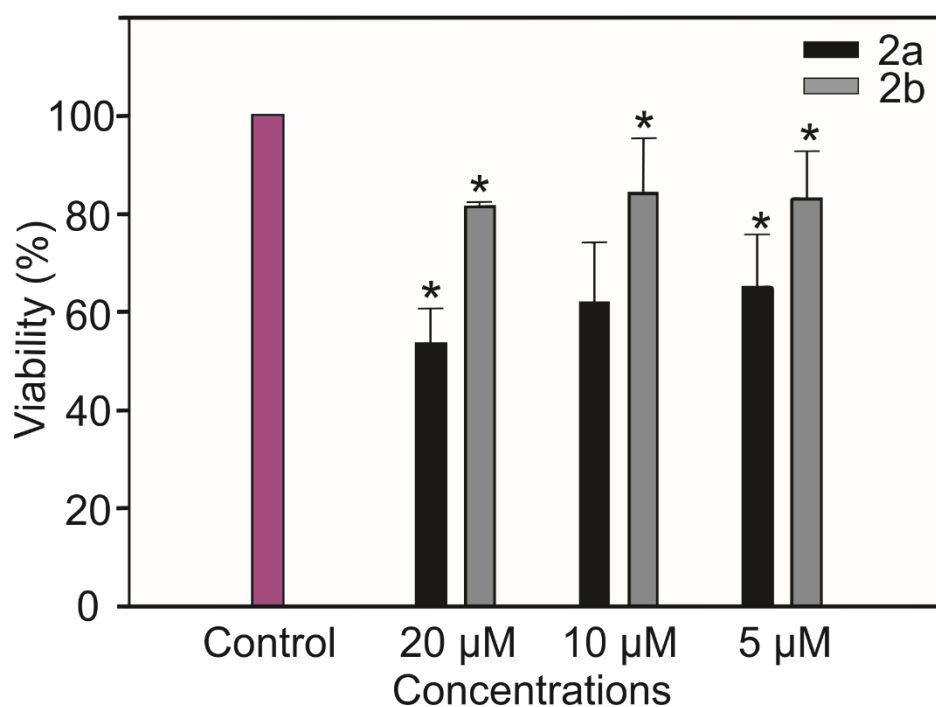


**Figure 4.** Most significant correlations provided by COSY and HMBC 2D spectra of compounds **2a** and **2b**

**2a** and **2b** were analyzed to determine their bioactive potential, as follows.

The viability assay was performed on HeLa cells after 48 of exposure time (Figure 5). The results clearly show differences between the effectiveness of two structural variants **2a** and **2b**. **2a** exhibited a slightly stronger cytostatic effect at concentrations 20, 10, and 5  $\mu$ M than **2b**. There was no significant difference observed between the effectivity of both the compounds at lower concentrations (2.5, 1.25, 0.625  $\mu$ M) with no decrease in viability (data not shown). These results suggest that even the small conformational change due to the occurrence of alternative starter modules in APTs BGC can influence the effectiveness of the compounds against cancer cells. Substitutions of the amino acid residues are thought to determine the degree of bioactivity [37,38]. Furthermore, more detailed studies on Trp containing APTs are needed to establish the structure activity relationship.





**Figure 5.** Cell viability was assessed by MTT assay at 48 h exposure time. Cell viability was calculated as the percentage of viable cells in **2a** and **2b** compounds-treated cells related to control at 20, 10, and 5  $\mu\text{M}$  concentration. Data are presented as mean  $\pm$  SD; n = 3. \*  $p \leq 0.05$ .

### 3. Materials and Methods

#### 3.1 Cyanobacterial strain and culturing conditions

The *Brasilonema* strain was collected in Costa Rica, as black mat on a bamboo trunk, on March 18, 2010 by Jan Mares and isolated into culture by Josef Juran. Strain was grown in Z' medium in glass columns (300 mL) bubbled with air enriched in 1.5%  $\text{CO}_2$  at constant temperature and illumination of  $28^\circ\text{C}$  and  $50 \mu\text{mol photon m}^{-2} \text{s}^{-1}$  [5]. Biomass was harvested by centrifugation ( $3125 \times g$ ) and stored at  $-80^\circ\text{C}$  and subsequently lyophilized.

#### 3.2 Genome Sequencing, assembling, annotation and mining for identification of Anabaenopeptin gene cluster

Single filaments of *Brasilonema* sp. CT11 were isolated by glass capillary technique described previously [39]. Briefly, filaments were serially washed in ten drops of TE buffer and amplified by multiple-displacement amplification (MDA) using the Repli-G Mini Kit (Qiagen). Ten filaments which passed quality check by 16S rRNA sequencing, were pooled together and sent for commercial *de novo* genome sequencing (EMBL Genomics Core Facility, Heidelberg, Germany) using an Illumina MiSeq Pair-End library with 250 bp reads, 350 bp average insert length, and 1.4 Gbp data yield. The data from Illumina were assembled using SPAdes 3.14 *de novo* assembler [40] with the single cell option enabled. Protein coding genes were predicted using Prodigal in the assembled scaffolds longer than 999 bp. All predicted proteins were compared to the NCBI-nr database using MMSeqs26 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 [41]. The anabaenopeptin gene cluster was identified based on BLASTp searches against the *Brasilonema* sp. CT11 genome assembly, using previously published *apt* genes as queries. Predicted open reading frames in the target genomic scaffold

were translated and subjected to BLASTp, NCBI conserved-domain search and antiSMASH 5.0 [42] analysis for functional annotation.

### 3.3 Crude extract preparation

Lyophilized biomass (~100 mg) was grinded (with the sea sand) and extracted with 10 mL of following solvents; hexane, chloroform, chloroform: methanol (1:1) and 70% methanol in water. All the extractions with different solvents were done in separate batches. Extracts were sonicated for 10 min (ultrasonication bath) and filtered through glass microfibre filter (1.2 $\mu$ m). All filtrates containing organic phase solvents were evaporated under vacuum using a rotary vacuum evaporator (Heidolph, Germany)

### 3.4 HPLC-HRMS/MS analysis

The bioactive extracts and fractions were analyzed on Thermo Scientific DionexUltiMate 3000 UHPLC (Thermo Scientific) equipped with a diode array detector (DAD) and high-resolution mass spectrometry with electrospray ionization source (ESI-HRMS; Impact HD Mass Spectrometer, Bruker). HPLC separation was performed on reversed phase Kinetex Phenomenex C<sub>18</sub> column (150  $\times$  4.6 mm, 2.6  $\mu$ m) with H<sub>2</sub>O/acetonitrile acidified with 0.1 % HCOOH as a mobile phase. Flow rate during analysis was 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 25min.) and 85/15 (in 30min.). The HPLC was connected to a high-resolution mass spectrometer (Bruker Impact HD) with following settings: 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 spectra rate 2 Hz. And the CID was set as a ramp from 20 to 60 eV on masses 200-1200, respectively. Calibration was performed using LockMass 622 as internal calibration solution and CH<sub>3</sub>COONa at the beginning of each analysis.

### 3.5 Molecular networking

The raw data files obtained from HPLC-HRMS/MS analysis were converted to mzXML format using MSConvert from the ProteoWizard suite (<http://proteowizard.sourceforge.net/tools.shtml>) [43]. The molecular network was created using the Global Natural Products Social Molecular Networking (GNPS) online workflow[44]. The 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 6 peaks in the +/- 50 Da window throughout the spectrum. The 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. Further, consensus spectra that contained less than 2 spectra were discarded. A network was then created where edges were filtered to have a cosine score above 0.65 and more than 4 matched peaks. Further edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. The spectra in the network were then searched against GNPS' spectral libraries. The library spectra were filtered in the same manner as the input data. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 4 matched peaks. Analogue search was enabled against the library with a maximum mass shift of 200 Da.

### 3.6 Isolation of compound 2a and 2b from *Brasilonema* CT11

In order to obtain larger biomass volume, strain was cultivated in 100 L cultivation unit twice to obtain 30 g of freeze-dried biomass. Crude extract was prepared by as described previously

in section 3.2 to obtain 5 g of crude extract. The obtained crude extracts were fractionated using reversed-phase flash column chromatography, eluting with a mixture of H<sub>2</sub>O/CH<sub>3</sub>CN (from 0 to 100%) and then with 100% of MeOH, to afford twenty fractions (Fr1–Fr20) [45]. Fraction 15 eluting with 80% of CH<sub>3</sub>CN was found to be enriched with APT compounds and hence, was further fractionated using Sephadex LH20 to obtain ten fractions using CH<sub>2</sub>Cl<sub>2</sub>/MeOH (1:1). Fraction 3 was further purified using semi-preparative reversed-phase column chromatography, eluting with 60% of CH<sub>3</sub>CN, to obtain compounds **2a** and **2b**.

Compound (**2a**) HRESIMS: tR = 20 min; [M + H]<sup>+</sup> m/z 803.4520, calcd. 803.4411 (Figure S2); <sup>1</sup>H NMR (700 MHz, CD<sub>3</sub>OD): δ 9.21 (1H, d, J=8.8), δ 8.97 (1H, d, J=3.2), δ 8.40 (2H, br s), δ 7.90 (1H, s), δ 7.82 (1H, d, J=8.0), δ 7.78 (1H, br s), δ 7.57 (1H, d, J=7.9), δ 7.31 (1H, d, J=8.1), δ 7.21 (2H, t, J=7.6), δ 7.14 (1H, t, 7.6), δ 7.11 (2H, d, J=7.6), δ 7.09 (1H, br d J=7.6), δ 7.05 (1H, s), δ 7.02 (1H, t, J=7.4), δ 5.05 (1H, dt, J=11.2, 5.0, 3.7), δ 4.60 (1H, q, J=6.7), δ 4.50 (1H, ddd, J=12.5, 8.7, 2.9), δ 4.18 (1H, br s), δ 4.15 (1H, t, J=4.6), δ 4.04 (1H, dd, J=9.6, 4.9), δ 3.75 (1H, m), δ 3.38 (1H, dd, J=13.7, 3.1), δ 3.27 (2H, m), δ 2.92 (1H, br d, J=12.6), δ 2.86 (1H, t, J=13.2), δ 2.25 (1H, m), δ 1.96 (1H, m), δ 1.93 (1H, m), δ 1.75 (1H, m), δ 1.57 (3H, s), δ 1.51 (2H, m), δ 1.45 (1H, m), δ 1.38 (2H, m), δ 1.26 (1H, m), δ 1.17 (3H, d, J=6.8), δ 0.98 (3H, d, J=6.9), δ 0.93 (3H, t, J=7.4), δ -0.26 (3H, d, J=6.7).

Compound (**2b**): HRESIMS: tR = 20 min; [M + H]<sup>+</sup> m/z 803.4510, calcd. 803.4411 (Figure S2); <sup>1</sup>H NMR (700 MHz, CD<sub>3</sub>OD): δ 10.4 (1H, s, NH-Trp), δ 9.2 (1H, d, J=8.7, Phe-NH), δ 9.0 (1H, d, J=3.4, NH-Trp), δ 8.2 (1H, br s, NH-Val, NH-CO-2H), δ 7.9 (1H, d, J=8.6, NH-Lys), δ 7.7 (1H, br s, NH-Leu), δ 7.6 (1H, d, J=8.1), δ 7.3 (1H, d, J=8.1), δ 7.2 (2H, t, J=7.4), δ 7.14 (1H, t, J=7.4), δ 7.09 (t, 3H, overlapped), δ 7.04 (1H, s), δ 7.0 (1H, t, J=7.4), δ 5.1 (1H, m, CH-α Trp), δ 4.6 (1H, q, J=6.7, CH-α NMeAla), δ 4.47 (1H, ddd, J=3.1, 8.5, 12.1, CH-α Phe), δ 4.3 (1H, m, CH-α Leu), δ 4.2 (1H, d, J=3.8), δ 4.16 (1H, t, J=4.9, CH-α Lys), δ 3.8 (1H, m), δ 3.4 (1H, dd, J=3.1, 13.5), δ 3.3 (1H, d, J=5.1), δ 2.9 (1H, br d), δ 2.85 (1H, t, J=13.3), δ 2.25 (1H, m), δ 1.95 (1H, tt, J=3.4, 13.8), δ 1.87 (1H, m), δ 1.76 (2H, overlapped), δ 1.73 (1H, overlapped), δ 1.62 (1H, m), δ 1.58 (3H, s, N-CH<sub>3</sub>), δ 1.55 (1H, m), δ 1.47 (1H, m), δ 1.27 (1H, m), δ 1.05 (3H, d, J=6.4), δ 1.0 (6H, dd, J=1.9, 6.6), δ 0.92 (3H, d, J=6.8), δ -0.25 (3H, d, J=6.6).

### 3.8 Antiproliferative activity

The human cervical cancer line HeLa was obtained from Mgr. David Sedlák, Ph.D. (Institute of Molecular Genetics of ASCR, v.v.i. Czech Republic) and were maintained at 37°C in a humidified incubator with 5% CO<sub>2</sub>. Cells were cultured in DMEM cultivation medium (Gibco Life Technologies) supplemented with 10% FBS (Gibco Life Technologies), 1% antibiotics (Gibco Life Technologies) and L-glutamine 2 mM (Gibco Life Technologies). The HeLa cells were plated (transparent 96-well cell culture plate, flat bottom) at a concentration of 1 × 10<sup>4</sup> cells per well one day prior to cytotoxicity experiment. The exposure solutions were prepared to obtain final concentration of the compounds 20, 10, 5, 2.5, 1.25 and 0.625 μM per well. Two technical replicates in one experiment were performed and final concentration of DMSO in the tested wells did not exceed 0.5%. After 48 h of incubation MTT assay [46] was performed as the endpoint measurement. Three independent biological experiments were performed for each compound and finally the viability index was expressed as a ratio between the absorbance values of the treated and control cells in percent.

### 3.9 Data deposition

The strain *Brasilonema* sp. CT11 has been deposited to culture collection of autotrophic organism (CCALA) under the strain number CCALA 1130. The genomic assembly is available under NCBI accession number JABXYX010000000, and the anabaenopeptin biosynthetic gene cluster is available under accession number MT670293. The mass spectrometry data was deposited on [MassIVE](https://massive.ucsd.edu/) public repository (MSV000085797). The molecular networking job can be publicly accessed at <https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=537f43651a084553ad560eac52895539>.

**Supplementary Materials:** The following are available online at [www.mdpi.com/xxx/s1](http://www.mdpi.com/xxx/s1),

**Author Contributions:** Conceptualization, S.S. and K.S.; methodology, S.S., K.S., G.E., A.C., J.M., and P.H.; validation, X.X., Y.Y. and Z.Z.; formal analysis, J.J., K.S. and J.M.; investigation, S.S., K.D. and K.S.; data curation, J.M., J.J., V.C., and K.S.; writing—original draft preparation, S.S., G.E., A.C., J.M., V.C., and K.S.; writing—review and editing, K.S., P.H., and V.C.; supervision, K.S.; project administration, K.S.; funding acquisition, K.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Ministry of Education, Youth and Sports of the Czech Republic MSCA IF II project (CZ.02.2.69/0.0/0.0/18\_070/0010493), Czech Science foundation (GAČR)- project no. 19-17868Y) and OncoLab, Regione Campania, PO FESR 2014-2020, O.S. 1.2, Project “Campania Oncoterapie” No. B61G18000470007.

**Conflicts of Interest:** “The authors declare no conflict of interest.”

## References

1. Mazard, S.; Penesyan, A.; Ostrowski, M.; Paulsen, I.T.; Egan, S. Tiny Microbes with a Big Impact: The Role of Cyanobacteria and Their Metabolites in Shaping Our Future. *Mar Drugs* **2016**, *14*.
2. Schirromeister, B.E.; Antonelli, A.; Bagheri, H.C. The origin of multicellularity in cyanobacteria. *BMC Evolutionary Biology* **2011**, *11*, 45.
3. Demay, J.; Bernard, C.; Reinhardt, A.; Marie, B. Natural Products from Cyanobacteria: Focus on Beneficial Activities. *Mar Drugs* **2019**, *17*.
4. Dittmann, E.; Gugger, M.; Sivonen, K.; Fewer, D.P. Natural Product Biosynthetic Diversity and Comparative Genomics of the Cyanobacteria. *Trends Microbiol* **2015**, *23*, 642-652.
5. Saurav, K.; Macho, M.; Kust, A.; Delawska, K.; Hajek, J.; Hrouzek, P. Antimicrobial activity and bioactive profiling of heterocytous cyanobacterial strains using MS/MS-based molecular networking. *Folia Microbiol (Praha)* **2019**, *64*, 645-654.
6. Mares, J.; Hajek, J.; Urajova, P.; Kust, A.; Jokela, J.; Saurav, K.; Galica, T.; Capkova, K.; Mattila, A.; Haapaniemi, E., *et al.* Alternative Biosynthetic Starter Units Enhance the Structural Diversity of Cyanobacterial Lipopeptides. *Appl Environ Microbiol* **2019**, *85*.
7. Kust, A.; Mares, J.; Jokela, J.; Urajova, P.; Hajek, J.; Saurav, K.; Voracova, K.; Fewer, D.P.; Haapaniemi, E.; Permi, P., *et al.* Discovery of a Pederin Family Compound in a Nonsymbiotic Bloom-Forming Cyanobacterium. *ACS Chem Biol* **2018**, *13*, 1123-1129.
8. Teta, R.; Romano, V.; Sala, G.D.; Picchio, S.; Sterlich, C.D.; Mangoni, A.; Tullio, G.D.; Costantino, V.; Lega, M. Cyanobacteria as indicators of water quality in Campania coasts, Italy: a monitoring strategy combining remote/proximal sensing and in situ data. *Environmental Research Letters* **2017**, *12*, 024001.
9. Esposito, G.; Teta, R.; Marrone, R.; De Sterlich, C.; Casazza, M.; Anastasio, A.; Lega, M.; Costantino, V. A Fast Detection Strategy for Cyanobacterial blooms and associated cyanotoxins (FDSCC) reveals the occurrence of lyngbyatoxin A in campania (South Italy). *Chemosphere* **2019**, *225*, 342-351.

10. Rezanka, T.; Dembitsky, V.M. Metabolites produced by cyanobacteria belonging to several species of the family Nostocaceae. *Folia Microbiol (Praha)* **2006**, *51*, 159-182.
11. Tan, L.T. Filamentous tropical marine cyanobacteria: a rich source of natural products for anticancer drug discovery. *Journal of Applied Phycology* **2010**, *22*, 659-676.
12. Jones, A.C.; Monroe, E.A.; Podell, S.; Hess, W.R.; Klages, S.; Esquenazi, E.; Niessen, S.; Hoover, H.; Rothmann, M.; Lasken, R.S., *et al.* Genomic insights into the physiology and ecology of the marine filamentous cyanobacterium *Lyngbya majuscula*. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 8815-8820.
13. Nunnery, J.K.; Mevers, E.; Gerwick, W.H. Biologically active secondary metabolites from marine cyanobacteria. *Current Opinion in Biotechnology* **2010**, *21*, 787-793.
14. Zanchett, G.; Oliveira-Filho, E.C. Cyanobacteria and cyanotoxins: from impacts on aquatic ecosystems and human health to anticarcinogenic effects. *Toxins (Basel)* **2013**, *5*, 1896-1917.
15. Caso, A.; Esposito, G.; Della Sala, G.; Pawlik, J.R.; Teta, R.; Mangoni, A.; Costantino, V. Fast Detection of Two Smenamide Family Members Using Molecular Networking. *Marine Drugs* **2019**, *17*.
16. Mohimani, H.; Gurevich, A.; Mikheenko, A.; Garg, N.; Nothias, L.F.; Ninomiya, A.; Takada, K.; Dorrestein, P.C.; Pevzner, P.A. Dereplication of peptidic natural products through database search of mass spectra. *Nat Chem Biol* **2017**, *13*, 30-37.
17. Mohimani, H.; Gurevich, A.; Shlemov, A.; Mikheenko, A.; Korobeynikov, A.; Cao, L.; Shcherbin, E.; Nothias, L.F.; Dorrestein, P.C.; Pevzner, P.A. Dereplication of microbial metabolites through database search of mass spectra. *Nat Commun* **2018**, *9*, 4035.
18. Quinn, R.A.; Nothias, L.F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P.C. Molecular Networking As a Drug Discovery, Drug Metabolism, and Precision Medicine Strategy. *Trends Pharmacol Sci* **2017**, *38*, 143-154.
19. Harada, K.-i.; Fujii, K.; Shimada, T.; Suzuki, M.; Sano, H.; Adachi, K.; Carmichael, W.W. Two cyclic peptides, anabaenopeptins, a third group of bioactive compounds from the cyanobacterium *Anabaena flos-aquae* NRC 525-17. *Tetrahedron Letters* **1995**, *36*, 1511-1514.
20. Welker, M.; von Dohren, H. Cyanobacterial peptides - nature's own combinatorial biosynthesis. *FEMS Microbiol Rev* **2006**, *30*, 530-563.
21. Rounge, T.B.; Rohrlack, T.; Nederbragt, A.J.; Kristensen, T.; Jakobsen, K.S. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. *BMC Genomics* **2009**, *10*, 396.
22. Christiansen, G.; Philmus, B.; Hemscheidt, T.; Kurmayer, R. Genetic Variation of Adenylation Domains of the Anabaenopeptin Synthesis Operon and Evolution of Substrate Promiscuity. *Journal of Bacteriology* **2011**, *193*, 3822.
23. Rouhiainen, L.; Jokela, J.; Fewer, D.P.; Urmann, M.; Sivonen, K. Two alternative starter modules for the non-ribosomal biosynthesis of specific anabaenopeptin variants in *Anabaena* (Cyanobacteria). *Chem Biol* **2010**, *17*, 265-273.
24. Koketsu, K.; Mitsuhashi, S.; Tabata, K. Identification of homophenylalanine biosynthetic genes from the cyanobacterium *Nostoc punctiforme* PCC73102 and application to its microbial production by *Escherichia coli*. *Appl Environ Microbiol* **2013**, *79*, 2201-2208.
25. Lima, S.; Alvarenga, D.; Etchegaray, A.; Fewer, D.; Jokela, J.; Varani, A.; Sanz, M.; Dörr, F.; Pinto, E.; Sivonen, K., *et al.* Genetic Organization of Anabaenopeptin and Spumigin Biosynthetic Gene Clusters in the Cyanobacterium *Sphaerospermopsis torques-reginae* ITEP-024. *ACS chemical biology* **2017**, *12*.

26. Alvarenga, D.O.; Franco, M.W.; Sivonen, K.; Fiore, M.F.; Varani, A.M. Evaluating Eucalyptus leaf colonization by *Brasilonema octagenarum* (Cyanobacteria, Scytonemataceae) using in planta experiments and genomics. *PeerJ* **2020**, *8*, e9158.
27. Fiore, M.F.; Sant'Anna, C.L.; Azevedo, M.T.d.P.; Komárek, J.; Kaštovský, J.; Sulek, J.; Lorenzi, Adriana, S. The cyanobacterial genus *Brasilonema*, gen. Nov., a molecular and phenotypic evaluation. *Journal of Phycology* **2007**, *43*, 789-798.
28. Vaccarino, M.A.; Johansen, J.R. *Brasilonema angustatum* sp. Nov. (nostocales), a new filamentous cyanobacterial species from the hawaiian islands. *Journal of Phycology* **2012**, *48*, 1178-1186.
29. Villanueva, C.D.; Hašler, P.; Dvořák, P.; Pouličková, A.; Casamatta, D.A. *Brasilonema lichenoides* sp. nov. and *Chroococidiopsis lichenoides* sp. nov. (Cyanobacteria): two novel cyanobacterial constituents isolated from a tripartite lichen of headstones. *Journal of Phycology* **2018**, *54*, 224-233.
30. Sanz, M.; Andreote, A.P.; Fiore, M.F.; Dorr, F.A.; Pinto, E. Structural Characterization of New Peptide Variants Produced by Cyanobacteria from the Brazilian Atlantic Coastal Forest Using Liquid Chromatography Coupled to Quadrupole Time-of-Flight Tandem Mass Spectrometry. *Mar Drugs* **2015**, *13*, 3892-3919.
31. Schmidt, E.W.; Harper, M.K.; Faulkner, D.J. Mozamides A and B, Cyclic Peptides from a Theonellid Sponge from Mozambique. *Journal of Natural Products* **1997**, *60*, 779-782.
32. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **2019**, *47*, W81-W87.
33. Luesch, H.; Hoffmann, D.; Hevel, J.M.; Becker, J.E.; Golakoti, T.; Moore, R.E. Biosynthesis of 4-methylproline in cyanobacteria: cloning of nosE and nosF genes and biochemical characterization of the encoded dehydrogenase and reductase activities. *J Org Chem* **2003**, *68*, 83-91.
34. Hoffmann, D.; Hevel, J.M.; Moore, R.E.; Moore, B.S. Sequence analysis and biochemical characterization of the nostopeptolide A biosynthetic gene cluster from *Nostoc* sp. GSV224. *Gene* **2003**, *311*, 171-180.
35. D'Agostino, P.M.; Gulder, T.A.M. Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synth Biol* **2018**, *7*, 1702-1708.
36. May, D.S.; Chen, W.L.; Lantvit, D.D.; Zhang, X.; Kronic, A.; Burdette, J.E.; Eustaquio, A.; Orjala, J. Merocyclophanes C and D from the Cultured Freshwater Cyanobacterium *Nostoc* sp. (UIC 10110). *J Nat Prod* **2017**, *80*, 1073-1080.
37. Gesner-Apter, S.; Carmeli, S. Protease inhibitors from a water bloom of the cyanobacterium *Microcystis aeruginosa*. *J Nat Prod* **2009**, *72*, 1429-1436.
38. Spooft, L.; Blaszczyk, A.; Meriluoto, J.; Ceglowska, M.; Mazur-Marzec, H. Structures and Activity of New Anabaenopeptins Produced by Baltic Sea Cyanobacteria. *Mar Drugs* **2015**, *14*, 8.
39. Mares, J.; Hajek, J.; Urajova, P.; Kopecky, J.; Hrouzek, P. 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* **2014**, *9*, e111904.
40. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D., et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **2012**, *19*, 455-477.
41. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **2015**, *25*, 1043-1055.

42. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research* **2019**, *47*, W81-W87.
43. Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. ProteoWizard: open source software for rapid proteomics tools development. *Bioinformatics* **2008**, *24*, 2534-2536.
44. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapon, C.A.; Luzzatto-Knaan, T., *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34*, 828-837.
45. Saurav, K.; Borbone, N.; Burgsdorf, I.; Teta, R.; Caso, A.; Bar-Shalom, R.; Esposito, G.; Britstein, M.; Steindler, L.; Costantino, V. Identification of Quorum Sensing Activators and Inhibitors in The Marine Sponge *Sarcotragus spinosulus*. *Marine Drugs* **2020**, *18*.
46. Mosmann, T. Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. *Journal of Immunological Methods* **1983**, *65*, 55-63.