Biosynthesis of microcystin hepatotoxins in the cyanobacterial genus *Fischerella*

Karina Heck[a], Danillo O. Alvarenga[a,c], Tania K. Shishido[b], Alessandro M. Varani[c], Felipe A. Dörr[d], Ernani Pinto[d], Leo Rouhiainen[b], Jouni Jokela[b], Kaarina Sivonen[b], Marli F. Fiore[a,*]

[a]University of São Paulo (USP), Center for Nuclear Energy in Agriculture, Piracicaba-SP, Brazil

[b]University of Helsinki, Department of Food and Environmental Sciences, Division of Microbiology and Biotechnology, Helsinki, Finland

[c]São Paulo State University (UNESP), College of Agricultural and Veterinary Science, Department of Technology, Jaboticabal-SP, Brazil

[d]University of São Paulo (USP), School of Pharmaceutical Sciences, São Paulo-SP, Brazil

*Corresponding Author

E-mail: fiore@cena.usp.br

Address: University of São Paulo, Center for Nuclear Energy in Agriculture, Av. Centenário 303, 13400-970, Piracicaba, SP, Brazil.

ABSTRACT

Microcystins (MCs) are serine/threonine phosphatase inhibitors synthesized by several members of the phylum Cyanobacteria. Mining the draft genome sequence of the nostocalean MC-producing *Fischerella* sp. strain CENA161 led to the identification of three contigs containing *mcy* genes. Subsequent PCR and Sanger sequencing allowed the assembling of its complete biosynthetic *mcy* gene cluster with 55,016 bases in length. The cluster encoding ten genes (*mcyA-J*) with a central bidirectional promoter was organized in a similar manner as found in other genera of nostocalean cyanobacteria. However, the nucleotide sequence of the *mcy* gene cluster of *Fischerella* sp. CENA161 showed significant differences from all the other MC-producing cyanobacterial genera, sharing only 85.2 to 78.2% identities. Potential MC variants produced by *Fischerella* sp. CENA161 were predicted by the analysis of the adenylation domain binding pockets and further investigated by LC-MS/MS analysis. To our knowledge, this study presents the first complete *mcy* cluster characterization from a strain of the genus *Fischerella*, providing new insight into the distribution and evolution of MCs in the phylum Cyanobacteria.

Keywords: cyanotoxins, phosphatase inhibitors, genome mining, Nostocales

1. Introduction

Microcystins (MCs) are small cyclic heptapeptides synthesized by several members of the phylum Cyanobacteria with global significance due to their toxicity to humans and other animals (Jochimsen et al., 1998; Sivonen and Jones, 1999). Their toxicity is exerted through inhibition of members of the protein phosphatase families PP1 and PP2A (MacKintosh et al. 1990; Gulledge et al., 2002). Despite best known for their acute hepatotoxicity, MCs are of interest as possible anti-cancer drug development targets (Niedermeyer et al., 2014; Kounnis et al., 2015). The general structure of MCs can be summarized as cyclo-D-Ala^1-X^2-D-$MeAsp^3$-Z^4-$Adda^5$-D-Glu^6-$Mdha^7$ (see Figure 1) (Botes et al., 1985), where X and Z are variable L-amino acids, while D-MeAsp corresponds to D-erythro-β-methyl-aspartic acid, Mdha to N-methyl-α-β-dehydroalanine and Adda to (2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-(4*E*,6*E*)-dienoic acid. The latter is exclusive to these toxins and nodularins and contributes to the molecule toxicity (Gulledge et al., 2002; Kounnis et al., 2015).

MCs are synthesized through enzymatic modification of short precursor peptides in the nonribosomal pathway. This process is driven by a multifunctional modular enzyme complex consisted of a combination of nonribosomal peptide synthetases (NRPS), type I polyketide synthases (PKS-I), hybrid NRPS/PKS-I and tailoring enzymes (Nishizawa et al., 2000; Tillett et al., 2000; Christiansen et al., 2003; Rouhiainen et al., 2004; Fewer et al., 2013). The microcystin gene cluster (*mcy*) is composed of nine to ten genes depending on taxa and the involvement of several *mcy* genes in MC biosynthesis was established by gene inactivation studies (Dittmann et al., 1997; Pearson et al., 2004; Christiansen et al., 2008; Fewer et al., 2008). The closely related nodularin (*nda*) synthetase gene cluster from *Nodularia* was also elucidated, indicating that it derived from MC synthetase genes through a deletion event and a change in substrate specificity (Moffitt and Neilan, 2004; Rantala et al., 2004). The biological role of cyanobacterial MC is not currently understood, but several hypotheses have been suggested such as contributing in photosynthesis, environmental adaptation, protection against oxidative stress, nutrient metabolism and storage, quorum sensing, colony formation, defense against zooplanktonic grazers, iron uptake or transfer and allelopathy (Omidi et al., 2017). These authors stated that conflicting results, unstandardized experimental design, strain-specific behavior and differences between conditions in laboratory and nature hinder generalizations on microcystin functions.

Despite several MC-producing strains have been found in the genera *Microcystis*, *Anabaena*, *Nostoc*, *Fischerella*, *Hapalosiphon*, *Oscillatoria*/*Planktothrix*, and *Phormidium* (Bishop et al., 1959; Botes et al., 1984; Krishnamurthy et al., 1986; Eriksson et al., 1988; Meriluoto et al., 1989; Sivonen et al., 1990; Harada et al., 1991; Prinsep et al., 1992; Izaguirre et al., 2007; Fiore et al., 2009), the MC biosynthetic pathway was only characterized in few strains of the genera *Microcystis*, *Anabaena*, *Planktothrix* and *Nostoc* (Tillett et al., 2000; Christiansen et al., 2003; Rouhiainen et al., 2004; Rounge et al., 2009; Fewer et al., 2013). The *mcy* gene clusters of these distantly related cyanobacterial genera have revealed a highly conserved set of multidomain proteins depicting the same basic reaction steps. Differences among these clusters have been observed in gene arrangements, localization and orientation of promoter regions, and in genes coding for tailoring enzymes. Interestingly, a cyanobacterium containing one *mcy* gene cluster can produce more than one MC variant mainly due to the relaxed specificity of adenylation (A) domains of McyB-A_1 and McyC-A (amino acid positions **2** and **4**, Figure 1). Therefore, the description of novel MC gene clusters from different cyanobacterial taxa offers high potential for isolating variants with unique properties.

Although MC production and fragments of biosynthetic genes have already been identified in strains of the genus *Fischerella* (Fiore et al., 2009; Cirés et al., 2014) and even prediction of an incomplete gene cluster has been reported (Shih et al., 2013), the entire gene cluster remains unsolved. Here we used a genomics-based approach to characterize the complete biosynthetic gene cluster in the MC-producing strain *Fischerella* sp. CENA161. Prediction analysis based on the amino acid residues lining the substrate-binding pockets in NRPS A domains were performed and potential structural variants investigated by high performance liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS).

2. Method

2.1. Cyanobacterial strain

The cyanobacterium *Fischerella* sp. CENA161 was isolated from a water sample collected from a small concrete dam of spring water in the municipality of Piracicaba, São Paulo state, Brazil, as previously described (Fiore et al., 2009). This strain is maintained under culture in CENA/USP, located in Piracicaba, SP, Brazil, in BG−11 (Allen, 1968) liquid medium without inorganic nitrogen ($BG{-}11_0$), at 25±1 ºC, with a 14:10 h light/dark photoperiod, and photon flux density of 40 µmol photons/m^2/s.

2.2. DNA extraction, PCR amplification and Sanger sequencing

Cells from the cyanobacterial culture were collected and processed as previously described (Heck et al., 2016). Total genomic DNA was extracted using the AxyPrep™ Bacterial Genomic DNA Miniprep Kit (Axygen Biosciences) according to manufacturer instructions. The integrity of the total genomic DNA extracted was verified using 1% agarose gel electrophoresis. The extracted DNA was purified with Axyprep™ PCR Clean-up Kit (Axygen) according to manufacturer instructions. Microcystin codifying genes (*mcy*) were amplified by polymerase chain reaction using a combination of primer sets previously described in literature and designed for this work (Table 1). PCR products were ligated to pGEM®-T Easy Vector Systems (Promega) and inserted into chemically competent *Escherichia coli* DH5α cells. Plasmids that received the PCR products were extracted from cells by alkaline hydrolysis (Birnboim and Doly, 1979). Sequencing reactions were performed using the BigDye Terminator Cycle Sequencing Kit (GE Healthcare), with vector primers M13F/M13R in a Techne TC-412 thermocycler (Bibby Scientific Limited) for 25 cycles at 95 ºC for 20 s, 52 ºC for 15 s, and 60 ºC for 1 min. Purified reactions were analyzed in an ABI PRISM 3500 genetic analyzer (Life Technologies). The sequenced reads had their base quality analyzed and consensus sequences were generated with the Phred/Phrap/Consed software package (Ewing and Green, 1998; Ewing et al., 1998; Gordon et al., 1998). Sequences were aligned and compared to other sequences available in NCBI GenBank (http://www.ncbi.nlm.nih.gov/) using BLASTn (Altschul et al., 1997).

2.3. Whole genome sequencing and assembly

Genomic DNA extracted from the cells was quantified using Qubit dsDNA Broad BR Assay kit and Qubit®2.0 Fluorometer (Thermo Fisher Scientific). Paired-ends libraries were prepared with the Nextera XT Sample Prep Kit (Illumina), which were sequenced in the MiSeq (Illumina) platform using the MiSeq 600 cycle Reagent Kit v3 (Illumina) according to manufacturer instructions. The quality of the raw Illumina sequence reads were initially assessed using FastQC v0.10.1 (Andrews, 2010). Bases with quality indices lower than Phred 20 and sequences shorter than 50 bp were removed using the program SeqyClean 1.8.10 (Zhbannikov et al., 2015). Overlapping read pairs were merged with PEAR 0.9.6 (Zhang et al., 2014) and genome assembly was performed using SPAdes 3.1.1 (Bankevich et al., 2012).

The complete nucleotide sequence of the *mcy* gene cluster of *Fischerella* sp. CENA161 was deposited in GenBank under accession number KX891213.

2.4. Microcystin gene cluster annotation and phylogenetic analysis

The MC synthetase gene cluster was identified by using BLASTn alignments between the 10 *mcy* gene sequences obtained in Sanger sequencing and the assembled genome file. Manual annotation was performed using Artemis 15.1.10 (Rutherford et al., 2000). The identification of motifs and adenylation domains was performed using NRPS Predictor2 (Rausch et al., 2005; Röttig et al., 2011). Amino acid sequences from the $McyB_2$ and McyC adenylation domains and for the 10 genes found in complete microcystin gene clusters available in the NCBI GenBank database were independently aligned with MUSCLE 3.8.31 (Edgar, 2004) and evolutionary models were estimated with ProtTest 3.2 (Darriba et al., 2011). Phylogenetic trees were reconstructed from alignments by Bayesian inference with MrBayes 3.2.5 (Ronquist and Huelsenbeck, 2003) using 5,000,000 generations, four chains and two independent runs.

2.5. Liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS)

The intracellular content of 60 mg of freeze-dried cells from the culture sample was extracted with MeOH:H_2O 70/30 (v/v) at ultrasound probe (Sonic Ruptor 400, Omni) during 1 min on ice. The supernatant was collected after centrifugation (10.000 × *g* for 10 min) and the extract diluted to 10% MeOH with ultrapure water. The sample was applied to a solid phase extraction cartridge (Sep-Pak 500 mg, Waters Corp.), previously conditioned by the sequential passage of 5 ml of MeOH and 5 ml of MeOH:H_2O 10/90 (v/v). After a washing step with 5 ml of MeOH:H_2O 10/90 (v/v), elution proceeded with 5 ml of MeOH:H_2O 90/10 (v/v). The solvent was evaporated under a stream of nitrogen and the concentrate reconstituted in 500 µl of MeOH:H_2O 50/50 (v/v), filtered (0.45 µm, PVDF, Millipore) and transferred to appropriate vials. Chromatography was performed in a Prominence HPLC (Shimadzu) employing a Fusion-RP column (150 x 2 mm, 4 µm; Phenomenex) with a gradient of (A) 2 mM ammonium formate containing 0.1% formic acid and (B) acetonitrile:water 90/10 (v/v) with the same additives, at a flow rate of 0.2 ml/min. Gradient elution proceeded as follows: 35 to 60% B in 10 min; 60 to 100% B in 6 min; 100% B in 2 min; 100 to 35% B in 0.5 min and finally kept in 25% B for 6.5 min. Collision-induced dissociation experiments for MC detection and characterization were performed in an Esquire HCT ion trap mass spectrometer (Bruker Daltonics) equipped with an electrospray ion source.

UPLC-QTOF analyses were performed with Acquity I-Class UPLC - Synapt G2-Si HDMS (Waters Corp.) system. Two µl filtered cyanobacterial methanol extract were injected into a Kinetex C8 column (50 x 2.1 mm, 1.7 µm, Phenomenex) which was eluted at 40 °C with a flow rate of 0.3 ml/min using (A) 0.1% formic acid and (B) acetonitrile/isopropanol 50/50 (v/v) containing 0.1% formic acid. Gradient elution proceeded as follows: 25 to 65% B in 5 min; 65 to 100% B in 0.01 min; 100% B in 1.99 min; 100 to 20% B in 0.5 min and finally kept in 25% B for 2.5 min. The mass spectrometer was calibrated with sodium formate giving a calibrated mass range from *m/z* 91.055 to 1921.759. Leucine enkephalin was used at 10 s interval as a lock mass reference compound. Mass spectral data was accumulated in positive electrospray ionization at a scan range from *m/z* 50 to 2000.

3. Results and Discussion

The search for conserved regions from *mcy* genes in the *Fischerella* sp. CENA161 genomic DNA by PCR amplification and Sanger sequencing returned positive for all 10 genes. High-throughput sequencing with Illumina MiSeq resulted in approximately 24 million raw reads. The raw data was assembled into 443 contigs, which constituted a draft genome size of 7,210,502 bp with ca. 220× coverage and GC content at 40.18%. Both genome size and GC content were in agreement with other recently sequenced cyanobacterial members of the genus *Fischerella* (Dagan et al., 2012; Shih et al., 2013; Hirose et al., 2016).

With the draft genome data in hand, we searched for contig(s) that coharbor *mcy* genes using the *mcy* PCR fragments as *in silico* probes. This effort led to the identification of three contigs (45,504 bp, 5,557 bp and 2,626 bp). Bioinformatics analyses identified that the two gaps separating the three contigs were located within the *mcyB* and *mcyC* genes, which were closed using PCR amplification and Sanger sequencing. The three contigs were assembled into a 55,137 kb contiguous region that contained the 55,016 bp *mcy* gene cluster.

The CENA161 *mcy* gene cluster encompasses ten genes (*mcyA*-*J*) with a central bidirectional promoter (Figure 2, Table 2). The genes encoding the McyG, D, E, A, B and C enzymes are responsible for the stepwise assembly and cyclization of peptide intermediates to form MC, while McyF, I and J are tailoring enzymes and McyH is an ABC transporter hypothetically involved in the efflux of the toxin (Tillett et al., 2000; Christiansen et al., 2003; Rouhiainen et al., 2004; Fewer et al., 2013). MC biosynthesis by *Fischerella* sp. CENA161 follows the collinearity rule, i.e., the order of genes is the same as the order of the single enzymatic steps (Marahiel et al., 1997; von Döhren et al., 1997) as occurs in other nostocalean cyanobacteria, such as *Anabaena* sp. 90 and *Nostoc* sp. 152, and NOD in *Nodularia spumigena* NSOR10. In these cyanobacteria, MC assembly is believed to initiate with the A-PCP domains of the hybrid NRPS/PKS-I enzyme McyG loading phenyllactate (Hicks et al., 2006). Then, the four PKS-I modules of McyG, D and E complete the formation of the Adda skeleton, and the *O*-methyltransferase McyJ (Christiansen et al., 2003) and the aminotransferase domain of McyE (Tillett et al., 2000) incorporate Adda side chain modifications. The NRPS modules of McyG, A, B, and C incorporate the six remaining amino acids (Tillett et al., 2000). The enzymes 2-hydroxy acid dehydrogenase McyI (Pearson et al., 2007) and the aspartate racemase McyF (Sielaff et al., 2003) are involved in the biosynthesis of D-erythro-β-methyl-aspartate. Finally, the elongated peptide is released from the enzyme complex by the thioesterase domain of McyC.

The nucleotide sequence of the *Fischerella* sp. CENA161 *mcy* gene cluster showed significant differences from clusters thus far characterized. Hits from the BLAST analyses in descending order of identity were *Anabaena* sp. 90 (coverage 95%, identity 85.2%), *Nostoc* sp. 152 (coverage 89%, identity 83.0%), *Nodularia spumigena* NSOR10 (coverage 78%, identity 80.8%), *Planktothrix rubescens* NIVA-CYA 98 (coverage 71%, identity 78.2%) and *Microcystis aeruginosa* NIES-843 (coverage 71%, identity 74.1%). Similarly, for amino acid sequences, top BLAST hits observed were with other nostocalean strains (Table 2). Comparisons of the *mcy* gene cluster from

Fischerella sp. CENA161 with other nostocalean genera (*Anabaena*, *Nostoc* and *Nodularia*) revealed, in general, the same structural organization with differences in the arrangement of certain genes coding for tailoring enzymes (Figure 2). On the other hand, major differences were observed in comparisons with distinct cyanobacterial orders (*Planktothrix* and *Microcystis*). These results reinforce previous observations that *mcy* gene arrangements are almost identical among cyanobacteria according to their taxonomic position (Rantala et al., 2004; Jungblut and Neilan, 2006; Kurmayer et al., 2006). In fact, the phylogenetic tree reconstruction based on concatenated amino acid sequences of all complete *mcy* gene clusters so far known (Figure 3) supported a correlation between cyanobacterial taxonomy and MC acquisition in these strains. Taking into account the still low number of available sequences, the Bayesian inference sustained the hypothesis that *mcy* genes might have evolved from a common ancestor and their irregular distribution in phylogenetically related taxa is due to repeated loss processes rather than horizontal transfer (Kurmayer et al., 2004; Rantala et al., 2004).

Multiple alignments of the amino acid residues surrounding the substrate-binding pockets in the NRPS adenylation domains observed in the MC gene cluster of CENA161 with other cyanobacteria are shown in Table 3. Adenylation domains in NRPS modules contain ten highly conserved core motifs, named A1 to A10. Lining the binding pocket, ten conserved amino acid residues are believed to recognize a specific substrate, allowing the prediction of the amino acid likely to be selected for activation (Marahiel et al., 1997; Stachelhaus et al., 1999; Challis et al., 2000; Mikalsen et al., 2003). In the case of *Fischerella* sp. CENA161, the amino acid sequences of the McyG and McyE A domains are highly conserved, containing identical residues as observed in all the MC gene clusters so far known (Table 3). These two modules are responsible for the partial formation of Adda in position **5** and the incorporation of glutamic acid in position **6** of the MC molecule, respectively. In a similar way, the McyA-A_1 and McyA-A_2 modules, responsible for the incorporation of amino acids in positions **7** (serine/MdhA) and **1** (alanine), respectively, are highly conserved. Likewise, McyB-A_2 binding pocket, responsible for the incorporation of amino acids in position **3** (MeAsp), was shown be conserved. On the other hand, McyB-A_1 and McyC-A sequences showed high diversity in the strain analyzed. These two modules are responsible for incorporating amino acids into positions $\mathbf{X^2}$ and $\mathbf{Z^4}$, respectively. Indeed, it is well known that positions $\mathbf{X^2}$ and $\mathbf{Z^4}$ show the highest degree of structural variation when compared to other positions in the molecule (Fewer at al., 2007). Such fluctuations in McyB-A_1 and McyC-A sequences are the major contributors for the diversity in MC biosynthesis in different species. Currently, over a hundred MC structural variants are known, differing in the type of amino acids incorporated or modifications to the peptide backbone (Dittmann et al., 2015).

The LC-MS/MS analysis of the strain CENA161 cell extract allowed the identification of seven MC variants: MC-LR (*m*/*z* 995, $[M+H]^+$), MC-FR (*m*/*z* 1029), MC-LA (*m*/*z* 910), MC-LAba (*m*/*z* 923), MC-LM (*m*/*z* 970), MC-LV (*m*/*z* 938) and MC-LL (*m*/*z* 952) (Supplementary Information Figures S1 and S2). The identification of MC *m/z* 938 was challenging as the product ion spectrum of *m/z* 938 can be well fitted to three isobaric variants. After thorough comparison of the product ion assignments and intensities it seems that variant MC-LV best explained the high resolution spectrum data (Supplementary Table S1). Exact ion masses, accuracies and intensities of MC isoforms are presented in Supplementary Table S2. MC-LR, MC-LV and MC-LL were found to be the major variants produced by strain CENA161, while the remaining variants were produced in trace amounts. The most studied and common variant MC-LR has been previously reported in the extract of *Fischerella* sp. CENA161 (Fiore et al., 2009) and *Fischerella* sp. NQAIF311 (Cirés et al., 2014), whereas the latter also produces the MC-LA and MC-FR variants. The other four variants were previously found in *Microcystis* strains (Craig et al., 1993; Sivonen and Jones, 1999; Diehnelt et al., 2006).

The origin of MC diversity within species has been attributed to recombination events in the *mcy* genes, to the activity of specific tailoring enzymes and to the low substrate specificity of NRPS modules $McyB_1$ and McyC (Kurmayer and Gumpenberger, 2006; Fewer et al., 2007; Fewer et al., 2008; Tooming-Klunderud et al., 2008; Kaasalainen et al., 2012; Fewer et al., 2013; Calteau et al., 2014). As a consequence of this relaxed substrate specificity, different amino acids can be incorporated in positions $\mathbf{X}^2$ and $\mathbf{Z}^4$, allowing a cyanobacterial strain to simultaneously produce several microcystin variants. In this regard, amino acid availability has been associated to the production of different MCs (Tonk et al., 2008; Van de Waal, 2010; Liu et al., 2016). These authors suggested that environmental factors affect the intracellular free amino acid levels which, ultimately, result in changes in the biosynthesis of MCs. Whether such correlations apply to the nitrogen-fixing strain CENA161 remains to be further investigated. Nevertheless, the MC production profile of CENA161 cultivated without inorganic nitrogen was dominated by hydrophobic variants with high C:N ratios. Considering the frequencies of amino acids incorporated in position $\mathbf{X}^2$, leucine was found in six of the seven variants described. These results are in line with the substrate prediction for McyB-A_1 (Table 3). However, McyC-A demonstrated a greater flexibility in terms of substrate selection, loading distantly related amino acids in position $\mathbf{Z}^4$, from hydrophilic arginine (MC-LR and MC-FR) to hydrophobic leucine (MC-LL).

The toxicity of MCs can be associated with the hydrophobicity of their constitutional amino acids. This probably occurs due to the increased ability of hydrophobic variants to get into cells by OATP-mediated transport or due to membrane interactions (Vesterkvist and Meriluoto, 2003; Feurstein et al., 2009; Faassen and Lürling, 2013). More toxic variants may harm cell membranes, damage mitochondrial dehydrogenases, and cause lactate dehydrogenase leakage (Monks et al., 2007; Fischer et al., 2010; Vesterkvist et al., 2012). Therefore, new variants of MC and the elucidation of their respective biosynthetic genes are considered of high interest for public health and pharmaceutical development since peptide structural diversities are reflected in different biological activities (Gupta et al., 2003; Zurawell et al., 2005; Monks et al., 2007; Feurstein et al., 2009; Fischer et al., 2010; Vesterkvist et al., 2012).

4. Conclusions

Identifying and annotating gene clusters responsible by the production of harmful toxic molecules that also have potential pharmacological application is important to the understanding of the process of their synthesis and the rules that govern their evolution, and to exploit their capabilities. The finding that MCs are also produced by subaerophytic cyanobacteria and not only by planktonic species may contribute to bring new insights into the cellular function of these heptapeptides, and as more gene clusters from different cyanobacteria are described it will facilitate the identification of amenable genus to genetic manipulation in order to advance our knowledge of genetic and biology of this toxin.

Funding sources

This work was supported by grants from the São Paulo Research Foundation (FAPESP, 2013/50425-8) to MFF, the Academy of Finland (1273798) to KS and FAPESP (2014/50420-9) and the National Council for Scientific and Technological Development (CNPq, 311048/2016-1) to EP. KH and DOA were supported by graduate fellowships from the Brazilian Federal Agency for the Support and Evaluation of Graduate Education (CAPES, 99999.008344/2014) and FAPESP (2011/08092-6), respectively. MFF and AMV would also like to thank to the National Council for Scientific and Technological Development (CNPq) for research fellowships (310244/2015-3 and 302599/2016-9, respectively). We acknowledge the Center of Functional Genomics Applied to Agriculture and Agroenergy (University of São Paulo, Campus “Luiz de Queiroz”) for generating Illumina MiSeq data.

Conflict of Interest

The authors declare no conflicts of interest.

References

Allen, M.B., 1968. Simple conditions for growth of unicellular blue-green algae on plates. J. Phycol. 4, 1–4.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J., 1997. Gapped BLAST and PSI BLAST: a new generation of protein database search programs. Nucleic Acid Res. 25, 3389–3402.

Andrews S., 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc (accessed 12 December 2016).

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. 19, 455–477.

Birnboim, H.C., Doly, J., 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7, 1513–1518.

Bishop, C.T., Anet, E.F.L.J., and Gorham, P.R., 1959. Isolation and identification of the past-death factor in *Microcystis aeruginosa* NRC-1. Can. J. Biochem. Physiol. 37, 453–471.

Botes, D.P., Tuinman, A.A., Wessels, P.L., Viljoen, C.C., Kruger, H., Williams, D.H., Santikarn, S., Smith, R.J., Hammond, S.J., 1984. The structure of cyanoginosin-LA, a cyclic heptapeptidetoxin from the cyanobacterium *Microcystis aeruginosa*. J. Chem. Soc., Perkin Trans. 1, 2311–2318.

Botes, D.P., Wessels, P.L., Kruger, H., Runnegar, M.T.C., Santikarn, S., Smith, R.J., Barna, J.C.J., Williams, D.H., 1985. Structural studies on cyanoginosins-LR, -YR, -YA, and -YM, peptide toxins from *Microcystis aeruginosa*. J. Chem. Soc., Perkin Trans. 1, 2747–2748.

Calteau, A., Fewer, D.P., Latifi, A., Coursin, T., Laurent, T., Jokela, J., Kerfeld, C.A., Sivonen, K., Piel, J., Gugger, M., 2014. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in Cyanobacteria. BMC Genomics 15, 977.

Challis, G.L., Ravel, J., Townsend, C.A., 2000. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. Chem. Biol. 7, 211–224.

Christiansen, G., Fastner, J., Erhard, M., Börner, T., Dittmann, E., 2003. Microcystin biosynthesis in *Planktothrix*: genes, evolution, and manipulation. J. Bacteriol. 185, 564–572.

Christiansen, G., Molitor, C., Philmus, B., Kurmayer, R., 2008. Nontoxic strains of cyanobacteria are the result of major gene deletion events induced by a transposable element. Mol. Biol. Evol. 25, 1695–1704.

Cirés, S., Alvarez-Roa, C., Wood, S.A., Puddick, J., Loza, V., Heimann, K., 2014. First report of microcystin-producing *Fischerella* sp. (Stigonematales, Cyanobacteria) in tropical Australia. Toxicon 88, 62–66.

Craig, M, McCready, T.L., Luu, H.A., Smillie, M.A., Dubord, P., Holmes, C.F.B., 1993. Identification and characterization of hydrophobic microcystins in Canadian freshwater cyanobacteria. Toxicon 31, 1541−1549.

Dagan, T., Roettger, M., Stucken, K., Landan, G., Koch, R., Major, P., Gould, S.B., Goremykin, V. V., Rippka, R., de Marsac, N.T., Gugger, M., Lockhart, P.J., Allen, J.F., Brune, I., Maus, I., Pühler, A., Martin, W.F., 2012. Genomes of stigonematalean Cyanobacteria (Subsection V) and the evolution of pxygenic photosynthesis from prokaryotes to plastids. Genome Biol. Evol. 5, 31–44.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

Diehnelt, C.W., Dugan, N.R., Peterman, S.M., Budde, W.L., 2006. Identification of microcystin toxins from a strain of *Microcystis aeruginosa* by liquid chromatography introduction into a hybrid linear ion trap Fourier transform ion cyclotron resonance mass spectrometer. Anal. Chem. 78, 501–512.

Dittmann, E., Neilan, B.A., Erhard, M., von Döhren, H., Börner, T., 1997. Insertional mutagenesis of a peptide synthetase gene that is responsible for hepatotoxin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Mol. Microbiol. 26, 779–787.

Dittmann, E., Gugger, M., Sivonen, K., Fewer, D.P., 2015. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. Trends Microbiol. 23, 642–652.

Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5, 113.

Eriksson, J.E., Meriluoto, J.A.L., Kujari, H.P., Skulberg, O.M., 1988. A comparison of toxins isolated from the cyanobacteria *Oscillatoria agardhii* and *Microcystis aeruginosa*. Comp. Biochem. Physiol. 89, 207–210.

Ewing, B., Hillier, L., Wendl, M.C., Green, P., 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res. 8 (3), 175–185.

Ewing, B., Green, P., 1998. Base-calling of automated sequencer traces using Phred. II, Error probabilities. Genome Res. 8, 186–194.

Faassen, E. J., Lürling, M., 2013. Occurrence of the microcystins MC-LW and MC-LF in Dutch surface waters and their contribution to total microcystin toxicity. Mar. Drugs 11, 2643-2654.

Feurstein, D., Holst, K., Fischer, A., Dietrich D.R., 2009. Oatp-associated uptake and toxicity of microcystins in primary murine whole brain cells. Toxicol. Appl. Pharmacol. 234, 247–255.

Fewer, D.P., Rouhiainen, L., Jokela, J., Wahlsten, M., Laakso, K., Wang, H., Sivonen, K., 2007. Recurrent adenylation domain replacement in the microcystin synthetase gene cluster. BMC Evol. Biol. 7, 183.

Fewer, D.P., Tooming-Klunderud, A., Jokela, J., Wahlsten, M., Rouhiainen, L., Kristensen, T., Rohrlack, T., Jakobsen, K.S., Sivonen, K., 2008. Natural occurrence of microcystin synthetase deletion mutants capable of producing microcystins in strains of the genus *Anabaena* (Cyanobacteria). Microbiology 154, 1007–1014.

Fewer, D.P., Wahlsten, M., Österholm, J., Jokela, J., Rouhiainen, L., Kaasalainen, U., Rikkinen, J., and Sivonen, K., 2013. The genetic basis for *O*-acetylation of the microcystin toxin in Cyanobacteria. Chem. Biol. 20, 861–869.

Fiore, M.F., Genuário, D.B., Silva, C.S.P., Shishido, T.K., Moraes, L.A.B., Cantúsio-Neto, R., Silva-Stenico, M.E., 2009. Microcystin production by a freshwater spring cyanobacterium of the genus *Fischerella*. Toxicon 53, 754–761.

Fischer, A., Hoeger, S.J., Stemmer, K., Feurstein, D.J., Knobeloch, D., Nussler, A., Dietrich, D.R., 2010. The role of organic anion transporting polypeptides (OATPs/SLCOs) in the toxicity of different microcystin congeners *in vitro*: A comparison of primary human hepatocytes and OATP-transfected HEK293 cells. Toxicol. Appl. Pharmacol. 245, 9–20.

Gordon, D., Abajian, C., Green, P., 1998. Consed: A graphical tool for sequence finishing. Genome Res. 8, 195–202.

Gulledge, B.M., Aggen, J.B., Huang, H.B., Nairn, A.C., Chamberlin, A.R., 2002. The microcystins and nodularins: cyclic polypeptide inhibitors of PP1 and PP2A. Curr. Med. Chem. 9, 1991–2003.

Gupta, N., Pant, S.C., Vijayaraghavan, R., Rao, P.V., 2003. Comparative toxicity evaluation of cyanobacterial cyclic peptide toxin microcystin variants (LR, RR, YR) in mice. Toxicology 188, 285–296.

Harada, K.-I., Ogawa, K., Kimura, Y., Murata, H., Suzuki, M., Thorn, P.M., Evans, W.R., Carmichael, W.W., 1991. Microcystins from *Anabaena flos-aquae* NRC 525-17. Chem. Res. Toxicol. 4, 535–540.

Heck, K., Machineski, G.S., Alvarenga, D.O., Vaz, M.G.M.V., Varani, A.M., Fiore, M.F., 2016. Evaluating methods for purifying cyanobacterial cultures by qPCR and high-throughput Illumina sequencing. J. Microbiol. Meth. 129, 55-60.

Hicks, L.M., Moffitt, M.C., Beer, L.L., Moore, B.S., Kelleher, N.L., 2006. Structural characterization of in vitro and in vivo intermediates on the loading module of microcystin synthetase. ACS Chem. Biol. 1, 93–102.

Hirose, Y., Fujisawa, T., Ohtsubo, Y., Katayama, M., Misawa, N., Wakazuki, S., Shimura, Y., Nakamura, Y., Kawachi, M., Yoshikawa, H., Eki, T., Kanesaki, Y., 2016. Complete genome sequence of cyanobacterium *Fischerella* sp. NIES-3754, providing thermoresistant optogenetic tools. J. Biotechnol. 220, 45–46.

Izaguirre, G., Jungblut, A.D., Neilan, B.A., 2007. Benthic cyanobacteria (Oscillatoriaceae) that produce microcystin-LR, isolated from four reservoirs in southern California. Water Res. 41, 492–498.

Jochimsen, E.M., Carmichael, W.W., An, J., Cardo, D., Cookson, S.T., Holmes, C.E.M., Antunes, M.B.C., Melo Filho, D.A., Lyra, T.M., Barreto, V., Azevedo, S.M.F.O., Jarvis, W.R., 1998. Liver failure and death following exposure to microcystin toxins at a hemodialysis center in Brazil. N. Engl. J. Med. 13, 873–878.

Jungblut, A.D., Neilan, B.A., 2006. Molecular identification and evolution of the cyclic peptide hepatotoxins, microcystin and nodularin, synthetase genes in three orders of cyanobacteria. Arch. Microbiol. 185, 107–114.

Kaasalainen, U., Fewer, D.P., Jokela, J., Wahlsten, M., Sivonen, K., Rikkinen, J., 2012. Cyanobacteria produce a high variety of hepatotoxic peptides in lichen symbiosis. Proc. Natl. Acad. Sci. USA 109, 5886–5891.

Kounnis, V., Chondrogiannis, G., Mantzaris, M.D., Tzakos, A.G., Fokas, D., Papanikolaou, N.A., Galani, V., Sainis, I., Briasoulis, E., 2015. Microcystin LR shows cytotoxic activity against pancreatic cancer cells expressing the membrane OATP1B1 and OATP1B3 transporters. Anticancer Res. 35, 5857–5865.

Krishnamurthy, T., Carmichael, W.W., Sarver, E.W., 1986. Investigation of freshwater cyanobacteria (blue-green algae) toxic peptide. I. Isolation, purification and characterization of peptides from *Microcystis aeruginosa* and *Anabaena flos-aquae*. Toxicon 24, 865–873.

Kurmayer, R., Christiansen, G., Fastner, J., Börner, T., 2004. Abundance of active and inactive microcystin genotypes in populations of the toxic cyanobacterium *Planktothrix* spp. Environ. Microbiol. 6, 831–841.

Kurmayer, R., Christiansen, G., Gumpenberger, M., Fastner, J., 2005. Genetic identification of microcystin ecotypes in toxic cyanobacteria of the genus *Planktothrix*. Microbiology 151, 1525-1533.

Kurmayer, R., Gumpenberger, M., 2006. Diversity of microcystin genotypes among populations of the filamentous cyanobacteria *Planktothrix rubescens* and *Planktothrix agardhii*. Mol. Ecol. 15, 3849–3861.

Liu, J., Van Oosterhout, E., Faassen, E.J., Lürling, M., Helmsing, N.R., Van de Waal, D.B., 2016. Elevated $p\mathrm{CO}_2$ causes a shift towards more toxic microcystin variants in nitrogen-limited *Microcystis aeruginosa*. FEMS Microbiol. Ecol. 92, fiv159.

MacKintosh, C., Beattie, K.A., Klumpp, S., Cohen, P., Codd, G.A., 1990. Cyanobacterial microcystin-LR is a potent and specific inhibitor of protein phosphatases 1 and 2A from both mammals and higher plants. FEBS Lett. 264, 187–192.

Marahiel, M.A., Stachelhaus, T., Mootz, H.D., 1997. Modular peptide synthetases involved in nonribosomal peptide synthesis. Chem. Rev. 97, 2651–2674.

Meriluoto, J.A.O., Sandstrom, A., Eriksson, J.E., Remaud, G., Craig, A.G., Chattopadhyaya, J., 1989. Structure and toxicity of a peptide hepatotoxin from the cyanobacterium *Oscillatoria agardhii*. Toxicon 27, 1021–1034.

Mikalsen, B., Boison, G., Skulberg, O.M., Fastner, J., Davies, W., Gabrielsen, T.M., Rudi, K. Jakobsen, K.S., 2003. Natural variation in the microcystin synthetase operon mcyABC and impact on microcystin production in *Microcystis* strains. J. Bacteriol. 185, 2774–2785.

Moffitt, M.C., Neilan, B.A., 2004. Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. Appl. Environ. Microbiol. 70, 6353–6362.

Monks, N.R., Liu, S., Xu, Y., Yu, H., Bendelow, A.S., Moscow, J.A., 2007. Potent cytotoxicity of the phosphatase inhibitor microcystin LR and microcystin analogues in OATP1B1- and OATP1B3-expressing HeLa cells. Mol. Cancer Ther. 6, 587–598.

Niedermeyer, T.H.J., Daily, A., Swiatecka-Hagenbruch, M., Moscow, J.A., 2014. Selectivity and potency of microcystin congeners against OATP1B1 and OATP1B3 expressing cancer cells. PLoS ONE 9(3*)*, e91476.

Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K., Ochi, K., Shirai, M., 2000. Polyketide synthase gene coupled to the peptide synthetase module involved in the biosynthesis of the cyclic heptapeptide microcystin. J. Biochem. 127, 779–789.

Omidi, A., Esterhuizen-Londt, M., Pflugmacher, S., 2017. Still challenging: the ecological function of the cyanobacterial toxin microcystin – what we know so far. Toxin Reviews, http://dx.doi.org/10.1080/15569543.2017.1326059.

Pearson, L.A., Hisbergues, M., Börner, T., Dittmann, E., Neilan, B.A., 2004. Inactivation of an ABC transporter gene, *mcyH*, results in loss of microcystin production in the cyanobacterium *Microcystis aeruginosa* PCC 7806. Appl. Environ. Microbiol. 70, 6370–6378.

Pearson, L.A., Barrow, K.D., Neilan, B.A., 2007. Characterization of the 2-hydroxy-acid dehydrogenase McyI, encoded within the microcystin biosynthesis gene cluster of *Microcystis aeruginosa* PCC7806. J. Biol. Chem. 282, 4681–4692.

Prinsep, M.R., Caplan, F.R., Moore, R.E., Patterson, G.M.L., Honkanen, R.E., Boynton, A.L., 1992. Microcystin-LA from a blue-green alga belonging to the Stigonematales. Phytochemistry 31, 1247–1248.

Rantala, A., Fewer, D.P., Hisbergues, M., Rouhiainen, L., Vaitomaa, J., Börner, T., Sivonen, K., 2004. Phylogenetic evidence for the early evolution of microcystin synthesis. Proc. Natl. Acad. Sci. USA 101, 568–573.

Rausch, C., Weber, T., Kohlbacher, O., Wohlleben, W., Hudson, D.H., 2005. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. 33, 5799–5808.

Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Röttig, M., Medema, M.H., Bin, K., Weber, T., Rausch, C., Kohbacher, O., 2011. NRPSpredictor2 — a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, 362–367.

Rouhiainen, L., Vakkilainen, T., Lumbye-Siemer, B., Buikema, W., Haselkorn, R., Sivonen, K., 2004. Genes coding for hepatotoxic heptapeptides (microcystins) in the cyanobacterium *Anabaena* strain 90. Appl. Environ. Microbiol. 70, 686–692.

Rounge, T.B., Rohrlack, T., Nederbragt, A.J., Kristensen, T., Jakobsen, K.S., 2009. A genome-wide analysis of nonribosomal peptide synthetase gene clusters and their peptides in a *Planktothrix rubescens* strain. BMC Genomics 10, 396.

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., Rice, P., Rajandream, M.A., Barrell, B., 2000. Artemis: sequence visualization and annotation. Bioinformatics 16(10), 944–945.

Shih, P.M., Wu, D., Latifid, A., Axen, S.D., Fewer, D.P., Talla, E., Calteau, A., Cai, F., De Marsac, N.T., Rippka, R., Herdman, M., Sivonen, K., Coursin, T., Laurent, T., Goodwin, L., Nolan, M., Davenport, K.W., Han, C.S., Rubin, E.M., Eisen, J.A., Woyke, T., Gugger, M., Kerfeld, C.A.,

2013. Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. Proc. Natl. Acad. Sci. USA 110, 1053–1058.

Sielaff, H., Dittmann, E., de Marsac, N.T., Bouchier, C., von Döhren, H., Börner, T., Schwecke, T., 2003. The *mcyF* gene of the microcystin biosynthetic gene cluster from *Microcystis aeruginosa* encodes an aspartate racemase. Biochem. J. 373, 909–916.

Sivonen, K., Carmichael, W.W., Namikoshi, M., Rinehart, K.L., Dahlem, A.M., Niemela, S.I., 1990. Isolation and characterization of hepatotoxic microcystin homologs from the filamentous freshwater cyanobacterium *Nostoc* sp. strain 152. Appl. Environ. Microbiol. 56, 2650–2657.

Sivonen, K., Jones, G., 1999. Cyanobacterial toxins, in: Chorus, I., Bartram, J. (Eds.). Toxic cyanobacteria in water: a guide to their public health significance, monitoring and management. E & FN Spon, London, pp. 41–111.

Stachelhaus, T., Mootz, H.D., Marahiel, M.A., 1999. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chem. Biol. 6, 493–505.

Tillett, D., Dittman, E., Erhard, M., von Döhren, H., Borner, T., Neilan, B., 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC 7806: as integrated peptide-polyketide synthetase system. Chem. Biol. 7, 753–764.

Tonk, L., Van de Waal, D.B., Slot, P., Huisman, J., Matthijs, H.C.P., Visser, P.M., 2008. Amino acid availability determines the ratio of microcystin variants in the cyanobacterium *Planktothrix agardhii*. FEMS Microbiol. Ecol. 65, 383–390.

Tooming-Klunderud, A., Mikalsen, B., Kristensen, T., Jakobsen, K.S., 2008. The mosaic structure of the *mcyABC* operon in *Microcystis*. Microbiology 154, 1886–1899.

Vesterkvist, P.S.M. Meriluoto, J.A.O., 2003. Interaction between microcystins of different hydrophobicities and lipid monolayers. Toxicon 41, 349–355.

Vesterkvist, P.S.M., Misiorek, J.O., Spoof, L.E.M., Toivola, D.M., Meriluoto, J.A.O., 2012. Comparative cellular toxicity of hydrophilic and hydrophobic microcystins on Caco-2 cells. Toxins 4, 1008–1023.

Van de Waal, D.B., Ferreruela, G., Tonk, L., Van Donk, E., Huisman, J., Visser, P.M., Matthijs, H.C.P., 2010. Pulsed nitrogen supply induces dynamic changes in the amino acid composition and microcystin production of the harmful cyanobacterium *Planktothrix agardhii*. FEMS Microbiol. Ecol. 74, 430-438.

von Döhren, H., Keller, U., Vater, J., Zocher, R., 1997. Multifunctional peptide synthetases. Chem. Rev. 29, 2675–2706.

Zhang, J., Kobert, K., Flouri, T., Stamatakis, A., 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30(5), 614–620.

Zhbannikov, I.Y., Hunter, S.S., Foster, J., Settles, M.L., 2015. SeqyClean: a Software Tool for Comprehensive Preprocessing of Sequence Data. https://github.com/ibest/seqyclean (accessed 15 February 2017).

Zurawell, R.W., Chen, H., Burke, J.M., Prepas, B.R., 2005. Hepatotoxic cyanobacteria: A review of the biological importance of microcystins in freshwater environment. J. Toxicol. Environ. Health B 8, 1–37.

Table 1. Thermal cycling programs used to amplify microcystin synthetase gene fragments.

Gene	Program	Primers	Reference
mcyA	94 ºC/4 min; 30x 94ºC/20 s; 55 ºC/30 s; 72 ºC/1 min; 72 ºC/7 min	OMET-F, OMET-R MSF, MSR	Tillet et al., 2000
mcyB	95 ºC/3 min; 30x 94 ºC/30 s; 52 ºC/30 s; 72ºC/1 min; 72ºC/10 min	PB3F, pB9R	Fewer et al., 2007
mcyB (gap)	95 ºC/3 min; 30x 94ºC/30 s; 52 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min		This work
mcyC	95 ºC/3 min; 30x 94 ºC/30 s; 52 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min	pC1F, pC13R	Fewer et al., 2007
mcyC (gap)	95 ºC/3 min; 30x 94 ºC/30 s; 52 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min		This work
mcyD	95 ºC/3 min; 30x 94 ºC/30 s; 56 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min	mcyDF, mcyDR	Rantala et al., 2004
mcyE	95 ºC/3 min; 30x 94 ºC/30 s; 56 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min	mcyEF2, mcyER4	Rantala et al., 2004
mcyF	95 ºC/3 min; 30x 94 ºC/30 s; 50 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min	mcyFKF, mcyFKR	This work
mcyG	95 ºC/3 min; 30x 94 ºC/30 s; 56 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min.	mcyGF, mcyGR	Fewer et al., 2007
mcyH	95 ºC/3 min; 30x 94 ºC/30 s; 56ºC/30 s; 72 ºC/1 min; 72 ºC/10 min.	mcyHKF, mcyHKR	This work
mcyI	95 ºC/3 min; 30x 94 ºC/30 s; 56 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min	mcyIdgenF, mcyIdgenR	Pearson et al., 2007
mcyJ	95 ºC/3 min; 30x 94 ºC/30 s; 56 ºC/30 s; 72 ºC/1 min; 72 ºC/10 min	mcyJKF, mcyJKR	This work

Table 2. Functions of proteins encoded in the microcystin biosynthetic gene cluster.

Protein	Lengths (amino acids)	Functions	Top BLAST Hit		
			Organism	**Identity (%)**	**Accession Number**
McyH	590	ABC transporter	*Anabaena* sp. 90	81.8	AAO62579
McyI	342	putative dehydrogenase	*N. spumigena* NSOR10	86.1	AAO62580
McyF	262	amino acid racemase	*Nostoc* sp. 152	81.6	AGZ05271
McyE	3,511	NRPS-PKS (KS-AT-PCP-AMT-C-A-PCP-C)	*Nostoc* sp. 152	80.1	AGZ05272
McyD	3,872	PKS (KS-DH-CM-KR-PCP-KS-AT-DH-KR-PCP)	*Anabaena* sp. 90	79.2	AAO62584
McyG	2,639	NRPS-PKS (A-PCP-KS-AT-CM-KR-PCP)	*Anabaena* sp. 90	79.9	AAO62585
McyA	2,787	NRPS (A-NMT-PCP-C-A-PCP-E)	*Anabaena* sp. 90	81.5	AAO62586
McyB	2,136	NRPS (C-A-PCP-C-A-PCP)	*Anabaena* sp. 90	83.1	AAO62587
McyC	1,283	NRPS (C-A-PCP-Te)	*Anabaena* sp. 90	83.3	AAO62588
McyJ	313	*O*-acetyltransferase	*N. spumigena* NSOR10	84.8	AAO64406

Table 3. Conservation of the seven adenylation domain binding pockets of Mcy of *Fischerella* sp. CENA161and other cyanobacteria.

Strain/Binding pocket	$McyA_1$	$McyA_2$	$McyB_1$	$McyB_2$	McyC	McyE	McyG
Fischerella sp. CENA161	DVWHISLIDK Ser 100%	DLFNNALTYK Ala 100%	DVLIFGLIYK Leu 70%	DARHVGIFVK Tyr 60%	DVWFFGLVDK Ser 80%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Anabaena sp. 90	---------- Ser 100%	---------- Ala 100%	--WF---VD- Leu 80%	---------- Tyr 60%	---C------ Ser 80%	DPRHSGVVGK Glu 100%	**LWVAASGK Tcl 60%
Nostoc sp. 152	---------- Ser 100%	---------- Ala 100%	-A-F------ Leu 80%	---------- Tyr 60%	---N—FI--- Gln 70%	DPRHSGVVGK Glu 100%	**LWVAASGK Tcl 60%
Planktothrix agardhii NIVA-CYA 126/8	---------- Ser 100%	-------S-- Ala 90%	-A-F---VD- Leu 70%	---------- Glu 60%	-P-G------ Gln 70%	DPRHSGVVGK Glu 100%	AILWVAASG* Tcl 60%
Planktothrix rubescens NIVA-CYA 98	-F-N-GMVH- Thr 100%	---------- Ala 100%	-A-F---VD- Leu 100%	-P------I- Glu 60%	-P-G------ Gln 70%	DPRHSGVVGK Glu 100%	AILWVAASG* Tcl 60%
Microcystis aeruginosa SPC777	----F----- Ser 100%	---------- Ala 100%	-GWTI-AVE- Arg 90%	---------- Tyr 60%	---TI-A--- Arg 100%	#	#
Microcystis aeruginosa NIES-843	----F----- Ser 100%	---------- Ala 100%	-GWTI-AVE- Arg 90%	---------- Tyr 60%	---TI-A--- Arg 100%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Microcystis aeruginosa PCC 7806	----F----- Ser 100%	---------- Ala 100%	-AWFL-NVV- Arg 90%	---------- Tyr 60%	---TI-A--- Arg 100%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Microcystis aeruginosa DIANCHI 905	----F----- Ser 100%	---------- Ala 100%	-AWFL-NVV- Leu 100%	---------- Tyr 60%	---TI-A--- Arg 100%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Microcystis aeruginosa K-139	----F----- Ser 100%	---------- Ala 100%	-AWFL-NVV- Leu 100%	---------- Tyr 60%	---TI-A--- Arg 100%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Microcystis aeruginosa PCC7941	----F----- Ser 100%	---------- Ala 100%	-AWFL-NVV- Leu 100%	---------- Tyr 60%	---TI-A--- Arg 100%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Microcystis aeruginosa PCC9807	----F----- Ser 100%	---------- Ala 100%	-AWFL-NVV- Leu 100%	---------- Tyr 60%	---TI-A--- Arg 90%	DPRHSGVVGK Glu 100%	**LWVAASG* Tcl 50%
Nodularia spumigena NSOR10[1]	-F-N-GMVH- Thr 100%	-	-	---------- Tyr 60%	---N—-F--- Glu 70%	DPRHSGVVGK Glu 100%	--LWVAASGK Tcl 60%

The adenylation domain is responsible for recognition and activation of amino acids in the microcystins. The probability of incorporation for each amino acid is shown.

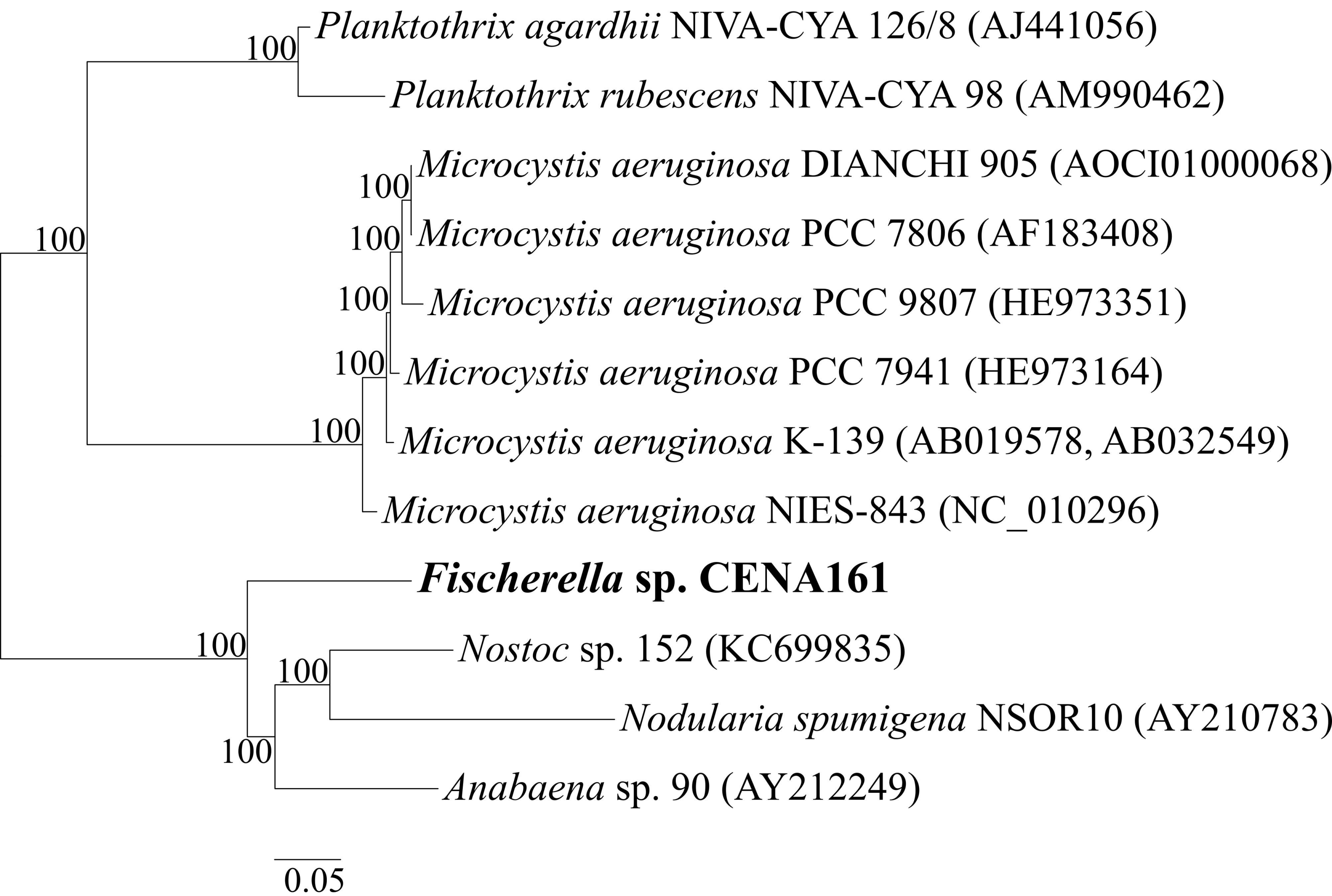
There is no available information about that sequence; * Unknown amino acid; **-** There is no homologue gene; [1] Nodularin-producing strain.

Figure Legends

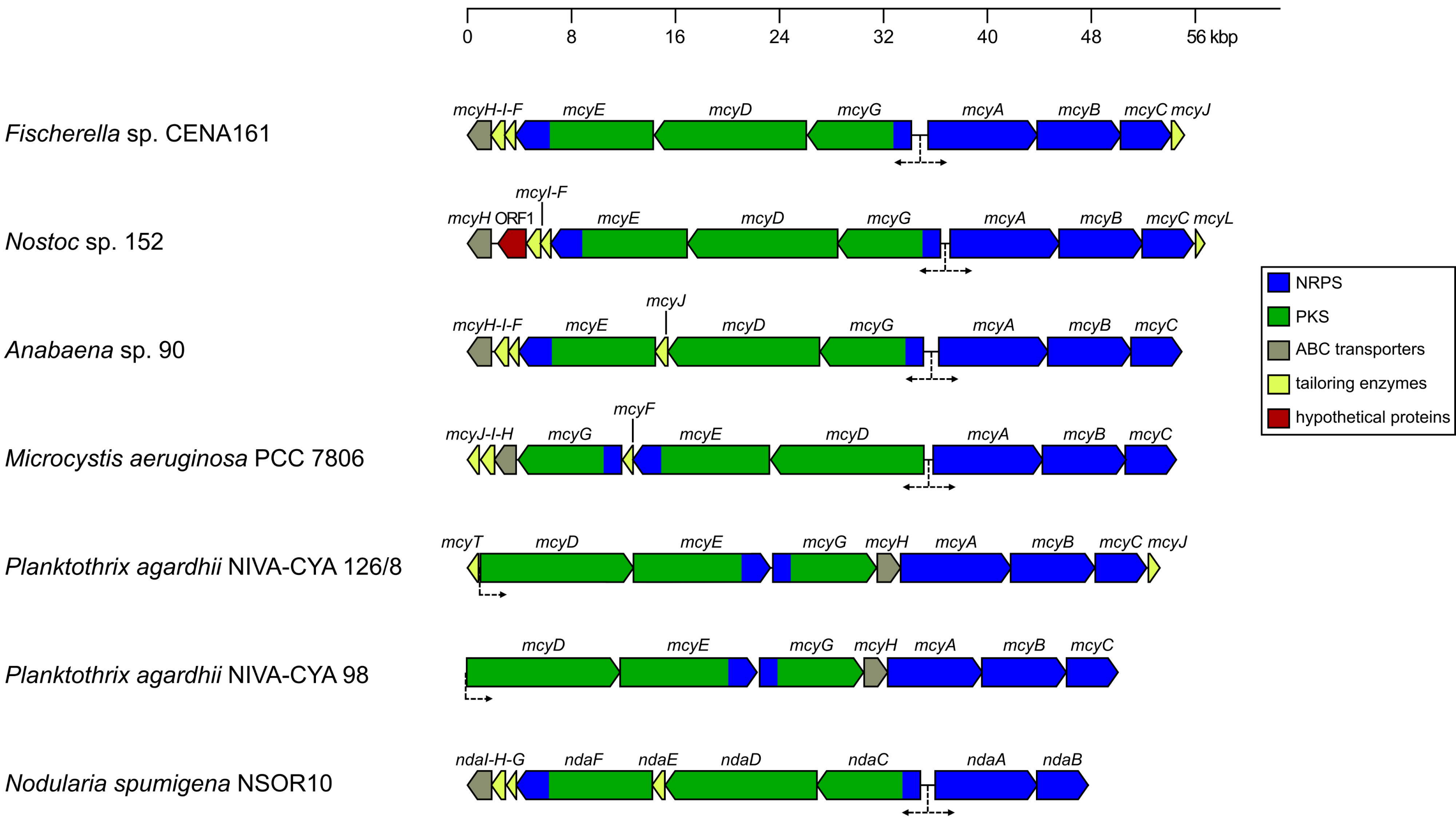
Figure 1. Structure of MC-LR. The *Fischerella* sp. CENA161 MC variations encountered are shown schematically according to their position. Abbreviations: Adda, (2S,3S,8S,9S)-3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-(4*E*,6*E*)-dienoic acid; D-Glu, glutamic acid; Mdha, N-methyl-α-β-dehydroalanine; Ala, alanine, Leu, leucine, Phe, phenylalanine, D-MeAsp, D-erythro-β-methyl-aspartic acid, Aba, Aminoisobutyric acid, Arg, arginine, Val, valine and Met, methionine.

Figure 2. Arrangement of gene clusters coding for the biosynthesis of microcystin in *Fischerella* sp. CENA161, *Nostoc* (Fewer et al., 2013), *Anabaena* (Rouhiainen et al., 2004), *Microcystis* (Nishizawa et al., 2000; Tillett et al., 2000), *Planktothrix* (Christiansen et al., 2003; Rounge et al., 2009), and of nodularin in *Nodularia* (Moffitt and Neilan, 2004). Arrows indicate the transcriptional start sites from the putative promoter regions.

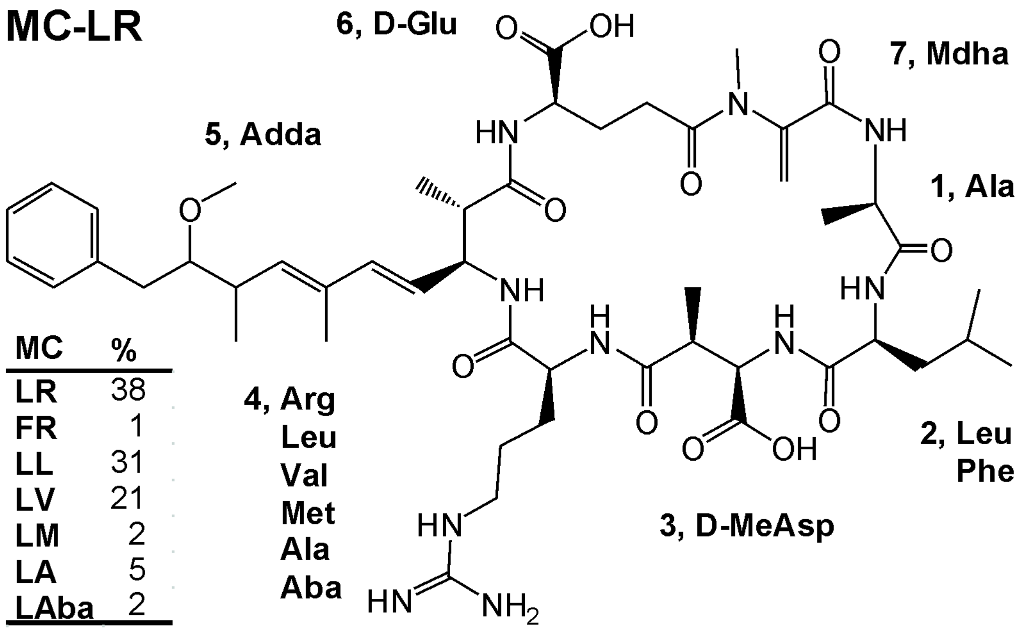
Figure 3. Bayesian inference phylogenetic tree reconstructed from concatenated Mcy and Nda amino acid sequences from strains presenting the microcystin gene cluster or nodularin gene cluster. Posterior probabilities are shown in the nodes.



100
Planktothrix agardhii NIVA-CYA 126/8 (AJ441056)
Planktothrix rubescens NIVA-CYA 98 (AM990462)
100
Microcystis aeruginosa DIANCHI 905 (AOCI01000068)
100
Microcystis aeruginosa PCC 7806 (AF183408)
100
Microcystis aeruginosa PCC 9807 (HE973351)
100
Microcystis aeruginosa PCC 7941 (HE973164)
100
Microcystis aeruginosa K-139 (AB019578, AB032549)
100
Microcystis aeruginosa NIES-843 (NC_010296)
100
***Fischerella* sp. CENA161**
100
Nostoc sp. 152 (KC699835)
100
Nodularia spumigena NSOR10 (AY210783)
100
Anabaena sp. 90 (AY212249)
0.05



0
8
16
24
32
40
48
56 kbp
Fischerella sp. CENA161
mcyH-I-F
mcyE
mcyD
mcyG
mcyA
mcyB
mcyC
mcyJ
Nostoc sp. 152
mcyI-F
mcyH
ORF1
mcyE
mcyD
mcyG
mcyA
mcyB
mcyC
mcyL
Anabaena sp. 90
mcyJ
mcyH-I-F
mcyE
mcyD
mcyG
mcyA
mcyB
mcyC
Microcystis aeruginosa PCC 7806
mcyF
mcyJ-I-H
mcyG
mcyE
mcyD
mcyA
mcyB
mcyC
Planktothrix agardhii NIVA-CYA 126/8
mcyT
mcyD
mcyE
mcyG
mcyH
mcyA
mcyB
mcyC
mcyJ
Planktothrix agardhii NIVA-CYA 98
mcyD
mcyE
mcyG
mcyH
mcyA
mcyB
mcyC
Nodularia spumigena NSOR10
ndaI-H-G
ndaF
ndaE
ndaD
ndaC
ndaA
ndaB
NRPS
PKS
ABC transporters
tailoring enzymes
hypothetical proteins



MC	%
LR	38
FR	1
LL	31
LV	21
LM	2
LA	5
LAba	2

Supplementary Information

Biosynthesis of Microcystin Hepatotoxins in the Cyanobacterial Genus *Fischerella*

Karina Heck[1], Danillo O. Alvarenga[1], Tania K. Shishido[2], Alessandro M. Varani[3], Felipe A. Dörr[4] , Ernani Pinto[4], Leo Rouhiainen[2], Jouni Jokela[2], Kaarina Sivonen[2], Marli F. Fiore[1]

[1]University of São Paulo (USP), Center for Nuclear Energy in Agriculture, Piracicaba-SP, Brazil

[2]University of Helsinki, Department of Food and Environmental Sciences, Division of Microbiology and Biotechnology, Helsinki, Finland

[3]São Paulo State University (UNESP), School of Agricultural and Veterinary Sciences, Department of Technology, Jaboticabal-SP, Brazil

[4]University of São Paulo (USP), School of Pharmaceutical Sciences, São Paulo-SP, Brazil

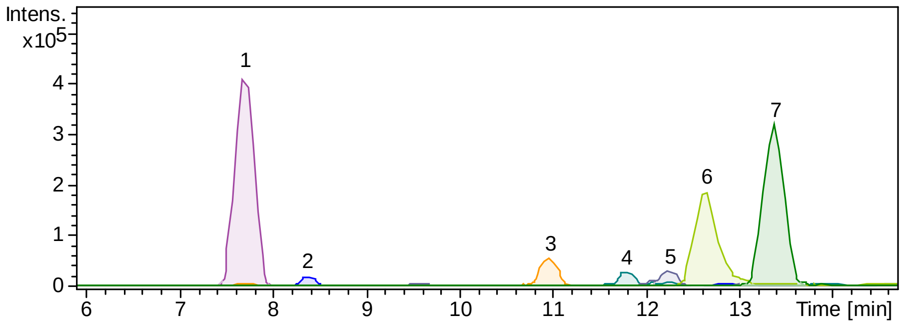


Figure S1. LC-MS representative extracted ion chromatograms for MC isoforms detected in *Fischerella* sp. CENA161 after negative mode electrospray ionization. $[M-H]^-$ ions were extracted for MC-LR at *m/z* 993 (**1**); MC-FR at *m/z* 1027 (**2**); MC-LA at *m/z* 908 (**3**); MC-LAba at *m/z* 922 (**4**); MC-LM at *m/z* 968 (5); (6) MC-LV at *m/z* 936 and (7) MC-LL at *m/z* 950. Relative amounts of each isoform were determined by peak area.

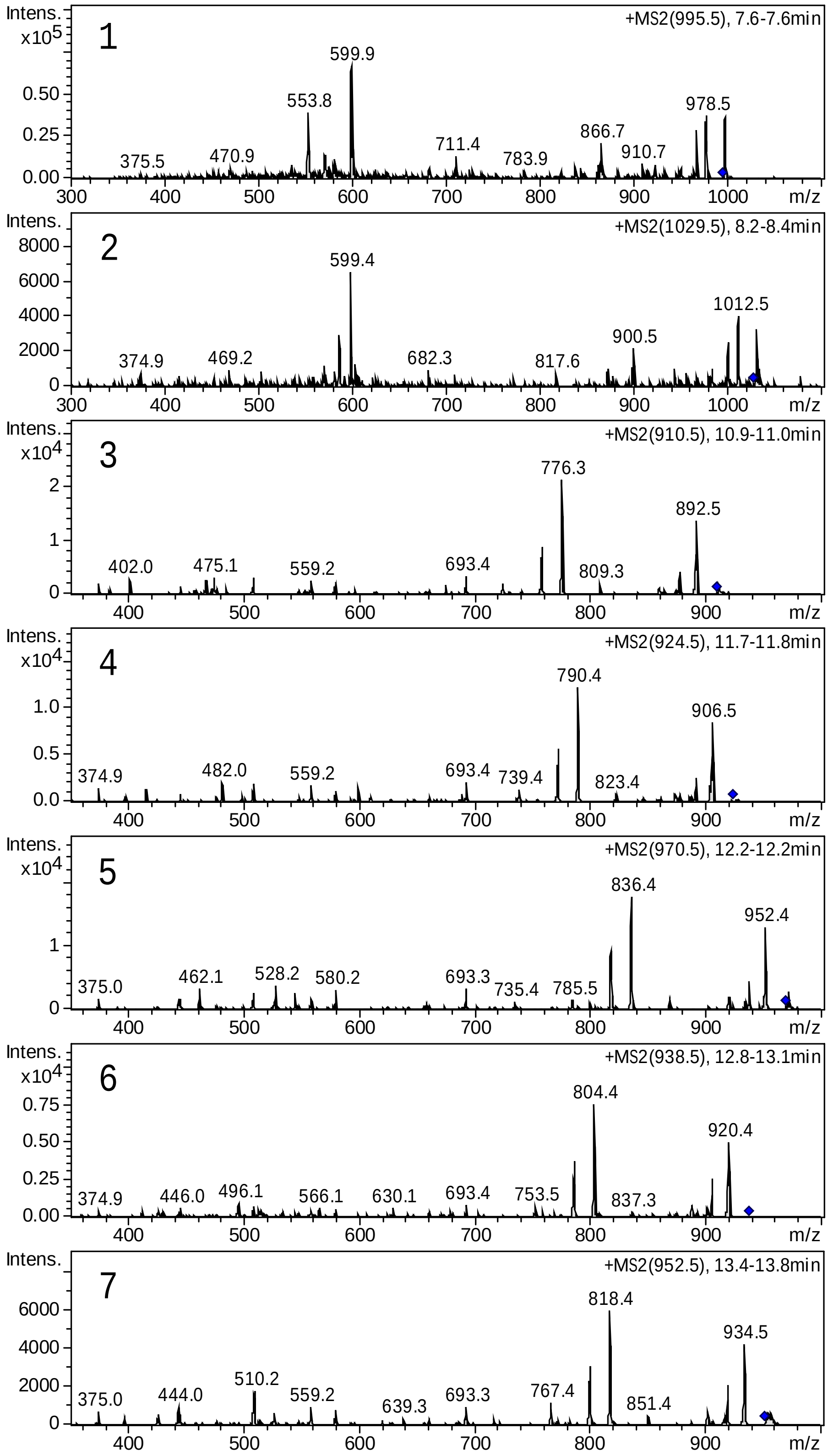


Figure S2. Characteristic ion trap collision-induced dissociation (CID) spectra for MC isoforms identified in *Fischerella* sp. CENA161 after positive mode electrospray ionization. (**1**) MC-LR at *m/z* 995; (**2**) MC-FR at *m/z* 1029; (**3**) MC-LA at *m/z* 910; (**4**) MC-LAba at *m/z* 924; (**5**) MC-LM at *m/z* 970; (**6**) MC-LV at *m/z* 938 and (**7**) MC-LL at *m/z* 952.

Supplementary Table S1. Ion assignments, accuracies (Δ, ppm) and intensities (I, %) of UPLC-ESI-QTOF product ion spectra of protonated MCs detected in *Fischerella* sp. CENA161. Blue bars show the relative intensities of product ions. Isomeric variants MC-LV, MC-VL and [D-Asp^3]MC-LL were fitted to spectral data from protonated MC ion at *m/z* 938 and abbreviations from the presence (or intensity) of product ions are marked with orange cells showing that ion data from *m/z* 938 is best explained by MC-LV structure.

Supplementary Table S1. Ion assignments, accuracies (Δ, ppm) and intensities (I, %) of UPLC-ESI-QTOF product ion spectra of protonated MCs detected in *Fischerella* sp. CENA161. Blue bars show the relative intensities of product ions. Isomeric variants MC-LV, MC-VL and [D-Asp^{3}]MC-LL were fitted to spectral data from protonated MC ion at *m/z* 938 and abbreviations from the presence (or intensity) of product ions are marked with orange cells showing that ion data from *m/z* 938 is best explained by MC-LV structure.

Product ion assignment: Aa^X-Aa^X... (X = 1 - 7)	Neutral losses	MC-LL Δ ppm	MC-LL I (%)	MC-LM Δ ppm	MC-LM I (%)	MC-LA Δ ppm	MC-LA I (%)	MC-Laba Δ ppm	MC-Laba I (%)	MC-LV Δ ppm	MC-LV I (%)	MC-VL Δ ppm	MC-VL I (%)	[Asp^3]MC-LL Δ ppm	[Asp^3]MC-LL I (%)
1-2-3-4-5-6-7	H_2O	-3,2	24	-0,7	24	-3,2	24	-0,3	21	-2,8	26	-2,8	26	-2,8	26
1-2-3-4-5-6-7	NH_3	0,9	25	5,5	22	-0,2	24	0,0	22	0,2	25	0,2	25	0,2	25
1-2-3-4-5-6-7	CH_3OH	-2,0	10	-11,2	10	-0,3	13	-10,6	8	-9,2	11	-9,2	11	-9,2	11
1-2-3-4	-	-1,9	19	-0,1	17	1,1	16	-1,0	12	-2,2	21	-2,2	21	-2,2	21
1-2-3-4-NH_2	-	-1,7	31	-3,4	37	0,7	35	-4,3	20	-4,6	32	-4,6	32	-4,6	32
1-2-3-4	CO	-2,7	15			-26,0	5	-5,1	5	-4,1	11	-4,1	11	-4,1	11
1-2-3	CO, H_2O	-3,6	17	0,5	14	1,2	26	-9,9	14	-5,7	19				
1-2	CO	1,0	4			6,4	5	2,9	3	-2,5	6			-2,5	6
2-3-4	-	0,6	10	2,6	9	4,6	13	-14,6	7	-2,0	12	-2,0	12	-2,0	12
2-3-4-NH_2	-	-0,9	12	-5,5	15	1,8	21	-8,0	8	-5,4	13	-5,4	13	-5,4	13
2-3-4	CO	-3,3	5			7,6	3	-1,3	2	2,0	10	2,0	10	2,0	10
2-3-4	CO, H_2O	-2,3	8	2,3	2	1,2	26	7,4	22			-6,1	9	-6,1	9
2-3	-	-1,0	27	0,5	8	-2,6	5	13,8	4	-6,9	7	-4,9	22	-4,9	22
2-3-NH_2	-	-2,2	15									7,4	44	7,4	44
2-3	CO	-0,6	45			5,5	1					-3,6	50	-3,6	50
2-3	CO, H_2O	-1,3	9			12,9	2			-7,6	2	-1,4	7	-1,4	7
2	CO	5,5	32	-0,3	17	3,2	30			-2,0	24	-8,7	8	-2,0	24
3-4	-	-1,0	27	-3,7	19	-3,4	25	5,4	15	-4,9	22	-6,9	7	-4,9	22
3-4-NH_2	-	-2,2	15	-5,4	22	1,2	27	0,1	14	7,4	44			7,4	44
3-4	CO	-0,6	45	-4,5	26	6,5	41	1,0	26	-3,6	50			-3,6	50
3-4	CO, H_2O	-1,3	9			1,3	36	26,3	5	-1,4	7	-7,6	2	-1,4	7
3	CO, H_2O	7,3	5	3,7	13	15,6	6	-15,4	7	-1,7	6	-1,7	6		
4-5-6-7-1	NH_3	-1,3	14									-5,3	17	-5,3	17
4	CO	5,5	32	7,2	5			-2,2	2	-8,7	8	-2,0	24	-2,0	24
5-6-7-1-2	NH_3	-1,3	14	1,7	12	9,2	14	0,9	9	-5,3	17			-5,3	17
5-6-7-1	NH_3	-0,4	20	-3,0	15			1,5	10	-2,7	24	-2,7	24	-2,7	24
5-6-7-1	H_2O, NH_3	1,7	2	-8,8	2			-4,6	2	-3,0	3	-3,0	3	-3,0	3
5-6-7	NH_3	-1,4	29	-2,8	26	-2,2	31	-3,4	17	-4,4	30	-4,4	30	-4,4	30
5	-	-4,5	4	-2,6	4	0,6	5	0,6	5	-10,8	10	-10,8	10	-10,8	10
5	NH_3	-0,7	5	-1,7	5	5,7	6	-20,5	3	-6,9	6	-6,9	6	-6,9	6
6-7-1-2-3-4	-	-0,2	13	-0,1	10	1,6	11	-13,6	11	-2,6	15	-2,6	15	-2,6	15
6-7-1-2-3-4	H_2O	-3,9	5	-8,3	5	9,0	6	-3,1	4	-2,1	6	-2,1	6	-2,1	6
6-7-1-2-3-4-NH_2	-	-2,2	2	-3,2	3	-7,0	3	5,9	2	-1,7	3	-1,7	3	-1,7	3
6-7-1-2-3-4	CO	-3,1	2			4,8	3			-6,9	2	-6,9	2	-6,9	2
6-7-1-2-3-4	CO, H_2O	3,0	4	0,6	4	-3,4	2	-2,9	4	-2,7	4	-2,7	4	-2,7	4
6-7-1-2-3	H_2O	-2,3	9			1,4	5	10,4	6	-5,6	9				
6-7-1-2	-	-1,9	60	-3,7	41	1,6	26	-4,4	23	-4,8	56			-4,8	56
6-7-1-2	CO, H_2O	-6,8	4	-5,1	4	-0,7	4	5,7	2	-8,9	4			-8,9	4
6-7	-	1,0	74	-0,4	66	1,5	81	-3,7	59	-2,5	82	-2,5	82	-2,5	82
6-7	H_2O	-0,6	26	-0,6	21	-2,7	33	5,5	17	-3,4	27	-3,4	27	-3,4	27
6-7	CO	6,1	1	2,3	24			16,9	2	11,8	2	11,8	2	11,8	2

Product ion assignment: Aa^X-Aa^X... (X = 1 - 7)	Neutral losses	MC-LL Δ ppm	MC-LL I (%)	MC-LM Δ ppm	MC-LM I (%)	MC-LA Δ ppm	MC-LA I (%)	MC-Laba Δ ppm	MC-Laba I (%)	MC-LV Δ ppm	MC-LV I (%)	MC-VL Δ ppm	MC-VL I (%)	[Asp^3]MC-LL Δ ppm	[Asp^3]MC-LL I (%)
6-7	CO, H_2O	1,2	14	6,0	16	6,6	15	-15,0	16	-0,3	16	-0,3	16	-0,3	16
6	CO, H_2O	7,3	5	3,7	13	15,6	6	-15,4	7	-1,7	6	-1,7	6	-1,7	6
7-1-2-3-4	-	-2,2	61	-2,0	49	1,1	51	-2,3	32	-2,9	62	-2,9	62	-2,9	62
7-1-2-3-4-NH_2	-	-1,3	15	2,9	19	-4,6	21	6,5	8	-0,5	18	-0,5	18	-0,5	18
7-1-2-3-4	CO	-4,8	11	-5,3	4			-2,7	5	-4,7	10	-4,7	10	-4,7	10
7-1-2-3-4	CO, H_2O	-5,9	4	-4,3	2	-3,8	2			-6,8	5	-6,8	5	-6,8	5
7-1-2-3	-	-1,9	60	-3,7	41	1,6	26	-4,4	23	-4,8	56				
7-1-2-3	CO, H_2O	-6,8	4	-5,1	4	-0,7	4	5,7	2	-8,9	4				
7-1-2	-	-3,6	17	0,5	14	1,2	26	-9,9	14	-5,7	19			-5,7	19
7-1	-	4,5	27	3,2	23	1,3	36	16,1	14	-2,3	28	-2,3	28	-2,3	28
7-1	CO	2,4	23	0,9	19	0,1	33	0,9	16	-1,1	29	-1,1	29	-1,1	29
7-1	NH_3	3,9	5	9,7	2	2,5	5	3,9	4	-3,7	6	-3,7	6	-3,7	6
7	-	7,3	5	3,7	13	15,6	6	-15,4	7	-1,7	6	-1,7	6	-1,7	6
1-2-3-4-5(?$C_9H_{10}O$)-6	H_2O	-0,1	5	-3,5	6	4,1	8	-8,9	4	-3,3	6	-3,3	6	-3,3	6
3-4-5(?$C_9H_{10}O$)-6-7-1	-	-7,3	3			-7,7	2					-0,7	2		
3-4-5(?$C_9H_{10}O$)-6-7-1	H_2O	-4,4	4			-1,8	1					-4,9	4		
3-4-5(?$C_9H_{10}O$)-6-7-1	NH_3	-1,5	7									-4,0	6		
3-4-5(?$C_9H_{10}O$)-6-7-1	H_2O, NH_3	-1,9	3									-4,6	2		
4-5(?$C_9H_{10}O$)-6-7-1	NH_3	-2,4	19									-4,4	21	-4,4	21
4-5(?$C_9H_{10}O$)-6-7-1	H_2O, NH_3	-4,4	3									-5,6	3	-5,6	3
5(?$C_9H_{10}O$)-6-7-1-2-3-4	-	-2,5	77	-5,8	61	-4,6	80	-2,4	44	-5,3	78	-5,3	78	-5,3	78
5(?$C_9H_{10}O$)-6-7-1-2-3-4	H_2O	-2,7	28	-4,5	27	-3,1	29	-3,1	17	-1,2	29	-1,2	29	-1,2	29
5(?$C_9H_{10}O$)-6-7-1-2-3-4	CO	-7,1	2			-27,9	1	8,2	2	-4,7	2	-4,7	2	-4,7	2
5(?$C_9H_{10}O$)-6-7-1-2-3-4	CO, H_2O	-6,4	4	0,7	4	-5,1	4			-1,0	4	-1,0	4	-1,0	4
5(?$C_9H_{10}O$)-6-7-1-2-3-4	NH_3	-0,7	43	4,5	32	5,2	41	2,5	20	-0,2	43	-0,2	43	-0,2	43
5(?$C_9H_{10}O$)-6-7-1-2-3-4	H_2O, NH_3	-2,3	8	5,6	5	-3,1	6	-6,9	7	-4,4	9	-4,4	9	-4,4	9
5(?$C_9H_{10}O$)-6-7-1-2-3	-	-7,3	3			-14,0	1	13,5	2	-0,7	2				
5(?$C_9H_{10}O$)-6-7-1-2-3	H_2O	-4,4	4	0,9	4	-1,9	7	15,0	5	-4,9	4				
5(?$C_9H_{10}O$)-6-7-1-2-3	NH_3	-1,5	7	-2,1	6	2,9	4			-4,0	6				
5(?$C_9H_{10}O$)-6-7-1-2-3	H_2O, NH_3	-1,9	3	5,7	3	-7,4	2			-4,6	2				
5(?$C_9H_{10}O$)-6-7-1-2	NH_3	-2,4	19			0,3	20	1,2	12	-4,4	21			-4,4	21
5(?$C_9H_{10}O$)-6-7-1-2	H_2O, NH_3	-4,4	3			2,7	4			-5,6	3			-5,6	3
5(?$C_9H_{10}O$)-6-7-1	-	2,3	4			6,7	4	5,8	3	-3,4	4	-3,4	4	-3,4	4
5(?$C_9H_{10}O$)-6-7-1	NH_3	-2,2	34	-1,0	37	-3,1	39	4,3	21	-3,4	39	-3,4	39	-3,4	39
5(?$C_9H_{10}O$)-6-7	NH_3	-0,7	90	-4,9	84	-0,7	95	-0,4	61	-2,9	100	-2,9	100	-2,9	100
5(?$C_9H_{10}O$)	NH_3	0,4	49	-0,3	43	2,2	57	2,2	35	-3,0	56	-3,0	56	-3,0	56
5-6	$C_9H_{10}O$, H_2O, NH_3	-4,6	4	-4,3	3	-12,3	6	-5,4	2	-4,1	5	-4,1	5	-4,1	5
5-6	$C_9H_{12}O$, CO_2, NH_3	-1,9	20	1,4	20	-1,0	24	-1,9	16	-9,0	44	-9,0	44	-9,0	44
5	CH_3OH	-2,6	15	-2,6	14	2,3	22	-6,9	22	-6,3	19	-6,3	19	-6,3	19
5	CH_3OH, NH_3	-2,6	15	1,5	13	0,0	19	-10,2	14	-6,6	20	-6,6	20	-6,6	20
C_3H_4O	Adda	-1,7	24	2,6	23	-2,5	30	4,1	12	-5,8	29	-5,8	29	-5,8	29

Supplementary Table S2. Exact ion masses (Calc, *m/z*), accuracies (Δ, ppm) and intensities (I, %) of UPLC-ESI-QTOF product ion spectra of protonated MCs identified in *Fischerella* sp. CENA161. Blue bars show the relative intensities of the product ions of MC isoforms.

No	Product ion assignment Aa^X-Aa^X... (X = 1 - 7)	Neutral losses	MC-LL Calc (*m/z*)	MC-LL Δ ppm	MC-LL I (%)	MC-LM Calc (*m/z*)	MC-LM Δ ppm	MC-LM I (%)	MC-LA Calc (*m/z*)	MC-LA Δ ppm	MC-LA I (%)	MC-Laba Calc (*m/z*)	MC-Laba Δ ppm	MC-Laba I (%)	MC-LV Calc (*m/z*)	MC-LV Δ ppm	MC-LV I (%)
1	1-2-3-4-5-6-7	H_2O	934,5284	-3,2	24	952,4849	-0,7	24	892,4815	-3,2	24	906,4971	-0,3	21	920,5128	-2,8	26
2	1-2-3-4-5-6-7	NH_3	935,5126	0,9	25	953,4690	5,5	22	893,4656	-0,2	24	907,4813	0,0	22	921,4969	0,2	25
3	1-2-3-4-5-6-7	CH_3OH	920,5128	-2,0	10	938,4692	###	10	878,4658	-0,3	13	892,4815	###	8	906,4971	-9,2	11
4	1-2-3-4	-	427,2551	-1,9	19	445,2115	-0,1	17	385,2082	1,1	16	399,2238	-1,0	12	413,2395	-2,2	21
5	1-2-3-4-NH_2	-	444,2817	-1,7	31	462,2381	-3,4	37	402,2347	0,7	35	416,2504	-4,3	20	430,2660	-4,6	32
6	1-2-3-4	CO	399,2602	-2,7	15	417,2166			357,2132	###	5	371,2289	-5,1	5	385,2445	-4,1	11
7	1-2-3	CO, H_2O	268,1656	-3,6	17	268,1656	0,5	14	268,1656	1,2	26	268,1656	-9,9	14	268,1656	-5,7	19
8	1-2	CO	157,1335	1,0	4	157,1335			157,1335	6,4	5	157,1335	2,9	3	157,1335	-2,5	6
9	2-3-4	-	356,2180	0,6	10	374,1744	2,6	9	314,1710	4,6	13	328,1867	###	7	342,2023	-2,0	12
10	2-3-4-NH_2	-	373,2445	-0,9	12	391,2010	-5,5	15	331,1976	1,8	21	345,2132	-8,0	8	359,2289	-5,4	13
11	2-3-4	CO	328,2231	-3,3	5	346,1795			286,1761	7,6	3	300,1918	-1,3	2	314,2074	2,0	10
12	2-3-4	CO, H_2O	310,2125	-2,3	8	328,1689	2,3	2	268,1656	1,2	26	282,1812	7,4	22	296,1969		
13	2-3	-	243,1339	-1,0	27	243,1339	0,5	8	243,1339	-2,6	5	243,1339	13,8	4	243,1339	-6,9	7
14	2-3-NH_2	-	260,1605	-2,2	15	260,1605			260,1605			260,1605			260,1605		
15	2-3	CO	215,1390	-0,6	45	215,1390			215,1390	5,5	1	215,1390			215,1390		
16	2-3	CO, H_2O	197,1285	-1,3	9	197,1285			197,1285	12,9	2	197,1285			197,1285	-7,6	2
17	2	CO	86,0964	5,5	32	86,0964	-0,3	17	86,0964	3,2	30	86,0964			86,0964	-2,0	24
18	3-4	-	243,1339	-1,0	27	261,0904	-3,7	19	201,0870	-3,4	25	215,1026	5,4	15	229,1183	-4,9	22
19	3-4-NH_2	-	260,1605	-2,2	15	278,1169	-5,4	22	218,1135	1,2	27	232,1292	0,1	14	246,1448	7,4	44
20	3-4	CO	215,1390	-0,6	45	233,0954	-4,5	26	173,0921	6,5	41	187,1077	1,0	26	201,1234	-3,6	50
21	3-4	CO, H_2O	197,1285	-1,3	9	215,0849			155,0815	1,3	36	169,0972	26,3	5	183,1128	-1,4	7
22	3	CO, H_2O	84,0444	7,3	5	84,0444	3,7	13	84,0444	15,6	6	84,0444	###	7	84,0444	-1,7	6
23	4-5-6-7-1	NH_3	693,3858	-1,3	14	711,3422			651,3388			665,3545			679,3701		
24	4	CO	86,0964	5,5	32	104,0528	7,2	5	44,0495			58,0651	-2,2	2	72,0808	-8,7	8
25	5-6-7-1-2	NH_3	693,3858	-1,3	14	693,3858	1,7	12	693,3858	9,2	14	693,3858	0,9	9	693,3858	-5,3	17
26	5-6-7-1	NH_3	580,3017	-0,4	20	580,3017	-3,0	15	580,3017			580,3017	1,5	10	580,3017	-2,7	24
27	5-6-7-1	H_2O, NH_3	562,2912	1,7	2	562,2912	-8,8	2	562,2912			562,2912	-4,6	2	562,2912	-3,0	3
28	5-6-7	NH_3	509,2646	-1,4	29	509,2646	-2,8	26	509,2646	-2,2	31	509,2646	-3,4	17	509,2646	-4,4	30
29	5	-	314,2115	-4,5	4	314,2115	-2,6	4	314,2115	0,6	5	314,2115	0,6	5	314,2115	###	10
30	5	NH_3	297,1849	-0,7	5	297,1849	-1,7	5	297,1849	5,7	6	297,1849	###	3	297,1849	-6,9	6
31	6-7-1-2-3-4	-	639,3348	-0,2	13	657,2912	-0,1	10	597,2879	1,6	11	611,3035	###	11	625,3192	-2,6	15
32	6-7-1-2-3-4	H_2O	621,3243	-3,9	5	639,2807	-8,3	5	579,2773	9,0	6	593,2930	-3,1	4	607,3086	-2,1	6
33	6-7-1-2-3-4-NH_2	-	656,3614	-2,2	2	674,3178	-3,2	3	614,3144	-7,0	3	628,3301	5,9	2	642,3457	-1,7	3
34	6-7-1-2-3-4	CO	611,3399	-3,1	2	629,2963			569,2930	4,8	3	583,3086			597,3243	-6,9	2
35	6-7-1-2-3-4	CO, H_2O	593,3293	3,0	4	611,2858	0,6	4	551,2824	-3,4	2	565,2980	-2,9	4	579,3137	-2,7	4
36	6-7-1-2-3	H_2O	508,2402	-2,3	9	508,2402			508,2402	1,4	5	508,2402	10,4	6	508,2402	-5,6	9
37	6-7-1-2	-	397,2082	-1,9	60	397,2082	-3,7	41	397,2082	1,6	26	397,2082	-4,4	23	397,2082	-4,8	56
38	6-7-1-2	CO, H_2O	351,2027	-6,8	4	351,2027	-5,1	4	351,2027	-0,7	4	351,2027	5,7	2	351,2027	-8,9	4
39	6-7	-	213,0870	1,0	74	213,0870	-0,4	66	213,0870	1,5	81	213,0870	-3,7	59	213,0870	-2,5	82
40	6-7	H_2O	195,0764	-0,6	26	195,0764	-0,6	21	195,0764	-2,7	33	195,0764	5,5	17	195,0764	-3,4	27
41	6-7	CO	185,0921	6,1	1	185,0921	2,3	24	185,0921			185,0921	16,9	2	185,0921	11,8	2
42	6-7	CO, H_2O	167,0815	1,2	14	167,0815	6,0	16	167,0815	6,6	15	167,0815	###	16	167,0815	-0,3	16
43	6	CO, H_2O	84,0444	7,3	5	84,0444	3,7	13	84,0444	15,6	6	84,0444	###	7	84,0444	-1,7	6
44	7-1-2-3-4	-	510,2922	-2,2	61	528,2486	-2,0	49	468,2453	1,1	51	482,2609	-2,3	32	496,2766	-2,9	62
45	7-1-2-3-4-NH_2	-	527,3188	-1,3	15	545,2752	2,9	19	485,2718	-4,6	21	499,2875	6,5	8	513,3031	-0,5	18
46	7-1-2-3-4	CO	482,2973	-4,8	11	500,2537	-5,3	4	440,2504			454,2660	-2,7	5	468,2817	-4,7	10
47	7-1-2-3-4	CO, H_2O	464,2867	-5,9	4	482,2432	-4,3	2	422,2398	-3,8	2	436,2554			450,2711	-6,8	5
48	7-1-2-3	-	397,2082	-1,9	60	397,2082	-3,7	41	397,2082	1,6	26	397,2082	-4,4	23	397,2082	-4,8	56
49	7-1-2-3	CO, H_2O	351,2027	-6,8	4	351,2027	-5,1	4	351,2027	-0,7	4	351,2027	5,7	2	351,2027	-8,9	4
50	7-1-2	-	268,1656	-3,6	17	268,1656	0,5	14	268,1656	1,2	26	268,1656	-9,9	14	268,1656	-5,7	19
51	7-1	-	155,0815	4,5	27	155,0815	3,2	23	155,0815	1,3	36	155,0815	16,1	14	155,0815	-2,3	28
52	7-1	CO	127,0866	2,4	23	127,0866	0,9	19	127,0866	0,1	33	127,0866	0,9	16	127,0866	-1,1	29
53	7-1	NH_3	138,0550	3,9	5	138,0550	9,7	2	138,0550	2,5	5	138,0550	3,9	4	138,0550	-3,7	6
54	7	-	84,0444	7,3	5	84,0444	3,7	13	84,0444	15,6	6	84,0444	###	7	84,0444	-1,7	6
55	1-2-3-4-5(?$C_9H_{10}O$)-6	H_2O	717,4182	-0,1	5	735,3746	-3,5	6	675,3712	4,1	8	689,3869	-8,9	4	703,4025	-3,3	6
56	3-4-5(?$C_9H_{10}O$)-6-7-1	-	705,3818	-7,3	3	723,3382			663,3348	-7,7	2	677,3505			691,3661		
57	3-4-5(?$C_9H_{10}O$)-6-7-1	H_2O	687,3712	-4,4	4	705,3276			645,3243	-1,8	1	659,3399			673,3556		
58	3-4-5(?$C_9H_{10}O$)-6-7-1	NH_3	688,3552	-1,5	7	706,3116			646,3083			660,3239			674,3396		
59	3-4-5(?$C_9H_{10}O$)-6-7-1	H_2O, NH_3	670,3447	-1,9	3	688,3011			628,2977			642,3134			656,3290		
60	4-5(?$C_9H_{10}O$)-6-7-1	NH_3	559,3126	-2,4	19	577,2690			517,2657			531,2813			545,2970		
61	4-5(?$C_9H_{10}O$)-6-7-1	H_2O, NH_3	541,3021	-4,4	3	559,2585			499,2551			513,2708			527,2864		
62	5(?$C_9H_{10}O$)-6-7-1-2-3-4	-	818,4658	-2,5	77	836,4223	-5,8	61	776,4189	-4,6	80	790,4345	-2,4	44	804,4502	-5,3	78
63	5(?$C_9H_{10}O$)-6-7-1-2-3-4	H_2O	800,4553	-2,7	28	818,4117	-4,5	27	758,4083	-3,1	29	772,4240	-3,1	17	786,4396	-1,2	29
64	5(?$C_9H_{10}O$)-6-7-1-2-3-4	CO	790,4709	-7,1	2	808,4273			748,4240	###	1	762,4396	8,2	2	776,4553	-4,7	2
65	5(?$C_9H_{10}O$)-6-7-1-2-3-4	CO, H_2O	772,4604	-6,4	4	790,4168	0,7	4	730,4134	-5,1	4	744,4291			758,4447	-1,0	4
66	5(?$C_9H_{10}O$)-6-7-1-2-3-4	NH_3	801,4393	-0,7	43	819,3957	4,5	32	759,3923	5,2	41	773,4080	2,5	20	787,4236	-0,2	43
67	5(?$C_9H_{10}O$)-6-7-1-2-3-4	H_2O, NH_3	783,4287	-2,3	8	801,3851	5,6	5	741,3818	-3,1	6	755,3974	-6,9	7	769,4131	-4,4	9
68	5(?$C_9H_{10}O$)-6-7-1-2-3	-	705,3818	-7,3	3	705,3818			705,3818	###	1	705,3818	13,5	2	705,3818	-0,7	2
69	5(?$C_9H_{10}O$)-6-7-1-2-3	H_2O	687,3712	-4,4	4	687,3712	0,9	4	687,3712	-1,9	7	687,3712	15,0	5	687,3712	-4,9	4
70	5(?$C_9H_{10}O$)-6-7-1-2-3	NH_3	688,3552	-1,5	7	688,3552	-2,1	6	688,3552	2,9	4	688,3552			688,3552	-4,0	6
71	5(?$C_9H_{10}O$)-6-7-1-2-3	H_2O, NH_3	670,3447	-1,9	3	670,3447	5,7	3	670,3447	-7,4	2	670,3447			670,3447	-4,6	2
72	5(?$C_9H_{10}O$)-6-7-1-2	NH_3	559,3126	-2,4	19	559,3126			559,3126	0,3	20	559,3126	1,2	12	559,3126	-4,4	21
73	5(?$C_9H_{10}O$)-6-7-1-2	H_2O, NH_3	541,3021	-4,4	3	541,3021			541,3021	2,7	4	541,3021			541,3021	-5,6	3
74	5(?$C_9H_{10}O$)-6-7-1	-	463,2551	2,3	4	463,2551			463,2551	6,7	4	463,2551	5,8	3	463,2551	-3,4	4
75	5(?$C_9H_{10}O$)-6-7-1	NH_3	446,2286	-2,2	34	446,2286	-1,0	37	446,2286	-3,1	39	446,2286	4,3	21	446,2286	-3,4	39
76	5(?$C_9H_{10}O$)-6-7	NH_3	375,1914	-0,7	90	375,1914	-4,9	84	375,1914	-0,7	95	375,1914	-0,4	61	375,1914	-2,9	100
77	5(?$C_9H_{10}O$)	NH_3	163,1117	0,4	49	163,1117	-0,3	43	163,1117	2,2	57	163,1117	2,2	35	163,1117	-3,0	56
78	5-6	$C_9H_{10}O$, H_2O, NH_3	274,1438	-4,6	4	274,1438	-4,3	3	274,1438	###	6	274,1438	-5,4	2	274,1438	-4,1	5
79	5-6	$C_9H_{12}O$, CO_2, NH_3	246,1489	-1,9	20	246,1489	1,4	20	246,1489	-1,0	24	246,1489	-1,9	16	246,1489	-9,0	44
80	5	CH_3OH	282,1852	-2,6	15	282,1852	-2,6	14	282,1852	2,3	22	282,1852	-6,9	22	282,1852	-6,3	19
81	5	CH_3OH, NH_3	265,1587	-2,6	15	265,1587	1,5	13	265,1587	0,0	19	265,1587	###	14	265,1587	-6,6	20
82	C_3H_4O	Adda	258,1852	-1,7	24	258,1852	2,6	23	258,1852	-2,5	30	258,1852	4,1	12	258,1852	-5,8	29