UNIVERSITY OF SÃO PAULO

Faculty of Pharmaceutical Sciences

*Postgraduate Programme in Pharmacy
(Pathophysiology and Toxicology)*

Area of study: Toxicology

Mariana de Almeida Torres

Ecotoxicological studies of metabolites produced by two Brazilian cyanobacterial strains

PhD thesis

São Paulo

2023

UNIVERSITY OF SÃO PAULO

Faculty of Pharmaceutical Sciences

Postgraduate Programme in Pharmacy (Pathophysiology and Toxicology)

Area of study: Toxicology

Mariana de Almeida Torres

Ecotoxicological studies of metabolites produced by two Brazilian cyanobacterial strains

Original Version

PhD thesis presented to the Postgraduate Programme of Pharmacy (Physiopathology and Toxicology) at the Faculty of Pharmaceutical Sciences of the University of São Paulo, to obtain the doctoral degree.

Research area: Toxicology

Advisor: Prof. Dr. Ernani Pinto

São Paulo

2023

Mariana de Almeida Torres

Ecotoxicological studies of metabolites produced by two Brazilian cyanobacterial strains.

Comissão Julgadora
da Tese
para obtenção do Título de Doutora em Ciências – Toxicologia

Prof. Dr. Ernani Pinto Junior
Orientador/presidente

_____

1º examinador(a)

_____

2º examinador(a)

_____

3º examinador(a)

_____

4º examinador(a)

São Paulo, _____ de _____ de 2023

Aos meus pais e ancestrais, pelo esforço e dedicação,

para que eu chegasse até aqui

AGRADECIMENTOS

ACKNOWLEDGEMENTS

"Nothing in life is to be feared; it is only to be understood."

Marie Curie

RESUMO

TORRES, M A. Estudos ecotoxicológicos de metabólitos produzidos por duas cepas de cianobactérias brasileiras. 2023. 114f. Tese (Doutorado). Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, 2023.

A proliferação de cianobactérias afeta os ecossistemas de água doce em todo o mundo, e uma das principais preocupações está relacionada aos seus metabólitos secundários estruturalmente diversos e bioativos, não limitados às já conhecidas toxinas, tais como as microcistinas (MC). No entanto, os riscos ecotoxicológicos associados a esses metabólitos coproduzidos permanecem, em sua maioria, desconhecidos. Neste estudo, avaliamos os metabólitos de duas cepas brasileiras do gênero *Microcystis*, a produtora de microcistina MIRS-04 e a não produtora NPCD-01, por LC-HRMS/MS. Estudos ecotoxicológicos foram conduzidos com os organismos de água doce *Thamnocephalus platyurus* (microcrustáceo) e *Danio rerio* (peixe-zebra) no estágio embrio-larval. A sobrevivência e o comportamento locomotor (natação) das duas espécies foram monitorados após a exposição a frações semipurificadas por HPLC dos extratos de cianobactérias. Para os embriões de peixe-zebra, também foram analisadas as alterações morfológicas e as funções cardiovasculares das larvas. Os extratos integrais da biomassa cianobactérias (frações combinadas) da cepa produtora de MCs e da cepa não produtora apresentaram toxicidade aguda dose-dependente para ambas as espécies modelo testadas. Para o microcrustáceo, o efeito agudo sobre a sobrevivência também foi observado para a fração de HPLC contendo a cianopeptolina micropeptina K139 (principal composto produzido pela cepa MIRS-04), e as frações apolares revelaram serem os principais contribuintes para a toxicidade aguda do extrato da NPCD-01. Além disso, efeitos subletais (representados como mobilidade reduzida) também foram observados nos microcrustáceos expostos às frações combinadas (extrato integral), a fração contendo majoritariamente micropeptina (MIRS-04) e às frações apolares (NPCD-01). Nos embriões de peixe-zebra, de modo similar, mortalidade foi registrada após 120 horas de exposição às frações combinadas de ambas as cepas, à fração dominada por micropeptina da cepa MIRS-04 e às frações apolares da NPCD-01. Além disso, foram observados edemas da região pericárdica em larvas expostas às frações de HPLC dominadas pela microginina nostoginina BN741 (não produtora) e micropeptina K139 (produtora de MC). Nossos resultados contribuem ainda mais para a crescente evidência de toxicidade de metabólitos de cianobactérias além das já conhecidas cianotoxinas.

Palavras-chave: Toxicidade em peixes. Microcrustráceos. Cianopeptídeos. Microgininas. Cianopeptolinas.

ABSTRACT

Cyanobacterial blooms affect freshwater ecosystems across the globe, and one major concern relates to their structurally diverse and bioactive metabolites not limited to known toxins such as microcystins (MC). Yet, the ecotoxicological risks associated with these co-produced metabolites remain mostly unknown. In this study, we assessed metabolites from two Brazilian *Microcystis* strains, the microcystin-producer MIRS-04 and the non-producer NPCD-01, by LC-HRMS/MS using CyanoMetDB (a comprehensive database of cyanobacterial secondary metabolites) for the suspect screening. Ecotoxicological studies were conducted with the freshwater organisms *Thamnocephalus platyurus* (fairy shrimp) and *Danio rerio* (zebrafish) in embryo-larval stage. The survival and locomotor (swimming) behaviour of the two species were monitored after exposure to HPLC-semi-purified fractions of the cyanobacterial extracts. For the zebrafish embryos, morphological alterations and cardiovascular functions of 5-day-old larvae were also analysed. The whole cyanobacterial extract (pooled HPLC fractions) from the MC-producer and non-producer pointed to acute dose-dependent toxicity in both model species tested. For the fairy shrimp, acute effect on survival was also observed for the HPLC-fraction of the MC-producer containing the cyanopeptolin micropeptin K139, and the apolar fractions revealed to be the main contributors to the acute toxicity of the microcystin-free extract. Additionally, sublethal effects (represented as reduced mobility) was also observed on the microcrustaceans exposed to the pooled, micropeptin-dominated (MIRS-04) and apolar fractions (NPCD-01). In the zebrafish embryos, likewise, mortality was recorded after 120h of exposure to the pooled fractions of both strains, the micropeptin-dominated fraction of the MIRS-04 strain, and apolar fractions of the NPCD-01. In addition, oedemas of the pericardial region were observed in larvae exposed to the HPLC-fractions dominated by the microginin nostoginin BN741 (non-producer) and micropeptin K139 (MC-producer). Our results further add to the growing evidence of toxicity of cyanobacterial metabolites other than microcystins.

Keywords: Fish toxicity. Microcrustaceans. Cyanopeptides. Microginins. Cyanopeptolins.

LIST OF FIGURES

SUMMARY

1 STATE OF THE ART

1.1 CYANOBACTERIA, EUTROPHICATION AND CYANOBACTERIAL BLOOMS

The population growth, the increase of industrial and agricultural activities and the agglomeration in large urban centres exert pressure on the ecosystems in the different forms of environmental stressors. One of the most common consequences is the discharge of effluents (such as fertilisers, industrial and domestic wastewater) into water bodies, which can promote an excessive increase in the concentration of nutrients, especially phosphate and nitrogen compounds, causing a process known as artificial eutrophication (CHORUS; WELKER, 2021; HUISMAN et al., 2018). These episodes of eutrophication prompted by anthropogenic pressure have been of concern and the subject of study by the scientific community and regulatory authorities around the world in different domains.

The increase in nutrient concentrations reflects directly on the increase of the primary productivity of phytoplankton communities (Figure 1). When this proliferation is dominated by one or only a few species, this phenomenon is known as **bloom**, especially when combined with favourable conditions, *e. g.* temperature, geographic location, and individual water body characteristics, generating complex scenarios for management (CHORUS; WELKER, 2021; HO; MICHALAK; PAHLEVAN, 2019; HUISMAN et al., 2018).

Figure 1. Cyanobacterial blooms in Brazil: pond located in the Minas Gerais state (A), and Billings Reservoir located in São Paulo state (B)

Source: courtesy of Rhuana Médice (A) and Melissa Medrano (B)

Dense biomass in aquatic ecosystems increases the turbidity, causing light shading for other species and increased oxygen consumption during biomass decay. Deleterious effects associated with blooms include reduced ecosystem biodiversity as well as impacts on the water usage for human consumption, crop irrigation, aquaculture, industrial processing, recreation, and tourism (MERILUOTO; SPOOF; CODD, 2018).

Depending on the climate and environmental conditions, they can be seasonal (*i.e.*, spring and summer, temperate climate), which is the case for Swiss peri-alpine lakes, for example (BAUMANN; JU, 2008; JONES; JANSSEN, 2022; MONCHAMP et al., 2018); or permanent, namely in tropical regions. In Brazil, cyanobacterial blooms in lakes and reservoirs are often a long-term issue and are closely linked with low water quality, reflecting in severe health, environmental, social, and economic impacts (MAY et al., 2021) – thus, the relevance of studying and further developing strategies to prevent cyanobacterial blooms, especially in tropical developing countries where these episodes are a long-lasting environmental and public health issue.

Regarding the biomass (phytoplankton) composition, the most commonly present in bloom events in eutrophic continental ecosystems are cyanobacteria, which stand out for their wide incidence (FALCONER, 1998).

1.2 CYANOBACTERIA

Cyanobacteria (Kingdom *Bacteria*, Phylum *Cyanobacteria*) belong to an ancient group of organisms whose origin is estimated to be approximately 3 billion years (WALTER et al., 2017). They are believed to be among the pioneer organisms in the early atmosphere, being the first primary producers to release elemental oxygen and to colonise never-inhabited rocks and soils (CHORUS; WELKER, 2021; HUISMAN et al., 2018).

Cyanobacteria are prokaryotic, single-celled, photoautotrophic microorganisms (CHORUS; WELKER, 2021), and their closest relatives are bacteria (CALIJURI; SANTOS; ALVES, 2006; CHORUS; WELKER, 2021; HUISMAN et al., 2018), albeit

differed by a thicker peptidoglycan layer, providing rigidity to the wall, protecting from osmotic cell lysis (BERGEY; HOLT, 1994)(HOICZYK; HANSEL, 2000).

Cyanobacteria often form part of pioneer communities and compose a group of cosmopolitan species with high adaptive skills, making them able to inhabit the most diverse environments, even the most hostile ones (*e.g.*, high salinity, extreme temperatures, and deep waters) (CHORUS; WELKER, 2021; GENUÁRIO et al., 2019).

Such characteristics have also given this group diverse morphological forms (Figure 2) that can be categorised into coccoids, with unicellular or colonial forms; and filamentous forms, the latter of which may or may not have the presence of specialised cells called heterocytes, capable of fixing atmospheric nitrogen (LOPES; SILVA; VASCONCELOS, 2022).

Figure 2. Photomicrographs of different cyanobacterial species
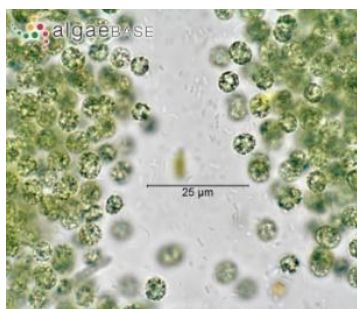
| *Microcystis aeruginosa* | *Anabaena flosaquae* | *Planktothrix isothrix* |
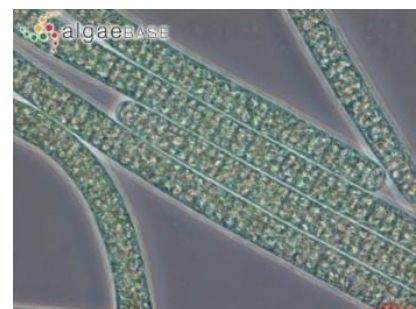
Source: Algae Base (http://www.algaebase.org/– accessed in January 2020)

Such high speciation capacity has resulted in a high genotypic and phenotypic diversity of cyanobacteria, so that this group is represented by more than 250 genera and more than 2700 identified species in the last survey conducted by Nabout and collaborators in 2013 (NABOUT et al., 2013). According to the authors of this study, this number continues to increase at the average rate of 15 new species per year. Regarding species richness, in Brazil, for instance, according to the study conducted by Sant'anna and collaborators in 2011, up until the end of the study, there had been recorded occurrences of 460 species of cyanobacteria in Brazilian territory (SANT'ANNA et al., 2011).

The cyanobacterial genera most commonly found in Brazilian ecosystems are *Microcystis*, *Raphidiopsis* and *Dolichospermum*, with the species *M. aeruginosa* and *R. raciborskii* (formerly *Cylindrospermopsis*) being the most widespread (SANT'ANNA et al., 2008).

In Brazil and worldwide, cyanobacterial blooms have increased in frequency and with greater intensity (HUISMAN et al., 2018), raising the question of whether climate change may favour the dominance of cyanobacteria through warming waters, greater vertical stratification of the water column, changes in nutrient and light availability (HUISMAN et al., 2018; PAERL; HUISMAN, 2008). However, the impact of climate change on phytoplankton growth and dominance is not straightforward, as climatic variabilities affect ecosystems directly but also indirectly, interfering in the seasonal appearance of zooplankton, changing grazer pressure on phytoplankton, CO_2 solubility, temperature stratification, rain regime, etc. Chapter 4 of the World Health Organization guideline for *Toxic Cyanobacteria in Water* and references therein explore the different aspects and outcomes that might be linked to a warmer climate and cyanobacterial biomass density (CHORUS; WELKER, 2021).

1.3 CYANOBACTERIAL SECONDARY METABOLITES

The main concern regarding cyanobacterial blooms is related to their ability to synthesise a wide range of bioactive secondary metabolites. Some of these compounds are known to promote toxic effects to aquatic biota and humans (cyanotoxins) (CARMICHAEL, 1992), and their mechanisms of action have already been elucidated. Improved targeted and non-targeted bioanalytical tools accompanied by increased availability of calibration standards have allowed the identification, in addition to cyanotoxins (extensively studied and structurally well known), of other bioactive cyanobacterial metabolites, which have revealed great biotechnological potential (biocides, biofertilisers, pharmaceutical applications, and photoprotectants, among many others). However, their role ecological role, bioactivity and toxicity are still not sufficiently explored (JANSSEN, 2019; WELKER; VON DÖHREN, 2006).

1.3.1 Cyanotoxins

Cyanotoxins are a diverse group of natural toxins, from the chemical and toxicological point of view. They can be classified according to their chemical structures, being divided into peptides, alkaloids and lipopolysaccharides, the first being the group with the highest share of compounds identified so far (so-called cyanopeptides) (JONES et al., 2021). Cyanotoxins can also be classified pharmacologically as neurotoxins (anatoxin-a, guanitoxin, and saxitoxins); hepatotoxins (microcystins, nodularins, and cylindrospermopsins, the latter also being classified as a cytotoxin); and dermatotoxins (lipolysaccharides) (CHORUS; WELKER, 2021; MEREL et al., 2013). Figure 3 presents the chemical structure of representatives of different classes of cyanotoxins.

Figure 3. Chemical structure of representative compounds from 5 classes of cyanotoxins (microcystins, cylindrospermopsins, anatoxins, saxitoxins and nodularins.
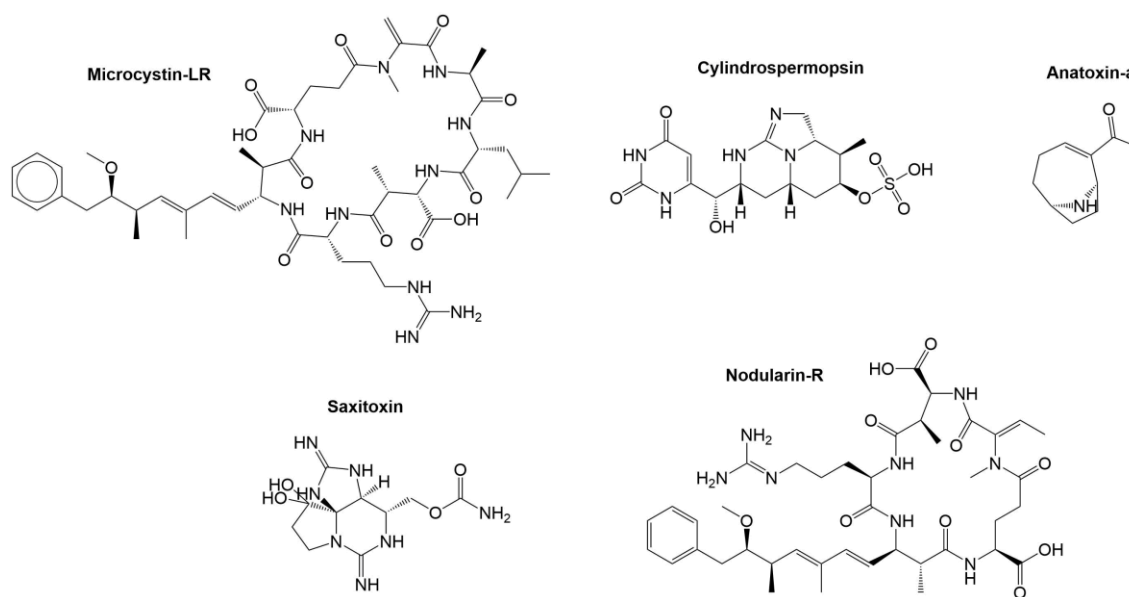
Source: Jones et al., 2021

The most common episodes involving cyanotoxin poisoning are caused by microcystins (hepatotoxins), which can cause death as a result of hypovolemic shock or intrahepatic haemorrhage. Microcystins compose a large group of different variants (formed by conserved structure and variable amino acid residues), with over 300 identified so far. The most striking and globally reported cyanotoxin poisoning episode occurred in Caruaru (Pernambuco, Brazil) in 1996, when 130 patients

receiving hemodialysis in a clinic developed acute liver failure, and 52 deaths were reported after intoxication. Subsequent studies showed the presence of microcystins and cylindrospermopsin in the activated carbon of the clinic's treatment system, as well as in blood and liver samples from patients (AZEVEDO et al., 2002).

This tragedy lead, later on, the World Health Organisation (WHO) to establish guideline values for the presence of microcystins (AZEVEDO et al., 2002; CHORUS; BARTRAM, 1999). In 2021, WHO updated the guidelines for water quality, including chronic and acute drinking water thresholds and recreational thresholds for microcystin, cylindrospermopsin, anatoxin-a and saxitoxin (CHORUS; WELKER, 2021).

These toxins enter the hepatocytes through the bile acid transporters, causing inhibition of the enzymes phosphatases 1 and 2A, with consequent hyperphosphorylation of proteins and disorganisation of the cytoskeleton. This disorganisation leads to hepatocyte retraction, which, in turn, leads to capillary retraction and increased intercellular spaces, causing tissue injury and liver haemorrhage (CHORUS; WELKER, 2021; CALIJURI; SANTOS; ALVES, 2006).

Microcystin-producing strains have been found in all orders (*Chroococcales*, *Oscillatoriales*, *Nostocales*, *Stigonematales Pleurocapsales*) (CHORUS; WELKER, 2021). However, there is no pattern with respect to orders and genera. The genera best known for producing microcystins are *Microcystis*, *Planktothrix*, *Dolichospermum*, and *Nostoc*; less frequently are the genera *Anabaenopsis*, *Arthrospira*, *Fischerella*, *Pseudanabaena*, *Phormidium*, *Synechococcus*, and *Radiocystis* (CHORUS; WELKER, 2021).

In Brazil, the Ministry of Health established the maximum value of 1 µg.L^{-1} for total microcystins in drinking water, taking as a basis the value recommended by the World Health Organization (CHORUS; WELKER, 2021; "PORTARIA DE CONSOLIDAÇÃO Nº 5, DE 28 DE SETEMBRO DE 2017 Consolidação", 2017).

1.3.2 Cyanobacterial metabolites beyond microcystins

Cyanobacteria represent one of the most promising groups of microorganisms for the isolation of new biochemically active natural compounds. A large proportion of

cyanobacterial secondary metabolites are oligopeptides (or have peptide substructures), synthesized by the NRPS (non-ribosomal peptide synthetase) or the combination NRPS/PKS (polyketide synthase), or by ribosomal synthesis with post-translational modifications (HUANG; ZIMBA, 2019). This variety results in a wide structural range of these metabolites, so that most can be divided into families based on the core fixed structures that characteride each class, including the well-known cyanotoxins (discussed in the previous section). Some examples of cyanopeptides other than the cyanotoxins include anabaenopeptins, cyanopeptolins, microviridins, aeruginosins, microginins, aerucyclamides, namalides, and spumigins (JONES et al., 2021; SANZ; SALINAS; PINTO, 2017; WELKER; VON DÖHREN, 2006) Board 1 presents the main classes of cyanopeptides known so far, their main structural characteristics and reported bioactivity.

Scientific interest in these compounds has grown in the last two decades due to their bioactivity, pointing to their potential biotechnological and pharmaceutical use (LOPES; SILVA; VASCONCELOS, 2022; RAJA et al., 2016). Biochemical activities already identified in some compounds include antiviral, anticancer, antibiotic, antitumor, antifungal, photoprotective, immunosuppressive, enzyme inhibitor, cytotoxic, anti-inflammatory activity, among others (BURJA et al., 2001; RAJA et al., 2016).

Although these bioactive compounds have drawn attention because of their commercial potential, recent evidence has raised the question of whether their presence in water bodies poses a risk to human health and ecosystems. Even though bloom episodes are common and recurrent worldwide, still the risk assessment is limited to cyanotoxins (notably microcystins), and the potential risks associated to other secondary metabolites produced by cyanobacteria remains limited (JANSSEN, 2019). This topic will be further discussed in the following session.
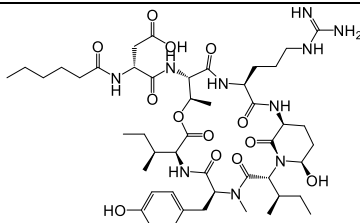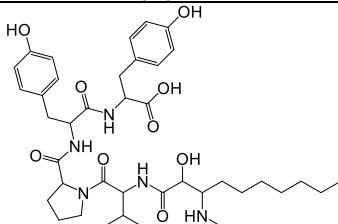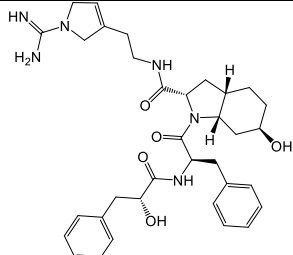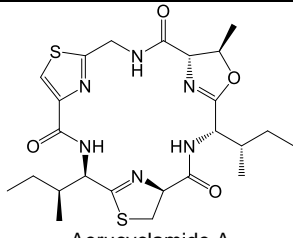
A number of cyanopeptides were already detected in different aquatic compartments, such as rivers, reservoirs, and lakes. In a review, Janssen (2019) compiled reports of detection of these compounds in different countries (Italy, Greece, Poland, Portugal, Israel, Switzerland, and Finland). In Brazil, several strains producing secondary metabolites not belonging to the traditional cyanotoxin classes

have already been identified in different geographic locations (JACINAVICIUS et al., 2021; SANZ et al., 2015; SILVA-STENICO et al., 2015)
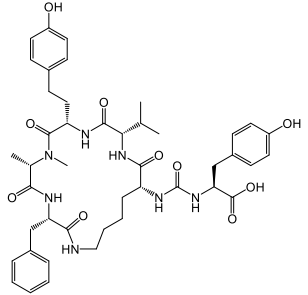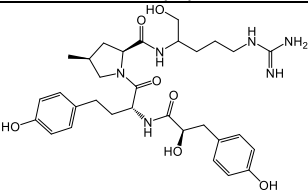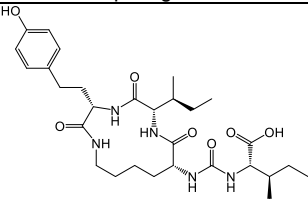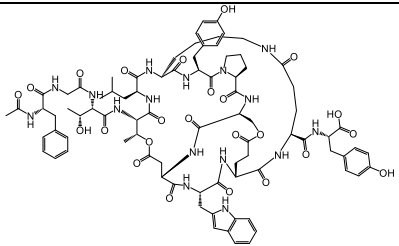
Unfortunately, the limitations regarding the quantification (such as demand for sophisticated analytical techniques, scarce and expensive analytical standards) make it difficult to monitor and report quantitative information of peptides beyond the traditional toxins; therefore, spatiotemporal quantification studies are limited.

Beversdorf et al. (2018) reported the detection of cyanopeptolins, anabaenopeptins, and microginins at micromolar concentration range (μg/L) in raw water of drinking water treatment plants in Wisconsin (USA) (BEVERSDORF et al., 2018). Filatova et al. (2021) analysed the cyanopeptide profile of different reservoirs in the United Kingdom, where it was also possible to detect cyanopeptides at a low micromolar range. In one specific case (Ingbirchworth reservoir), the concentrations of anabaenopeptin A and B, and anabaenopeptilide 202A (a cyanopeptolin variant) were comparable to those found for five different microcystins variants.

Board 1. Main classes of cyanopeptides and reported bioactivity.

Class	Main structural features	Bioactivity	Chemical structure of a class representative	Ref
Cyanopeptolins	Hexapeptides with the characteristic Ahp moiety (3-amino-6-methoxy-2-piperidone). The cyclization of the peptide ring by an ester bond of the β-hydroxy group of Thr with the carboxyl group of the terminal amino acid	Mainly protease inhibition (chymotrypsin, elastase, human kallikrein, matriptase, plasmin, thrombin, trypsin)	 Micropeptin K139	(1)
Microginins	Linear peptides with four to six amino acids and the characteristic Ahda portion (3-amino-2-hydroxy-decanoic acid) in the N-terminal side	Inhibitors metalloproteases, such as angiotensin-converting enzyme and leucine aminopeptidase	 Microginin FR6	(2) (3) (4)
Aeruginosins	Linear tetrapeptides with the characteristic portions Choi (2-carboxy-6-hydroxyoctahydroindol) and Hpla (p-hydroxyphenyl lactic acid)	Cathepsin B, plasmin, thrombin and trypsin inhibitors	 Aeruginosin A	(2) (3)
Cyclamide	Cyclic hexapeptides characterised by three azole rings	Activity against leukemia cells and human parasites	 Aerucyclamide A	(2) (3)

Board 1. Main classes of cyanopeptides and reported bioactivity (continued)

Class	Main structural features	Bioactivity	Chemical structure of a class representative	Ref
Anabaenopeptins	Cyclic peptides with a ureido bond connecting the primary amine of the R' group of Lys with the carboxyl group of the neighbouring amino acid to form an amide bond	Inhibitors of cathepsin G, chymotrypsin, carboxypeptidase A, elastase, phosphatase 1, phosphatase 2A, thrombin	Anabaenopeptin A	(2) (3)
Spumigins	Characterised by the Hpla portion at position 1, HTyr or HPhe at position 2, and an Arg or Lys derivative at the C-terminal portion. Unlike the aeruginosins, position 3 is composed of MePro or Pro	Variants containing methylproline reported to inhibit trypsin	Spumigin A	(5)
Namalides	Cyclic tetrapeptides with three-amino acid ring and side chain amino acid linked via the ureido group to the ᴅ-amino group of a Lys	Carboxypeptase A inhibition	Namalide B	(5)
Microviridins	Multi-cyclic, high molecular weight structures with peptide bonds and ester bonds, and a side chain of varying length. The main peptide ring consists of seven amino acids with an ester bond between the 4-carboxy group of Asp (position 10) and the hydroxy group of Thr (position 4), a peptide bond between the 6-amino group of Lys (position 6) and the 4-carboxy group of Glu (position 7).	Inhibitors of chymotrypsin, elastase, trypsin	Microviridin B	(2) (3)

Sources: (1) (KÖCHER et al., 2020); (2) (HUANG; ZIMBA, 2019); (3) (WELKER; VON DÖHREN, 2006); (4) (PAIVA et al., 2017); (5) (SANZ; SALINAS; PINTO, 2017).

24

Considering that the cyanobacterial strains studied herein produce mainly microginins and cyanopeptolins, the following sessions will discuss these classes in more detail.

1.3.2.1 Microginins

Microginins are linear peptides already detected in strains of *M. aeruginosa*, *Woronichinia naegeliana*, *Planktothrix*, *Oscillatoria* and *Nostoc* (ZERVOU et al., 2020 and references therein). However, most reports focus on the genus *Microcystis*, which is widely found in bloom episodes (CARNEIRO et al., 2012; FERNANDES et al., 2019b; MARTINS et al., 2009; PAIVA et al., 2017; RESHEF; CARMELI, 2001; STRANGMAN; WRIGHT, 2016).

This class of linear peptides has variants with 4 to 6 amino acid residues reported so far, with a predominance of two tyrosine units at the C-terminus, and is characterised by a decanoic acid derivative, 3-amino-2-hydroxy-decanoic acid (Ahda) in the N-terminal portion (ISHIDA et al., 2000; ZERVOU et al., 2020). Mono and dichlorinated variants in the Ahda portion have already been identified, as well as derivatives containing a modified octanoic acid (Ahoa) portion and derivatives containing N-methyl sulfoxide of methionine (see board 1) (STRANGMAN; WRIGHT, 2016; WELKER; VON DÖHREN, 2006). The identified congeners have different amino acids between the N-terminal β-amino acid and the C-terminal tyrosine residues, including isoleucine, leucine, valine, alanine, tryptophan, homotyrosine, proline, and serine, some of which are methylated (STEWART et al., 2018). To date, 87 variants have been identified with the structural information reported (JONES et al., 2021).

Many microginins reported so far exhibit inhibitory activity of zinc metalloproteases, such as angiotensin-converting enzyme (ACE) and leucine aminopeptidase (LAP). Consequently, such compounds are of pharmaceutical interest with regard to the discovery of new antihypertensive agents (ISHIDA et al., 2000).

Even though many microginin variants can have bioactive properties, data regarding their environmental concentration, fate and impacts are rare. Beversdorf et al. (2017) monitored the concentration of microginin 690 (among other cyanopeptides) for six months, in 6 lakes located in central and south central Wisconsin (USA), and found concentrations up to 2.21 µg/L. Filatova et al. (2021) reported concentrations in the low µg/L range for microginin KR604 in samples from Ingbirchworth reservoir (UK). In addition, Zervou et al. (2021) observed concentrations of microginin T1 of 47.0 µg/L in the Greek lake Vegoritis, which was the most abundant cyanopeptide.

No studies were found in which the concentration of microginins was monitored in Brazilian waters, despite the fact that many Brazilian cyanobacterial strains were reported to produce numerous microginin variants (CARNEIRO et al., 2012; KINOSHITA et al., 2020; SILVA-STENICO et al., 2011).

1.3.2.2 Cyanopeptolins

Cyanopeptolins comprise a broad class of cyclic peptides and are generically composed of 6 amino acids, among them the characteristic 3-amino-6-hydroxy-2-piperidone (Ahp) moiety at position 3 (HUANG; ZIMBA, 2019). The cyclisation of the peptide ring is given by an ester bond of the β-hydroxy group of threonine with the carboxyl group of the terminal amino acid (see Board 1) (CORNEL et al., 1993).

Cyanopeptolins have a side chain of varying length, attached to threonine via the amino grouping. Two main types of side chains are common, one consisting of one or two amino acids and an aliphatic fatty acid (formic acid to octanoic acid), and another with a glyceric acid unit at the N-terminus. The main variations in cyanopeptolins occur in the side chains with some variations being halogenated and sulfated, and all amino acids in the ring can vary except for threonine (position 1) and Ahp (position 3) (HUANG; ZIMBA, 2019; WELKER; VON DÖHREN, 2006).

As reviewed by Huang and Zimba (2019), the bioactivity related to this class of cyanopeptides includes inhibition of chymotrypsin, factor XIa, human kallikrein, plasmin, thrombin, and trypsin (HUANG; ZIMBA, 2019).

The literature reports the detection of these compounds in several genera of cyanobacteria, such as *Anabaena*, *Microcystis*, *Planktothrix*, *Scytonema*, *Symploca*, *Lyngbya*, *Woronichinia*, and *Aphanocapsa* (HUANG; ZIMBA, 2019; WELKER; VON DÖHREN, 2006). In the same monitoring study previously mentioned, Beversdorf et al. (2017) quantified 3 different cyanopeptolin variants with total maximum concentration of 13.93 µg/L (summing up the concentration of all three variants) in Lake Winnebago (Wisconsin, USA). Also, Filatova et al. (2021) observed that the cyanopeptolin variant anabaenopeptilide 202A was the most abundant cyanopeptide among all identified (including WHO-recognised toxins) in the Ingbirchworth reservoir (UK), reaching a concentration of 22 µg/L.

1.3.3 Ecotoxicological risk of cyanobacterial metabolites beyond the WHO-recognised toxins: why should they be of concern?

Despite the knowledge about the bioactivities of the cyanopeptides, the risks associated with most cyanopeptide classes in water bodies are not yet sufficiently explored. In addition, the fact that these are naturally occurring bioactive metabolites co-produced with known toxins like microcystins should be taken into account in terms of their potential adverse effect on aquatic ecosystems and human health.

While only a small share of the compounds has been assessed in terms of their toxicological risk, a recent number of studies revealed the toxicity of (semi)purified cyanobacterial metabolites beyond the well-known toxins for different levels of biological organization and in organisms from different phyla (single cultured cells, unicellular and microeukaryotes, plants, crustaceans, and fish).

For example, a fraction of extracts free of known cyanotoxins from *Nodularia spumigena* contained five spumigin and three aeruginosin variants, and induced high mortality rates (90%) in the shrimp *Artemia franciscana* (MAZUR-MARZEC et al., 2015). Variants of anabaenopeptins and cyanopeptolins were lethal towards the nematode *Caenorhabditis elegans* in the same low micromolar range as the microcystin variant MC-RR (LENZ; MILLER; MA, 2019). Acute toxicity towards the crustacean *Thamnocephalus platyurus* was also observed upon exposure to different aeruginosins (SCHERER; BEZOLD; GADEMANN, 2016), cyanopeptolins (*e.g.,*

cyanopeptolin 1020, (GADEMANN et al., 2010)), microginin FR3 (LC_{50} = 7.78 µg/L) (BOBER; BIALCZYK, 2017) and microviridin 1777 (SIEBER et al., 2020).

In freshwater impacted by cyanobacterial blooms, zooplankton can be affected by multiple reasons, such as deterioration of quality of the algal food; reduction of the filter efficiency of larger zooplankton species by filaments and colonies of cyanobacteria; and harmful effects of cyanobacterial metabolites on different zooplankton genera and species, especially after the collapse of cyanobacterial blooms where the concentration of dissolved metabolites increase considerably (PAWLIK-SKOWROŃSKA; TOPOROWSKA; MAZUR-MARZEC, 2019 and references therein). As previously shown, many cyanopeptides inhibit proteases, which also play a role in the digestive process, thus raising the hypothesis whether those compounds have anti-grazing function (CHORUS; WELKER, 2021; PAWLIK-SKOWROŃSKA; TOPOROWSKA; MAZUR-MARZEC, 2019; VILAR; DE ARAÚJO-CASTRO; MOURA, 2014). Thus, grazers can be good indicators of the toxicity of cyanopeptides.

During young life stages, zebrafish are more sensitive due to a combination of factors including a larger body surface area, high metabolic rate, and absence of mature detoxification pathways as well as limited motility, which can make them restricted to lentic areas where cyanobacteria frequently form scums or benthic populations (DI PAOLO et al., 2015; MALBROUCK; KESTEMONT, 2006). Zebrafish larvae have been broadly used to study the lethal, sublethal and molecular effects of cyanobacterial cultures, focusing on aqueous cyanobacterial extracts or the WHO-recognized toxins (BLAGOJEVI et al., 2021; COSTA et al., 2019; LI et al., 2021; LYDON et al., 2022; RATNAYAKE et al., 2020; YU et al., 2021).

The early evidence published by Oberemm et al. pointed to malformation and acute mortality of 6-day-old zebrafish larvae exposed to different microcystin variants and crude cyanobacterial extracts, the latter showing far more pronounced effects than the purified toxin standards (OBEREMM et al., 1999; OBEREMM; FASTNER; STEINBERG, 1997). More recently, Li et al. (2021) observed severe developmental effects with increased mortality, teratogenicity and decreased hatching and heart rate in 96 hpf-old zebrafish larvae exposed to *Microcystis* cells, which triggered oxidative damage and downregulation of the expression of genes related to the growth

hormone/insulin-like growth factor (LI et al., 2021). Blagojević et al. (2021) screened the toxicity of different cyanobacterial biomass extracts from *Nostoc*, *Anabaena* and *Microcystis* genera, observing acute mortality to 48 hpf zebrafish larvae for 8 strains in the low mg dry weight biomass per litre extract range, and the concentration of microcystins were all in the low µg/L range (assessed by ELISA assay). As to cyanopeptides other than the well-known toxins, Faltermann et al. (2014) found significant transcriptional effects in zebrafish eleuthero-embryos exposed to 100 µg/L cyanopeptolin 1020, altering a large number of transcripts that belonged to genes and pathways related to DNA damage and repair, response to light, neurophysiological process, and circadian rhythm. The authors further observed dose-dependent transcriptional alterations for DNA damage and repair genes, and circadian rhythm regulation. Acute toxicity to 96 hpf zebrafish embryos was also reported for 50 mg/L microcyclamide (60% mortality), and cardiac development was affected in the surviving organisms (FREITAS et al., 2023).

Fernandes et al (2019) attested the toxicity of aqueous cyanobacterial extracts of two Brazilian *Microcystis* strains (MIRS-04 and NPCD-01) towards larvae of native fish *Astyanax altiparanae,* the most common fish species in Brazilian waters. Both strains pointed to acute toxicity and morphological alterations in the low mg dry weight biomass per litre extract range, even though the NPCD-01 does not produce microcystins (FERNANDES et al., 2019a).

Although still incipient, the concern regarding the risk of cyanobacterial metabolites and their presence in aquatic ecosystems was also included in a chapter of the second edition of the guide published by the World Health Organization (WHO) (Toxic cyanobacteria in water - Second edition) (CHORUS; WELKER, 2021). There, the authors list two aspects related to toxicity beyond the already known cyanotoxins: the bioactivity of various metabolites observed in different biological systems; and the toxic effects detected in extracts of cyanobacteria that are not solely explained by the presence of cyanotoxins. Such factors are directly linked and are further reasons to avoid exposure to high concentrations of cyanobacterial biomass, regardless of the cyanotoxin concentrations.

2 RESEARCH QUESTIONS AND OBJECTIVES

Considering the presented context, this PhD project had as a basis the following research questions:

- Do cyanobacterial secondary metabolites other than recognised toxins pose an acute risk to grazers and in the development of fish in the embryo-larval stage?
- Do cyanobacterial metabolites affect these organisms in a sublethal level (*e.g.*, locomotor behaviour)?

In order to provide toxicological information and taking into account our research questions, the main research objective was to assess to acute toxicity of semi-purified cyanopeptides from two Brazilian cyanobacterial species *Microcystis aeruginosa* (strain NPCD-01) and *Microcystis panniformis* (strain MIRS-04).

To that, we made use of high-performance liquid chromatography (HPLC) to fractionate the cyanobacterial extracts, and used semi-purified fractions to expose the freshwater microcrustacean *Thamnocephalus platyurus* (fairy shrimp) and *Danio rerio* (zebrafish) embryos. We focused on the HPLC-fractions containing the highest purity of the main cyanopeptides, analysed by LC-HRMS/MS. Mortality was used to assess the acute toxicity, and the locomotor behaviour to study effects on the sublethal level. Additionally, for the zebrafish embryos, we analysed morphological endpoints (heart rate, body length, and eye and swim bladder sizes).

This work was based on the following hypothesis: (i) freshwater microcrustaceans and early-life stage fish are sensitive to cyanopeptides; (ii) semi-purified peptides can promote sublethal alterations and reflect in the locomotor impairment of microcrustaceans and fish.

The following sections of this thesis will present the materials and methods used in the experimental chapters (Results and Discussion), which in turn will be presented in three parts, briefly:

    I. CyanoMetDB, a comprehensive database of cyanobacterial metabolites;

II. Cyanobacterial metabolites identified on the two Brazilian *Microcystis* strains;

III. Toxicity of cyanobacterial metabolites produced by two Brazilian *Microcystis* strains in the freshwater microcrustacean *Thamnocephalus platyurus*;

IV. Toxicity assessment of cyanopeptides produced by two Brazilian *Microcystis* strains towards zebrafish larvae.

3 EXPERIMENTAL SECTION

3.1 MATERIALS

Solvents used in the LC (Methanol and acetonitrile) were purchased from Thermo Scientific™ (Optima® LC/MS 99.9%), and Formic acid (FA) (98–100%) was obtained from Sigma-Aldrich®. Nanopure water was from a NANOpure® 21 water purification system (Barnstead from Thermo Scientific). Ethanol was purchased from Sigma-Aldrich® (absolute for analysis, EMSURE® ACS,ISO,Reag. Ph Eur). Analytical standards for cyanopeptides were obtained from different sources, as follows: Nodularin and microcystin reference standards with >95% purity (MC-LR, MC-YR, MC-RR, MC-LF, MC-LA, MC-LW, MC-LY, [D-Asp^3]MC-LR, MC-HilR) were acquired from Enzo Life Science (Lausen, Switzerland) and [D-Asp^3,E-Dhb^7]MC-RR (also >95% purity) from CyanoBiotech GmbH (Berlin, Germany); Aerucyclamide A was obtained as a purified bioreagent in dimethyl sulfoxide by Prof. Karl Gademann (University Zurich, Switzerland); and bioreagents (all >90% purity) for aeruginosin 98B, cyanopeptolin A, cyanopeptolin D, anabaenopeptin A, anabaenopeptin B, and oscillamide Y were obtained from CyanoBiotech GmbH (Berlin, Germany). All salts used for culture medium preparation were analytical grade.

3.2 CYANOBACTERIAL CULTURES

The cyanobacterial strains *Microcystis panniformis* (MIRS-04, isolated from the Samuel Reservoir in Rondônia, Brazil), and *Microcystis aeruginosa* (NPDC-01, isolated from a sewage treatment plant, Cidade de Deus, Rio de Janeiro, Brazil) were obtained from the culture collection of the Laboratory of Ecophysiology and Toxicology of Cyanobacteria (LECT), at the Federal University of Rio de Janeiro (UFRJ -Brazil).

Figure 3. Cyanobacterial cultures at LTPNA (USP)
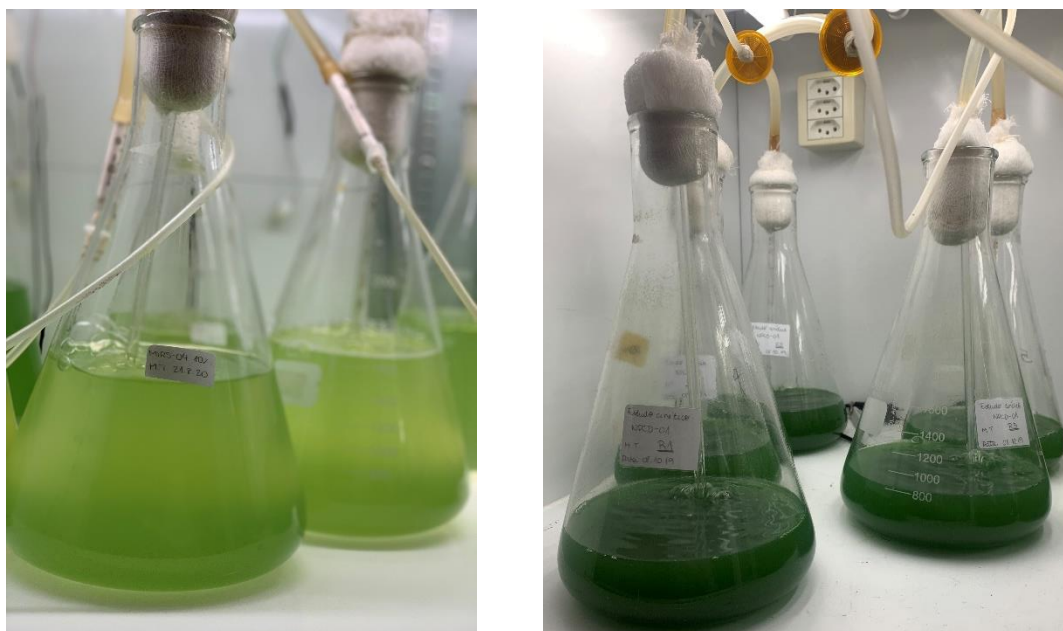
The strains MIRS-04 and NPCD-01 were cultivated in the Laboratory of Toxins and Natural Products of Algae and Cyanobacteria (LTPNA) at the University of São Paulo (FCF-USP) - Brazil, using ASM-1 medium with pH adjusted to 7.5, for 20 days (GORHAM et al., 1964) (Table A1, Appendix II). The culture flasks were maintained under a 12h light/dark photoperiod, with light intensity of 20 μmol photons $m^{-2}s^{-1}$, continuous aeration, and temperature of 24±2 ℃ (Figure 3). After 20 days of growth, the cells were harvested by centrifugation (9,000 rpm at 15 ℃, 10 min). The resulting biomass was stored at -20 ℃, lyophilised (− 55 ℃, 500 μHg, 24 h, L101 Liotop® lyophiliser, Liobras, Brazil) (Figure 4), and sent to the Swiss Federal Institute of Aquatic Science and Technology (Eawag), to perform the LC-HRMS analyses, fractionation, and toxicity tests.

Figure 4. Schematic representation of the cyanobacterial biomass centrifugation and freeze-dry process after the culture
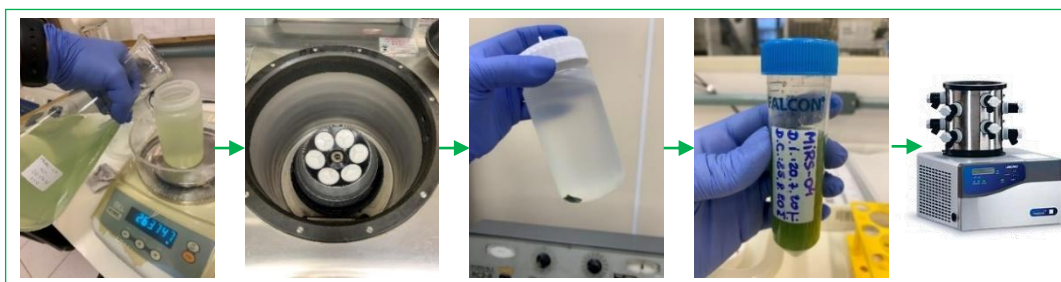
3.3 EXTRACTION OF LYOPHILISED CYANOBACTERIAL BIOMASS AND FRACTIONATION PROCEDURE

1g of freeze-dried cyanobacterial biomass was extracted with MeOH:H_2O (70:30 v/v) at a ratio of 20 µL mg$^{-1}_{dry,wt}$, homogenised by vortexing for 15 seconds and sonicated for 10 minutes (VWR, Ultrasonic cleaner USC-THD, level 6, 15-19 ℃). The solids were separated by centrifugation (Eppendorf Centrifuge 5427 R, 10 min, 10 ℃ and 4000 rcf), the supernatant was transferred to a new glass vial and the extraction procedure was repeated one more time and supernatants were combined. The solvent was evaporated to reduce the methanol content to less than 5% under a gentle stream of nitrogen at minimum flow rate (>0.8 L/min 40 ℃, TurboVap LV, Biotage). The extracts were then diluted with nanopure water, the exact mass recorded (Kern ABT analytical balance 220-5DM, weighing range of 1 mg - 82 g), centrifuged (Eppendorf Centrifuge 5427 R, 10 min, 10 ℃, and 4000 rcf), and filtered (PTFE syringe filters, 0.45µm) into fresh glass vials (Figure 5).

Figure 5. Schematic represetantion of the cyanobacterial biomass extraction procedure
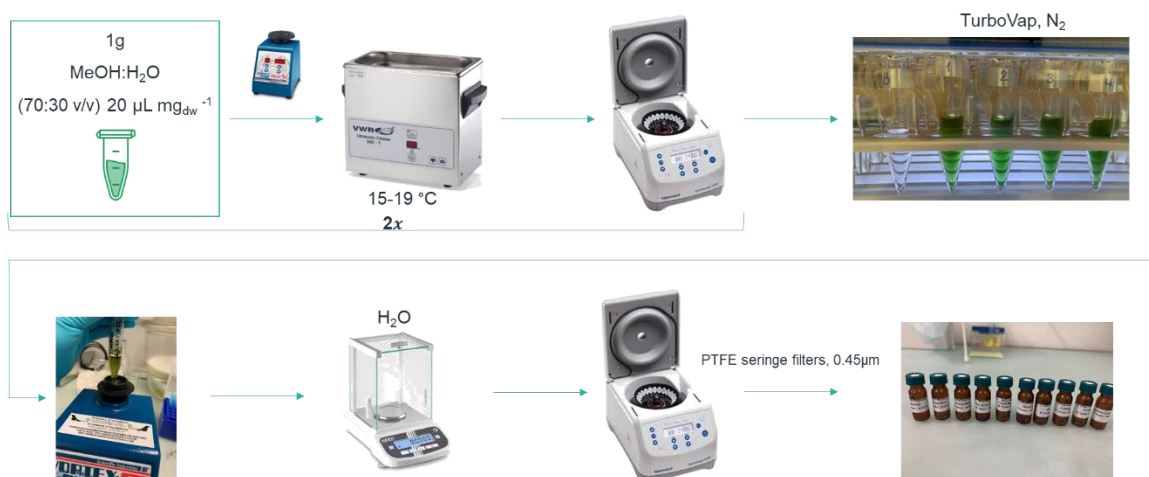
Filtered extracts were fractionated using an UltiMate™ 3000 UHPLC System (Thermo Scientific™) coupled with a FoxyJr® Fraction Collector, and fitted with a Kinetex® LC column (C18, 5 µm particle size, 100 Å pore size, 150 x 4.6 mm, Phenomenex®), with an in-line filter (BGB®). Eluents used for the fractionation were methanol and nanopure water in a gradient elution (20–95% B in 30 min) at a flow

rate of 1.2 mL/min and 100 µL injection volume. Fractions were collected in 50 mL amber glass vials in 11 intervals of 2 minutes each. The first and last 4 minutes of the chromatographic run were diverted to waste, aiming to exclude salts and highly polar compounds (beginning) and highly apolar compounds (end) from the fractions. Blank injections of nanopure water without collecting the fractions ran after every 5 samples, and an extra cleaning run for the analytical column was performed with ACN after 15 sample runs.

The collected fractions were evaporated to dryness using a Syncore® Analyst R-12, BÜCHI Labortechnik AG (shaker temperature 55 ℃, lid temperature 60 ℃, 175 rpm, 20 mbar, cooling system at 4 ℃). Materials were resuspended in ethanol (EMSURE®, absolute for analysis) overnight (LABWIT ZWYR-D2403 incubator, 100 RPM, 12 ℃). Then, the volume of each fraction was adjusted gravimetrically with nanopure water to a 80:20% w/w ethanol:water solution, and stored at -20 ℃ until further use.

Figure 6. Schematic represetantion of the fractionation of the cyanobacterial biomass extracts
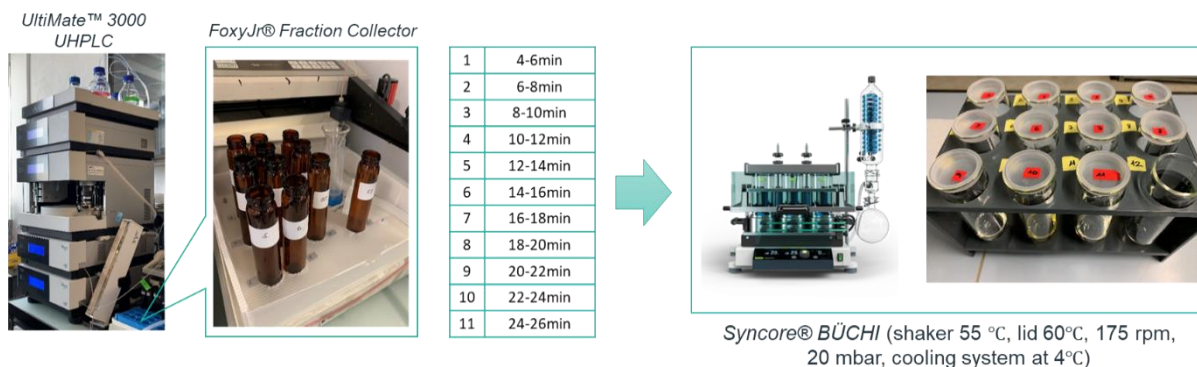
3.4 ANALYSIS OF CYANOPEPTIDES BY LC-HRMS/MS

The fractions from the cyanobacterial extracts of both strains were analysed by high-performance liquid chromatography (Dionex UltiMate3000 RS pump, Thermo Fisher Scientific), coupled with high-resolution tandem mass spectrometry (HRMS/MS, LumosFusion Orbitrap or Exploris 240, ThermoFisher Scientific). The samples were injected into the HPLC system (20µL) with a CTC Analytics autosampler, and the chromatographic separation was carried out on a Kinetex® LC

column (C18, 2.6μm particle size, 100 Å pore size, 100 x 2.1 mm, Phenomenex®) with pre-column (VanGuard Cartridge, Waters) and inline filter (BGB®). Samples were eluted with 0.1% v/v formic acid in nanopure water (solvent A) and methanol (solvent B) in a binary gradient (10 to 95% of B in 25 minutes) at a flow rate of 255 μL/min. After the chromatographic run, the eluates were introduced into an HRMS/MS system (Orbitrap Fusion Lumos or Exploris 240), by an electrospray ionization source (ESI) operated as follows: positive ionization mode with capillary potential of 3.5 kV, 320 ℃ capillary temperature, 40 arbitrary units (AU) sheath gas, 10 AU auxiliary gas, and 275 ℃ vaporizer temperature. Accurate mass spectra were acquired in full scan mode (450-1350 *m/z*) with a nominal resolving power of 120,000 at *m/z* 250, automated gain control (AGC) of $5 \cdot 10^4$, maximal injection time of 50 ms, 1 microscan, and 40% S-lens RF setting. Data-dependent acquisition of MS^2 scans was triggered by an inclusion list containing *m/z* values of the cyanopeptides present in the publicly available database CyanoMetDB (see section 4.1). Data-dependent MS^2 product ion spectra were acquired at a resolving power of 15,000 at 400 *m/z*, AGC of $1 \cdot 10^4$, and maximal injection time of 22 ms, by sequentially recording individual spectra with collision energies of 15%, 30%, and 45%.

3.5 DATA PROCESSING AND CYANOPEPTIDE QUANTIFICATION

HRMS data were first processed using Skyline 20.1 (MacCoss Lab Software) to assign tentative candidates from the database CyanoMetDB and 18 standards or bioreagents that were available (see materials). We followed the confidence of identification scheme (Level 1-5) established earlier for micropollutants (SCHYMANSKI et al., 2014), and adopted previously for cyanopeptides (FILATOVA et al., 2021; NATUMI; DIEZIGER; JANSSEN, 2021; NATUMI; JANSSEN, 2020; NATUMI; MARCOTULLIO; JANSSEN, 2021). The criteria chosen for identifying tentative candidates (level 3) was exact mass error (<3.0 ppm), accurate isotopic pattern (Skyline idotp value >0.93), and consistent RT across samples allowing to link the proposed molecular formula to compounds in CyanoMetDB considering but not differentiating between isomers; probable structures were identified (level 2) based on diagnostic evidence from the fragmentation data (2b) and spectral library match (2a); and we identify confirmed structures (level 1) when these criteria

matched one of our reference standards or bioreagents available. For most suspects, no analytical standard ware commercially available and no reference spectra were deposited in open source database (*e.g.*, MassBank Europe). For those tentative candidates identified by suspect screening, we manually annotated MS^2 spectra with support of in silico molecular fragmentation provided for CyanoMetDB in MetFrag (RUTTKIES et al., 2016) and previously published MS^2 spectra from the primary references listed in CyanoMetDB to increase the confidence of identification and differentiate between structural isomers.

The most abundant peptides were quantified by external calibration with the standard/bioreagent when possible. For those cyanopeptides for which no reference standard or bioreagent was available, the semi-quantification was performed by the MC-LR equivalent approach. Concentrations were only reported when the peak area was above the limits of quantification (LOQs) of the respective regression model.

For the most apolar HPLC fractions (fraction #9, 22 to 24 min; and fraction #10, 24 to 26 min) molecular networks were performed using the Global Natural Products Social Molecular Network (GNPS) platform (available at http://gnps.ucsd.edu, UC San Diego, La Jolla, CA, USA). The raw MS data were converted into mzXML file format data followed by upload to GNPS. GNPS molecular networking parameters were as follows: precursor ion mass tolerance, 0.02 Da; fragment ion mass tolerance, 0.02 Da; min pairs cos, 0.7; network topk, 10; minimum matched fragment ions, 6; minimum cluster size: 2. Additional filters included: precursor window, peaks in a 50 Da window, and spectra from G6 as blanks before networking. Spectra were automatically searched for annotation against the GNPS spectral libraries (thresholds fixed for cosine score >0.7 and at least six matched peaks). The molecular network outputs were enhanced with the MolNetEnhancer tool to provide a more comprehensive chemical overview of the metabolites matched in the given fractions (ERNST et al., 2019). Molecular networks were visualized on Cytoscape 3.8.2.

3.6 SAMPLE PREPARATION OF THE HPLC FRACTIONS FOR THE TOXICITY TESTS

The sample preparation for the toxicity tests consisted of the following: an exact aliquot of the given HPLC fraction (80:20% w/w ethanol:water solution) was transferred to the Büchi Syncore® Analyst R-12 evaporator vials with 1 mL appendix to evaporate the ethanol (shaker temperature 55 ℃, lid temperature 60 ℃, 175 rpm, 20 mbar, cooling system at 4 ℃), remaining only the water in the samples. Solvent controls were included using the same ethanol/water proportion without the added extract. After the evaporation process, the samples were transferred to a clean glass vial, the evaporator vials were rinsed two times with nanopure water, and the samples were gravimetrically adjusted. Note, that for testing the toxicity of the whole extract, an aliquot of each fraction was combined, herein referred to as the pool or pooled fractions, to account for sample preparation steps when comparing effects in the pool and individual fractions.

At this point, the samples were dissolved in nanopure water. Considering it was mandatory to have the exposure solutions diluted in aerated culture medium (described in the next sections), the same exact mass of freshly prepared, double-concentrated culture medium solution was added to the sample, resulting in the correct concentration of salts. The following steps consisted of preparing the exposure solution's dilution series with aerated culture medium.

The concentrations of the pooled and isolated fractions are shown as equivalents of mg of dry weight biomass per mL solution. For that, the initial concentration of each fraction was calculated dividing the amount of biomass injected into the HPLC for fractionation, by the final volume of the resuspended fraction. Then, when an aliquot was taken to prepare the fraction for the toxicity tests and the mass recorded, the final concentration was calculated considering the initial concentration and the final volume of the exposure solution in aerated culture medium.

3.7 ACUTE TOXICITY TEST WITH MICROCRUSTACEANS *THAMNOCEPHALUS PLATYURUS*

The acute toxicity of the fractions obtained from both cyanobacterial strains was screened with the microcrustacean *Thamnocephalus platyurus* (fairy shrimp), using standard operational procedures of the commercially available THAMNOTOXKIT F (Microbiotests Inc., Belgium).

Briefly, the text consists in exposing the Instar II-III larvae of fairy shrimp *T. platyurus* hatched from cysts for 24h in a 12-well plate, under controlled conditions. At the end of the test, the mortality was recorded, and the quantitative effect was evaluated through the lethal concentration of 50% of the total larvae population (LC_{50}). In addition to the acute lethal effects, locomotor effects on the remaining living larvae was also assessed, and will be described in the next section. Appropriate negative, positive and solvent controls were also performed.

3.8 LOCOMOTOR BEHAVIOUR AND DATA ANALYSIS OF THE MICROCRUSTACEANS *THAMNOCEPHALUS PLATYURUS*

To have quantitative data on the locomotor effects of the cyanobacterial extracts on the fairy shrimp, the swimming behaviour was recorded after the 24-h exposure using the DanioVision Observation Chamber (v. DVOC-0040T; Noldus, Netherlands), consisting of a Gigabit Ethernet video camera fitted with infrared and white-light sources and a transparent multi-well plate holder. A standard PC system with the EthoVision XT13 software (version 13.0.1220, Noldus, Netherlands) connected to the camera recorded the videos for later analysis of locomotor activity. All locomotor experiments were conducted in a temperature-controlled room maintained at 26 ± 1 ℃.

After the 24h exposure, each of the remaining living larvae were carefully transferred to an individual well (150µL, 96-well plate). The locomotor activity assay consisted in different phases, and it was an adaptation of previous work with the same set-up with *Artemia salina* (KOHLER; PARKER; FORD, 2021): acclimation period of 2 minutes in light, to settle baseline movement (I); 4 minutes of spontaneous swimming behaviour in light (hereafter referred to as "spontaneous")

(II); two alternated dark intervals of 2 minutes (III and V); and two alternated light intervals of 2 minutes (IV and VI).

Using the EthoVision XT13 software, the larvae movement was re-tracked in each recorded video using a non-live tracking mode, allowing a static subtraction of the background (reducing tracking artefacts) and tracking smooth profiles. Distance moved was chosen as parameter to cross-compare the experimental groups.

3.9 ZEBRAFISH MAINTENANCE AND EMBRYO-LARVAL EXPERIMENT PROCEDURES

Zebrafish, *Danio rerio* (strain WM, mixed wildtype) were maintained in the fish facility at the Swiss Federal Institute of Aquatic Science and Technology (Eawag), under standard controlled conditions, and in accordance with the Swiss animal protection law. Adult fish were maintained in 12 L tanks with a density of one fish per litre, at 26 ± 1 ℃ and under a photoperiod of 14/10h light/dark cycle. The tanks were connected to a recirculating flow-through system supplied with a mixture of tap and desalted water (1:1). Fish were fed twice a day (weekdays) and once a day (weekends) with granulate (ZebraFeed, Sparos, Portugal) and live food (*Artemia nauplii*). For the breeding, spawning trays were added to the tanks by the end of the day, and the eggs collected in the following morning, approximately 1 h post-fertilization. The eggs were rinsed and kept in aerated culture medium (294.0 mg/L $CaCl_2$ $2H_2O$, 123.2 mg/L $MgSO_4 \cdot 7H_2O$, 64.7 mg/L $NaHCO_3$, and 5.7 mg/L KCl in nanopure water, according to ISO-7346/3 guideline) (ISO, 1996). The collected eggs were then pre-exposed in Petri dishes (4 mL of pre-exposure solution), and sorted to add only fertilised eggs in the exposure plate (48-well plates, Greiner Bio-One, Austria). The final volume in each well was 500μL, and exposure plates were covered with adhesive foil and kept in the incubator at the same conditions mentioned above. We tested up to 5 different concentrations of pooled or individual fractions from the cyanobacterial extract. Relevant controls were performed in all tests, namely 4 mg/L 3,4-dichloroaniline (positive control), aerated culture medium (negative controls), and solvent control (from sample preparation explained above). 21 fertilised embryos were used per exposure concentration, 32 embryos for and negative control, 37 for the solvent control, and 16 embryos for the positive control.

Daily observations of lethal and sublethal effects were performed following the OECD guideline 236 for Fish Embryo Acute Toxicity (FET) tests (OECD, 2013) and Kimmel et al. (KIMMEL et al., 1995), using a Leica S8APO stereo microscope. The exposure lasted 120h, and on the final day, in addition to the lethal and sublethal endpoints, morphological, cardiological, and behaviour measurements were recorded, which will be detailed in the following section. pH, dissolved oxygen, and temperature were measured in the control media and the highest concentration of the exposure medium on the first and last day of the FET test. All experiments were carried out according to relevant animal protection guidelines.

3.10 HEART RATE AND MORPHOLOGICAL MEASUREMENTS OF ZEBRAFISH LARVAE

Heart rate and morphological measures of exposed fish and controls were performed on the final day of the experiment (day 5, 120h). Larvae were anaesthetised with 160 mg/L of ethyl 3-aminobenzoate methanesulfonate (MS222; Sigma-Aldrich), a level of anaesthesia previously shown not to influence zebrafish heart rates, and applied in previous work (FITZGERALD et al., 2019).

A 15-second video of each larva in the lateral position was recorded at 30 frames per second (Media Recorder 4 software, version 4.0; Noldus, Netherlands), using a Basler acA2000-165 µM camera installed in the Leica S8APO stereo microscope.

The acquired videos were read in the DanioScope Software (version 1.2.206; Noldus, Netherlands), where heart rate and morphology parameters were measured. The heart area was manually defined in each video, and the software calculated the number of beats per minute using a power plot spectrum of the frequencies extracted from an activity signal. Body length (from nose to tip of the tail; mm), eye size, and swim bladder area (both from the lateral view of the larvae; mm^2) were also measured manually, delimiting one picture frame from the video. The same calibration profile was used to analyse all images to minimise biases during image processing.

3.11 BEHAVIOUR TRACKING AND DATA ANALYSIS OF ZEBRAFISH LARVAE

To assess whether behavioural responses were affected in exposed larvae, behaviour was recorded at 120 hpf using the same DanioVision Observation Chamber (v. DVOC-0040T; Noldus, Netherlands), as previously described.

The behaviour embryo-larval assay, frequently referred to as light-dark transition test, consisted of different phases: acclimation period of 20 minutes in light, to allow the fish to adjust (settle baseline movement) (I); 20 minutes of spontaneous swimming behaviour in light (hereafter referred to as "spontaneous") (II); two alternated dark intervals of 10 minutes (III and V); and two alternated light intervals of 10 minutes (IV and VI).

Using the EthoVision XT13 software, fish were re-tracked in the recorded videos using a non-live tracking mode, allowing a static subtraction of the background and reducing tracking artefacts. Distance moved was also the parameter chosen to cross-compare the experimental groups for the fish.

3.12 STATISTICAL ANALYSES

Non-linear regression model (Prism 9.4.1, GraphPad Software, Inc., San Diego, California) was used to estimate the lethal concentration of 50% of the total population (LC_{50}) exposed to the HPLC-fractions of *M. aeruginosa* (NPCD-01) and *M. panniformis* (MIRS-04) and the slope factor (HillSlope) and goodness of fit (R^2) were assessed. Other statistical analyses (such as Kolmogorov–Smirnov's normality test, Mann Whitney U test, Kruskal–Wallis one-way analysis of variance, and post-hoc Tukey's honestly significant difference test), were also performed with the Prism 9.4.1. The significance level considered was *$p<0.05$*.

4 RESULTS AND DISCUSSION

4.1. CyanoMetDB, A COMPREHENSIVE DATABASE OF CYANOBACTERIAL METABOLITES

Challenges to be addressed for comprehensive risk assessment associated with cyanopeptides in aquatic environments include the high structural diversity of the compounds. Also, very few analytical standards are commercially available, often for elevated costs (JANSSEN, 2019).

To help mitigate the challenges related to the identification of these compounds, the CyanoMetDB, a comprehensive database of cyanobacterial metabolites, was created. This database contains information on over 2000 secondary metabolites produced by these organisms.

Chemical information on the compounds was compiled from existing public databases and in-house libraries of the research teams. From that, a master list of the compounds was created and manually curated, considering primary references in the literature published between 1967 and 2020. Information from open-access databases was checked, expanded and corrected and more than 820 publications were reviewed with a particular focus on providing structural codes (*e.g.*, SMILES codes). All the entries in the database were checked with the 4-eye principle.

This initiative involved research groups with expertise in cyanobacterial metabolites from different countries. The Brazilian team was composed of members of the Laboratory of Toxins and Natural Products from Algae (LTPNA), coordinated by Professor Dr Ernani Pinto.

The first version of the database is presented in a flat file, containing: compound identifier, name, class, molecular formula, molecular weight, monoisotopic mass, primary reference (in which the compound was first reported), SMILES string, InChI string, InChIKey and IUPAC name. For each entry in CyanoMetDB, the following additional fields were also filled in optionally to clarify aspects of the identification of a compound: use of Nuclear Magnetic Resonance (NMR); secondary

reference; genus, species and strain of cyanobacteria in which the compounds were identified; or whether they were identified in environmental samples.

CyanoMetDB covers a large chemical space in terms of molecular weight (from 118 to 2708 Da); of that, 69% are peptides or have a peptide bond in their structure. The compounds were divided into the different classes of cyanobacterial metabolites (such as microcystins, microginins, saxitoxins), making up 10 specific classes. More general classes were included to fit the other compounds where no specific metabolite class could be assigned (*e.g.*, "other cyclic peptides" and "other linear non-peptides").

Comprehensive structural information compiled by the database is useful to facilitate the detection and dereplication of secondary metabolites by targeted analysis, using it in a targeted suspect screening as a list of candidate compounds (using MS1 information). Also, the database can help identify new natural products from cyanobacteria (*e.g.*, using the exclusion method, similarity of MS/MS data, and mass defect). Regarding bioactivity, the database can be of help considering newly discovered compounds tend to share structural similarities of those already known, while changes in structure (even if small) can significantly alter the biological response to that compound. Finally, CyanoMetDB can aid the investigation of the abundance, persistence and toxicity of cyanobacterial secondary metabolites in natural environments, such as by using such structural information in models for predicting the behaviour of these compounds and possible mechanisms of biotransformation (quantitative structure-activity relationships).

The database is freely available in different open-access platforms (*e.g.*, Zenodo), and it was primarily published as open-access in the journal *Water Research* (Appendix I) (JONES *et al.*, 2021).

In response to the challenges posed by the ongoing collection of mass spectrometry data on cyanobacterial natural products, the CyanoMetDB team aims to continuously update and add new entries to the database, presenting new versions at least every two years.

To further improve confidence in compound identification, the next phase of the CyanoMetDB project is to systematically record mass spectral reference data

(including MS^2) of available standards, (semi)purified bioreagents and compounds in extracts of cyanobacterial biomass. Spectra is now being processed, cleaned up and undergoing manual quality checks, and will soon be uploaded to open access databases, as well as an up-to-date version (with reviewed and new entries) of CyanoMetDB.

4.2 CYANOBACTERIAL METABOLITES IDENTIFIED ON THE TWO *MICROCYSTIS* STRAINS

The cyanobacterial biomass from both *Microcystis* strains (NPCD-01 and MIRS-04) were extracted and fractionated. For each HPLC fraction analysed by HPLC-HRMS/MS, cyanopeptide structures were assigned based on suspect screening of full scan data (MS1) against CyanoMetDB (version 1, 2021). These analyses revealed several microcystin and cyanopeptolin variants in HPLC fractions derived from MIRS-04 (Table 1, Figure 1). In terms of total peak area, the four most abundant cyanopeptides in MIRS-04 were micropeptin K139 (MP K139, 74.9%), microcystin-LR (MC-LR, 11.9%), cyanopeptolin 972 (CP 972, 6.3%) and cyanopeptolin 958 (CP 958, 3.7%), with [D-Asp^3]MC-LR, [D-Asp^3,ADMAdda^5]MC-LHar and cyanopeptolin 1014 (CP 1014) being minor components.

Analysing the distribution of the most abundant compound (micropeptin K139) over the 11 HPLC fractions (Figure A1c, Appendix II), fraction #7 with RT 16-18 min, presented the highest share of the compound, and with the highest purity (93% by peak area).

As for HPLC fractions derived from NPCD-01, the most abundant compounds were nostoginin BN741 (NG BN741, 78.4%), microginin FR6 (MG FR6, 9.6%), cyanopeptolin 959 (CP 959, 6.8%), and microginin SD755 (MG SD755, 5.2%).

For the non-producer strain, two HPLC fractions presented the highest amount of two main cyanopeptides, the cyanopeptolin 959-dominated fraction (fraction #1 with RT 4-16 min, 73.3% by peak area, Figure A1b) and the nostoginin BN741-dominated fraction (fraction #3 with RT 8-10 min, 97.2% by peak area, Figure A1c)

The MS^2 spectral annotations of the main cyanobacterial metabolites identified in both strains are presented in Appendix II (Figures A2-A13).

The cyanopeptides identified herein mostly agree with those reported previously by Fernandes et al. (2019), despite differences in the extraction procedure. The only major difference was nostoginin BN741, which in the present study was found to be the major microginin in NPCD-01.

Table 1. Cyanopeptides annotated (level 2-3) or identified (level 1) in methanolic extracts of *Microcystis panniformis* (MIRS-04) and *Microcystis aeruginosa* (NPCD-01).

Microcystis panniformis (MIRS-04)						
Cyanopeptides	**class**	**molecular formula**	**precursor ion type**	**precursor m/z**	**% by peak area**	**level of confidence[1]**
Micropeptin K139	cyanopeptolin	$C_{47}H_{74}N_{10}O_{13}$	[M+H]+	987.5509	74.9	Level 2b
MC-LR	microcystin	$C_{49}H_{74}N_{10}O_{12}$	[M+H]+	995.5560	11.9	Level 1
Cyanopeptolin 972	cyanopeptolin	$C_{46}H_{72}N_{10}O_{13}$	[M+H]+	973.5353	6.3	Level 2b
Cyanopeptolin 958	cyanopeptolin	$C_{45}H_{70}N_{10}O_{13}$	[M+H]+	959.5197	3.7	Level 2b
[D-Asp^3]MC-LR	microcystin	$C_{48}H_{72}N_{10}O_{12}$	[M+H]+	981.5403	1.2	Level 1
Cyanopeptolin 1014	cyanopeptolin	$C_{49}H_{78}N_{10}O_{13}$	[M+H]+	1015.5822	0.7	Level 2b
[D-Asp^3,ADMAdda^5]MC-LHar	microcystin	$C_{50}H_{74}N_{10}O_{13}$	[M+H]+	1023.5509	0.7	Level 3
MC-LY	microcystin	$C_{52}H_{71}N_7O_{13}$	[M+H]+	1002.5182	0.3	Level 1
Microcystis aeruginosa (NPCD-01)						
Cyanopeptides	**class**	**molecule formula**	**precursor ion type**	**precursor m/z**	**% by peak area**	**level of confidence[1]**
Nostoginin BN741	microginin	$C_{39}H_{59}N_5O_9$	[M+H]+	742.4385	78.4	Level 2b
Microginin FR6	microginin	$C_{39}H_{57}N_5O_9$	[M+H]+	740.4229	9.6	Level 3
Cyanopeptolin 959	cyanopeptolin	$C_{43}H_{61}N_9O_{14}S$	[M+H]+	960.4131	6.8	Level 2b
Microginin SD755	microginin	$C_{40}H_{61}N_5O_9$	[M+H]+	756.4542	5.2	Level 2b

[1]According to Schymanski et al, 2014

Figure 7. Cyanopeptide concentration produced by (A) *Microcystis panniformis* MIRS-04 and (B) *Microcystis aeruginosa* NPCD-01 in log-scale semi-quantified as microcystin-LR equivalents (mg/L) representing the highest exposure concentration employed in the toxicity tests of the pooled extract.
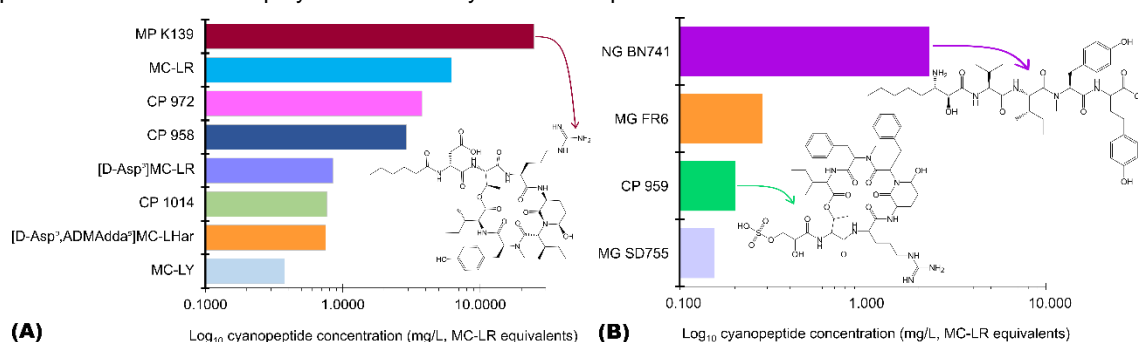
As will be shown and further discussed in the next chapters, the apolar fractions (Fraction #9+10, RT 20-24 min) from the NPCD-01 strain played a key role in the toxicity of the extract. Given that, they were screened for suspects against CyanoMetDB, which gave no conclusive matches of known cyanobacterial metabolites. However, untargeted screening of metabolites using the MolNetEnhancer tool from the Global Natural Products Social Molecular Networking (GNPS) (Table A2, Appendix II) pointed to the presence of glycerolipids, the most abundant one identified being glyceryl palmitate (Figure A14). It is worth mentioning that the same data analysis pipeline (molecular networking and MolNetEnhancer) was applied to analyse the apolar fractions of the MIRS strain, but no matches were found for any of the apolar compounds annotated in the NPCD strain, pointing to differences in their respective metabolite profiles.

4.3 TOXICITY OF CYANOBACTERIAL METABOLITES PRODUCED BY TWO BRAZILIAN *MICROCYSTIS* STRAINS IN THE FRESHWATER MICROCRUSTACEAN *THAMNOCEPHALUS PLATYURUS*

To gain a better understanding on the impact of cyanobacterial metabolites on freshwater organisms, the HPLC-fractions of both extracts (MIRS-04 and NPCD-01) were assayed against the microcrustacean *Thamnocephalus platyurus* for acute toxicity. The commercially available standard biotests on commonly occurring zooplankton crustaceans are widely used in preliminary assessments of cyanobacterial toxicity, and this crustacean species is proven to be particularly sensitive to cyanobacterial metabolites (BOBER; BIALCZYK, 2017; MARŠÁLEK; BLÁHA, 2004; MAZUR-MARZEC et al., 2015).

Although such tests are widely used, their results are limited to mortality as an endpoint, and sublethal effects such as deterioration of the physiological state manifested in reduced mobility is often neglected. Thus, in this study we adapted the zebrafish behaviour protocol (both described in the methods section) to obtain a quantitative measure of locomotor effects also on the microcrustaceans.

4.3.1 Mortality and locomotor impairment of larvae exposed to the MC-producing extract (MIRS-04)

The pooled fractions from the MIRS-04 (MC-containing) extract were tested against *T. platyurus*, and the denoted lethal concentration on 50% of individuals after 24-h exposure was 0.503 $mg_{dw\_biomass}$/mL. In order to inspect the contribution of the main peptide produced by this strain, exposure to the HPLC fraction dominated by the micropeptin K139 (fraction #7, RT 16-18 min 93% by peak area) was also performed separately, and the calculated LC_{50} was 0.958 $mg_{dw\_biomass}$/mL, indicating that the produced micropeptin also plays a role on the toxicity of the MC-containing extract. The remaining toxicity was observed when the pooled fraction was tested subtracting the MPK139-containing fraction (Figure 8A, orange up-triangles), which was right-shifted and showed and LC_{50} of 0.673 $mg_{dw\_biomass}$/mL.

Figure 8. Dose-response mortality curves of *T. platyurus* in terms of mg dry biomass equivalents ($mg_{dw\_biomass}$/mL) at 24h, following exposure to varying concentrations of (A) a pool of all HPLC fractions from MIRS-04 (blue circles), the specific HPLC fraction from MIRS-04 containing primarily micropeptin K139 (purple squares), and the pooled HPLC fractions subtracting the micropeptin K139-containing fraction (orange up-triangles); (B) a pool of all HPLC fractions from NPCD-01 (orange down-triangles), and apolar compounds ('apolar fractions', red squares).
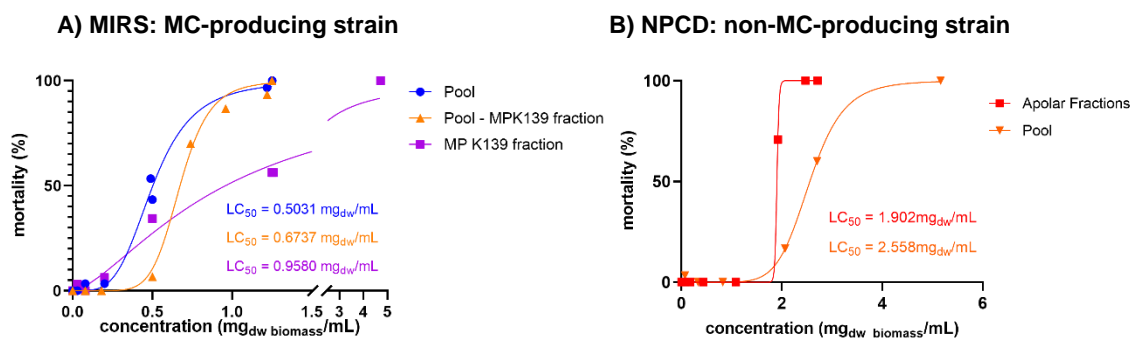
Sublethal effects were also observed in the locomotor assays of larvae exposed to the MIRS pooled fractions, the micropeptin-dominated fraction (fraction #7) and the pooled fractions subtracting fraction #7. When exposed to the pooled fractions, significantly reduced mobility was observed in the concentrations up to 0.20 $mg_{dw\_biomass}$/mL (Figure 9); when only the MPK139-dominated fraction was present (Figure 10), locomotor effects were observed in the two highest concentrations (1.28

48

and 0.51 mg$_{dw\_biomass}$/mL, respectively). Finally, when exposed to pooled fractions without fraction #7 (Figure 11 and Figure A15), mobility effects were noted on concentrations of 0.74 and 0.50 mg$_{dw\_biomass}$/mL, respectively.

Considering that mobility effects observed in the pool spanned to a concentration as low as 0.2mg/mL but not when subtracting fraction #7 (pool minus fraction #7), and locomotor effects were observed in the same concentration range when only the MPK139-fraction was tested, it is possible to affirm that the cyanopeptolin variant identified in the MC-producing strain contributes to the sublethal toxicity (measured as reduced locomotor effects) of the fairy shrimp.

As for the different light conditions of the assay, no marked effects on swimming distances were observed in the different light phases, which may indicate that *Thamnocephalus platyurus* does not present phototaxis nor thigmotaxis behaviour. Kohler et al.. (2021) also observed no significant differences in the swimming behaviour of *Artemia franciscana* in the different stages of the dark-light assay, with different light intensities, using the same behaviour set-up as our study. Additionally, the authors concluded that swimming speed provides a valid measure of the impacts in behaviour of organisms exposed to different psychotropic compounds. Nevertheless, we highlight that a more comprehensive study is needed to gain a better understanding on the phototaxis and thigmotaxis behaviour of *Thamnocephalus platyurus*, as this is not the focus of our study.

The inhibitory activity of cyanopeptides against animal digestive proteases (*e.g.*, elastase, chymotrypsin, trypsin) has been linked with anti-grazing strategies of cyanobacteria. As mentioned previously, many cyanopeptolin and microginin variants showed inhibitory activity against different serine proteases, and micropeptin K139 was proven to be a potent FVIIa-sTF and thrombin inhibitor (IC$_{50}$ values of 10.62 µM and 26.94 µM, respectively), but the inhibition of additional serine proteases were not tested for this micropeptin variant so far. For instance, the cyanopeptolin variant described by Gademann et al. (2010) (cyanopeptolin 1020) showed very potent picomolar trypsin inhibition (IC$_{50}$ of 670 pM) and low nanomolar values against human kallikrein (4.5 nM) and factor XIa (3.9 nM), in addition to acute toxicity (24h) towards *T. platyurus* (LC$_{50}$ of 8.8 µM, corresponding to 9 mg/L).

As for the contribution of microcystins to the toxicity, Esterhuizen-Londt et al. (2018) found a LC$_{50}$ value of 0.1± 0.2 mg/L in terms of total MCs in a *M. aeruginosa* strain isolated from Wangsong reservoir. The authors detected the MC variants MC-LR, MC-RR, and MC-YR in the cyanobacterial extract. TARCZYNSKA et al. (2001), in turn, found a LC$_{50}$ value of 1.8 mg/L of purified MC-LR; however, many studies have pointed to higher toxicity of cyanobacterial extracts when compared to purified toxins (KORPINEN; KARJALAINEN; VIITASALO, 2006; PAWLIK-SKOWROŃSKA; BOWNIK, 2022; TARCZYNSKA et al., 2001).

In our study, the total concentration of the main microcystins identified in the pool corresponded to approximately 0.72 mg/L at around 0.5 mg$_{dw\_biomass}$/mL (LC$_{50}$), reflecting the toxicity of the whole extract. In the micropeptin-dominated fraction, the total MC concentration was below the detection limit at the same biomass concentration range, indicating that the acute toxicity was prompted by the main cyanopeptolin. The concentration of the main cyanopeptides across the different exposure solutions of the MIRS strain is presented in Table 2.

Figure 9. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to different concentrations of MIRS-04 pooled fractions in terms of dry weight biomass. On the boxplot, alterations between groups obtained by Kruskal–Wallis one-way analysis of variance, followed by Tukey's HSD multiple comparison test (95% CI) are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001). NC: Negative control.

Figure 10. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to different concentrations of micropeptin-dominated fraction (#7) of the MIRS-04 strain, in terms of dry weight biomass. On the boxplot, alterations between groups obtained by Kruskal–Wallis one-way analysis of variance, followed by Tukey's HSD multiple comparison test (95% CI) are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001). NC: Negative control.

Figure 11. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to the highest concentration concentrations of the MIRS-04 pooled fractions subtracting the micropeptin-dominated fraction, in terms of dry weight biomass. On the boxplot, alterations between the negative control (NC) and exposed group obtained by Mann-Whitney U test (95% CI) and difference is indicated by asterisks (**p < 0.01).
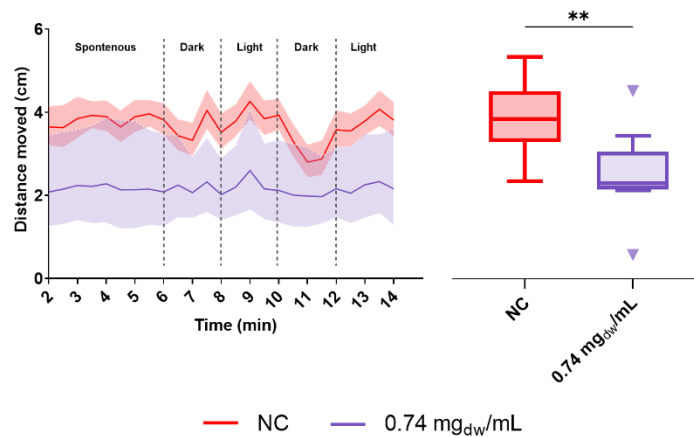
Table 2. Semi-quantification of the main cyanopeptides as MC-LR equivalents in the grazer toxicity test exposure solutions, in the experiments with MIRS-04 strain

Exposure solution	Biomass concentration (mg_{dw}/mL)	Cyanopeptide concentration (mg/L, MC-LR equivalents)								
		MPK139	MC-LR	CP972	CP958	[D-Asp^3]MC-LR	CP1014	[D-Asp^3,ADMAdda^5]MC-LHar	MC-LY	Total MCs
Pooled fractions	1.218	3.241	1.374	0.396	0.486	0.196	0.101	0.156	0.080	1.806
	0.487	1.296	0.550	0.159	0.194	0.069	0.040	0.063	0.043	0.724
	0.195	0.519	0.220	0.054	0.070	0.033	0.010	0.025	0.015	0.294
	0.078	0.181	0.086	0.033	0.037	0.010	0.003	0.009	0.006	0.110
	0.031	0.083	0.044	0.026	0.017	0.004	<LQ	0.004	<LQ	0.054
Fraction #7 (16to18min)	4.718	19.079	0.004	0.041	0.017	0.001	<LD	0.001	<LD	0.006
	1.275	3.014	<LD	0.005	0.002	<LD	<LD	<LD	<LD	<LD
	0.510	1.205	<LD	0.003	<LD	<LD	<LD	<LD	<LD	<LD
	0.204	0.482	<LD	0.001	<LD	<LD	<LD	<LD	<LD	<LD
	0.082	0.193	<LD	<LD	<LD	<LD	<LD	<LD	<LD	<LD
	0.033	0.103	<LD	<LD	<LD	<LD	<LD	<LD	<LD	<LD
Pool - Fraction#7	1.252	0.347	2.034	0.342	0.366	0.236	0.125	0.167	0.096	2.534
	0.960	0.202	0.439	0.288	0.142	0.103	0.068	0.086	0.051	0.679
	0.736	0.146	0.842	0.221	0.109	0.074	0.048	0.067	0.041	1.024
	0.501	0.081	0.814	0.137	0.146	0.068	0.023	0.062	0.046	0.989
	0.200	0.021	0.325	0.059	0.078	0.045	<LQ	0.043	0.017	0.430
	0.080	0.002	0.127	0.015	0.046	0.011	<LQ	0.006	0.007	0.151
	0.032	0.006	0.064	0.014	0.016	0.005	<LQ	0.004	<LQ	0.075

LOQ 0.003 mg/L; LOD 0.001 mg/L <LD: value below the detection limit; <LQ: value below the quantification limit.

4.3.2 Mortality and locomotor impairment of larvae exposed to the extract of the non-MC-producer strain (NPCD-01)

To study whether the metabolites produced by the non-MC producing strain possess acute toxicity towards *T. platyurus*, the pooled fractions were tested in different concentrations, as shown in Figure 8B (orange down-triangles). A dose-dependent mortality was observed after 24h of exposure, with LC_{50} of 2.55 $mg_{dw\_biomass}$/mL, five times higher than the LC_{50} obtained for the pooled fractions of the MC-producing strain.

Considering 100% mortality was achieved in the highest concentration, and earlier evidence already pointed to bioactivity/toxicity of the metabolites produced by this strain (FERNANDES et al., 2019a; LAUGHINGHOUSE et al., 2012; SILVA-STENICO et al., 2011), all the HPLC-fractions were screened separately at the fixed concentration of 2.71 $mg_{dw\_biomass}$/mL, in an attempt to identify what was the contribution of each fraction for the toxicity of the whole extract.

Screening of the different fractions revealed that the apolar fractions (fraction #9+10, RT 20-24 min) were the main contributors to the acute toxicity (100% mortality achieved in the highest concentration tested), resulting in LC_{50} value of 1.90 $mg_{dw\_biomass}$/mL (Figure 8B, orange down-triangles).

Comparing the mortality results from the non-MC-producing strain, it is possible to observe a left shift in the apolar fraction's curve and a lower LC_{50} value, which indicates that the apolar compounds were more toxic than when combined with the other cyanobacterial metabolites from the strain (*e.g.* microginins and cyanopeptolin present in the pool), marking an antagonistic effect.

All the other 9 HPLC fractions, including the cyanopeptolin 959-dominated fraction (fraction #1 with RT 4-16 min, 73.3% by peak area), and the nostoginin BN741-dominated fraction (fraction #3 with RT 8-10 min, 97.2% by peak area) did not induce any mortality in the microcrustaceans.

The observation of the locomotor behaviour also pointed to significant impaired mobility of the organisms exposed to the pooled fractions (Figure 12) and apolar fractions (Figure 13 and Figure A16), in the highest concentration where

enough organisms survived (*i.e.,* 2.06 $mg_{dw\_biomass}$/mL and 1.92 $mg_{dw\_biomass}$/mL, respectively).

Literature on the toxicity of apolar compounds from cyanobacteria in microcrustaceans is scarce. Egli et al. (2020) observed strong inhibition of aminopeptidase by the apolar extracts (85:15% v/v methanol: water) of two *Microcystis aeruginosa* strains, one being a non-MC-producer (knocked-out mutant). Both strains produced different cyanopeptolin and aerucyclamide variants, which were mainly present in the apolar fractions.

Studies showing that crude extracts from cyanobacteria have a greater impact on tested organisms than purified extracts have been often reported (as reviewed by CHORUS; WELKER, 2021). Adverse effects were also shown in crustaceans exposed to toxin-free cyanobacterial extracts or toxin-free fractions, as shown by Keil et al. (2002); Korpinen et al. (2006); Mazur-Marzec et al. (2015). Lastly, apolar fractions of cyanobacterial extracts also exerted developmental toxicity in zebrafish embryos, which will be shown and discussed in more detail in the following chapter.

Figure 12. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to different concentrations of HPLC-pooled fractions from the NPCD-01 strain, in terms of dry weight biomass. On the boxplot, alterations between groups obtained by Kruskal–Wallis one-way analysis of variance, followed by Tukey's HSD multiple comparison test (95% CI) are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001). NC: Negative control.
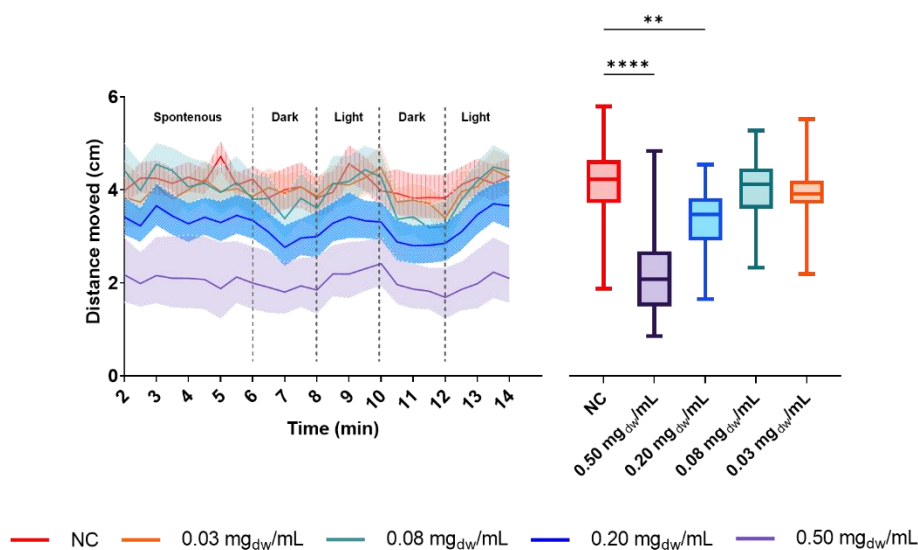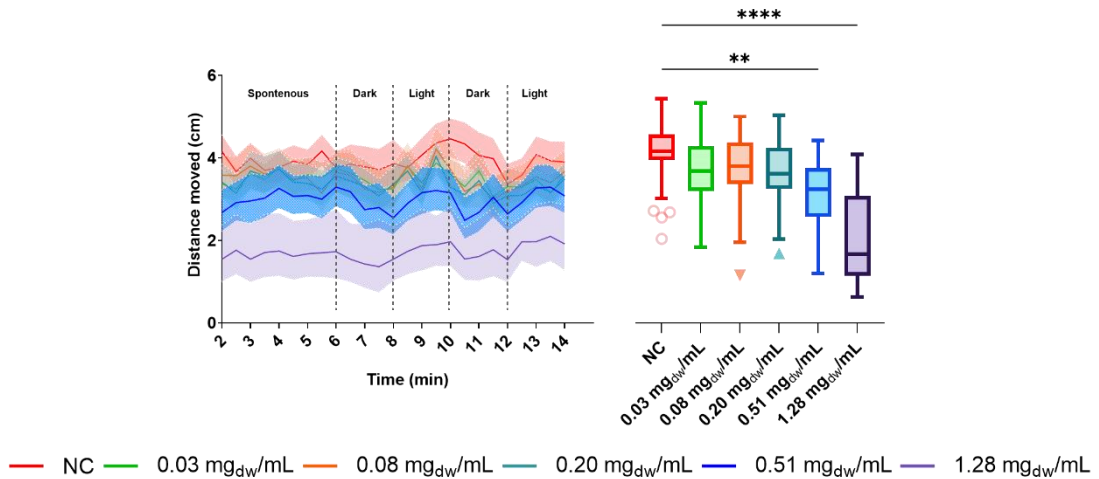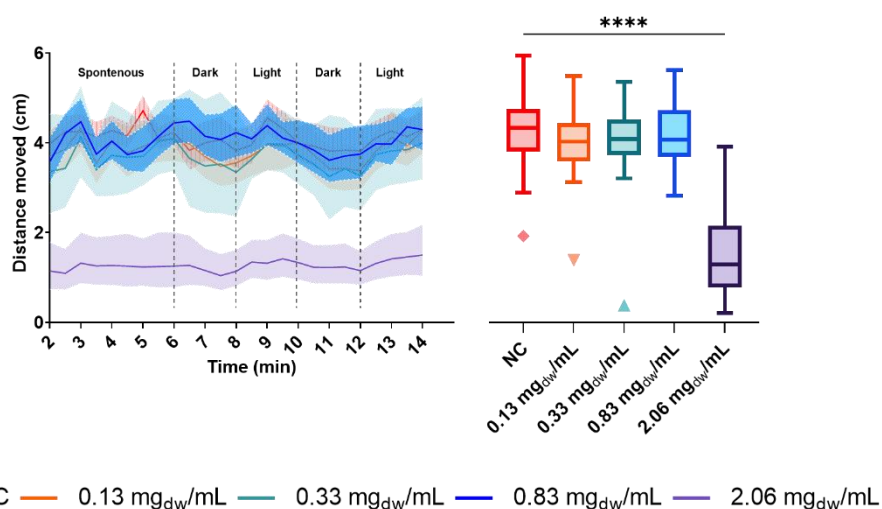

Figure 13. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to the highest concentration of the apolar fractions (F#9+10) from NPCD-01, in terms of dry weight biomass. On the boxplot, alterations between the

negative control (NC) and exposed group was obtained by Mann-Whitney U test (95% CI) and difference is indicated by asterisks (**p < 0.01).
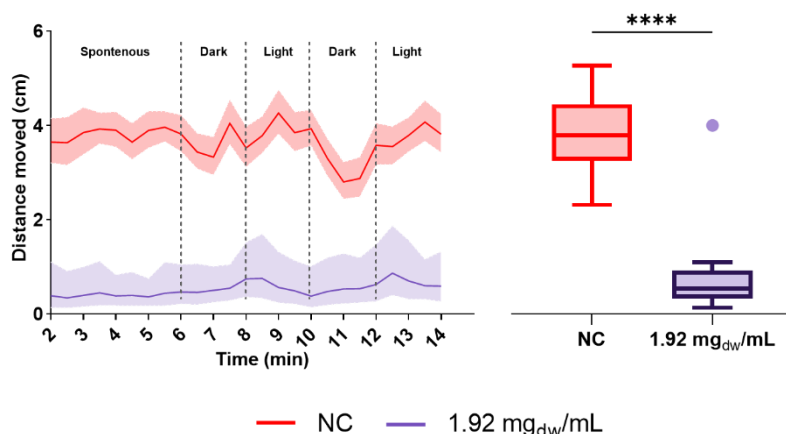

4.4 TOXICITY ASSESSMENT OF CYANOPEPTIDES PRODUCED BY TWO BRAZILIAN *MICROCYSTIS* STRAINS TOWARDS ZEBRAFISH LARVAE

Zebrafish embryos were exposed to HPLC fractions of MIRS-04 and NPCD-01 biomass, in accordance with OECD test guideline No. 236 (2013). Throughout the exposure, zebrafish were inspected for acute and sublethal effects, with behavioural and morphological effects also screened at the final exposure time point (120h). The pooled fractions for MIRS-04 and NPCD-01, prepared by combing aliquots of all associated HPLC fractions, were initially assessed for acute toxicological effects, before inspecting the contribution of individual fractions.

Exposure to the pooled fractions revealed a dose-dependent relationship between zebrafish mortality at the 120 hpf time point for both cyanobacterial strains (Figure 14A and Figure A17). The concentration of pooled fractions that induced mortality in 50% of larvae (LC_{50} value), was 0.49 (0.42-0.58) $mg_{dw\_biomass}$/mL for MIRS-04 and 0.98 (0.89-1.07) $mg_{dw\_biomass}$/mL for NPCD-01 at 120 hpf.

For both strains, mortality occurred by 24 hpf and did not further increase until the final time point at 120 hpf, as shown in the survival graphs (Figure A18, Appendix II). The LC_{50} values for MIRS-04 and NCPD-01 pooled fractions were 2-fold and 3-fold higher, respectively, compared to values measured previously using native Brazilian fish larvae *Astyanax altiparanae* (0.24 and 0.32 $mg_{dw\_biomass}$/mL of the aqueous extract, respectively) (Fernandes et al 2019). The present study further supports earlier observations that *Microcystis* strain NPCD-01 does not produce

microcystins nor any other recognized cyanotoxins (i.e., anatoxin-a, cylindrospemopsin, saxitoxin), yet causes significant acute and lethal toxicity in both fish species (Fernandes et al. 2019). To further investigate the contribution of individual cyanopeptides to the observed mortality, zebrafish embryos were exposed to individual HPLC fractions, dominated by different cyanopeptides.

For the MIRS-04 extract, the micropeptin K139-dominated fraction (fraction #7 with RT 16-18 min, 93% by peak area) induced dose-dependent mortality, with maximum mortality at 120 hpf of 52% at the two highest exposure concentrations (8.74 and 3.49 $mg_{dw\_biomass}$/mL equivalents). By contrast, the two highest concentrations of the MIRS-04's pooled fractions (5.43 and 2.17 $mg_{dw\_biomass}$/mL equivalents) induced 100% mortality at 120 hpf.

Despite 100% mortality having not been reached, an LC_{50} value at 120 hpf was estimated for the micropeptin K139-dominated fraction to be 6.21 (4.55-8.47) $mg_{dw\_biomass}$/mL, almost 15-fold higher compared to the pooled fractions resembling the total extract. Other than the pooled fractions discussed above, this micropeptin K139 fraction showed an increase in mortality over exposure time between 24-72 hpf at the two highest exposure concentrations (8.74 and 3.49 $mg_{dw\_biomass}$/mL, Figure A18, a and b). These observations suggest that micropeptin K139 contributes significantly to the observed toxicity in the pooled extract, and the remaining lethal toxicity may be linked to microcystins present. Studies exploring the acute toxicity of microcystins in zebrafish embryos are limited. Wei et al. (2020) reported LC_{50} of 2.79 mg/L for MC-LR (72 hpf zebrafish embryos). Here, the LC_{50} is 0.49 mg_{dw}/mL of the pooled fractions of MIRS-04, which contained 0.7 mg/L MC-LR and the contribution of the other microcystins was less than 3% of that. The LC_{50} value for MC-LR observed by Wei et al. (2020) is 2.6-fold higher than the concentration found in our study at the LC_{50} level, and further supports that micropeptin K139 contributed to mortality.

For the extract of the non-producer of microcystins (NPCD-01), the nostoginin BN741-dominated fraction (fraction #3 with RT 8-10 min, 97.2% by peak area) and the cyanopeptolin 959-dominated fraction (fraction #1 with RT 4-16 min, 73.3% by peak area) were inspected separately. The nostoginin BN741-dominated fraction triggered only low mortality, with the two highest concentrations (12.48 and 4.99

mg$_{dw\_biomass}$/mL, respectively) causing 14% lethality. The cyanopeptolin 959-dominated fraction could only be tested at low concentrations where it did not lead to acute effects (Figure 2C). We conclude that the two most abundant cyanopeptides nostoginin BN741 and cyanopeptolin 959 alone cannot explain the mortality inflicted by the extract of NPCD-01, which does not produce MCs or other recognized toxins. The concentration of the cyanopeptides of each strain in each of the FET test exposure solutions are presented in Table A3 (Appendix II).

Further screening of the remaining HPLC fractions revealed that the apolar fractions (fraction #9+10, RT 20-24 min) were the main contributors to the acute toxicity, resulting in LC$_{50}$ values 120 hpf of 1.52 mg$_{dw\_biomass}$/mL (Figure 2C). As previously mentioned, the suspect screening of these apolar fractions gave no conclusive matches of known cyanobacterial metabolites on CyanoMetDB, and the untargeted screening of metabolites using the MolNetEnhancer tool from the Global Natural Products Social Molecular Networking (GNPS) pointed to the presence of glycerolipids, the most abundant one identified being glyceryl palmitate.

In a recent study, Roegner et al. (2019) observed high mortality and developmental impairment in zebrafish embryos exposed to lipophilic extracts of *Microcystis* bloom samples from the La Plata river, Uruguay (extract of 1 mg dry with biomass/mL). Similarly, in a toxicity screening study, Berry et al. (2007) tested over 122 cyanobacterial polar (EtOH) and apolar (CHCl$_3$) extracts with the zebrafish developmental toxicity assay. Of those, the majority of the toxic extracts were lipophilic, and all embryos exposed to concentrations of 50 µg/mL or higher were found to die within 24 hours, regardless of developmental stage. Fish eggs and embryos can rapidly achieve high tissue concentrations of hydrophobic chemicals present at very low concentrations in the surrounding water (Incardona and Scholz 2016), and the more pronounced effect of lipophilic compounds from cyanobacterial extracts on the embryo development is presumably due to the ability of those compounds to more easily cross membranes (Berry et al. 2007; Roegner et al. 2019). It is important to note that in both studies mentioned, different levels of developmental toxicity were observed in embryos, dependent on the lipophilic extract tested (different bloom samples or different cyanobacterial strain biomasses). Such difference suggests that the developmental impairment observed in the fish is related

to the presence of specific metabolites, rather than a general toxic effect caused by the apolar extracts, albeit the yet unknown chemical nature of these compounds, as highlighted by Berry et al. (2009).

We emphasise that the acute toxicity observed in the apolar fractions might not only be explained or attributed to the metabolites identified in the untargeted screening in our study, considering that their composition could not be entirely elucidated, and toxicity was also observed in the aqueous extracts by Fernandes et al. (2019). To conclude, the mortality observed for the microcystin-free fraction of MIRS-04 and the extract of NPCD-01, a non-microcystin producer, further adds to the existing body of literature that cyanobacterial metabolites other than microcystins can cause acute (lethal) toxicological effects. We investigated further sublethal toxicity endpoints for those concentrations and fractions where no or low acute effects were observed and where enough individuals survived to facilitate investigation of morphological and behavioural effects.

Figure 14. Dose-response mortality curves of zebrafish larvae relative to mg dry biomass equivalents (mg$_{dw\_biomass}$/mL) at 120 hpf, following lifetime exposure to varying concentrations of (A) all pooled HPLC-fractions of MIRS-04 (blue up-triangles) and NPCD-01 (orange down-triangles); (B) a pool of all HPLC fractions from MIRS-04 (blue up-triangles) and the specific HPLC fraction from MIRS-04 containing primarily micropeptin K139 (dark-red squares); (C) a pool of all HPLC fractions from NPCD-01 and the individual HPLC fractions from NPCD-01 containing mainly nostoginin-BN741 (purple squares), cyanopeptolin-959 (green down-triangles) and apolar compounds ('apolar fractions'; red circles); and (D) microscope images (x3 magnification) of a healthy zebrafish larvae from the control wells (i) versus representative examples of larvae with heart oedema following exposure to MIRS-04 micropeptin K139 fraction (ii) and NPCD-01 nostoginin-BN741 fraction (iii).
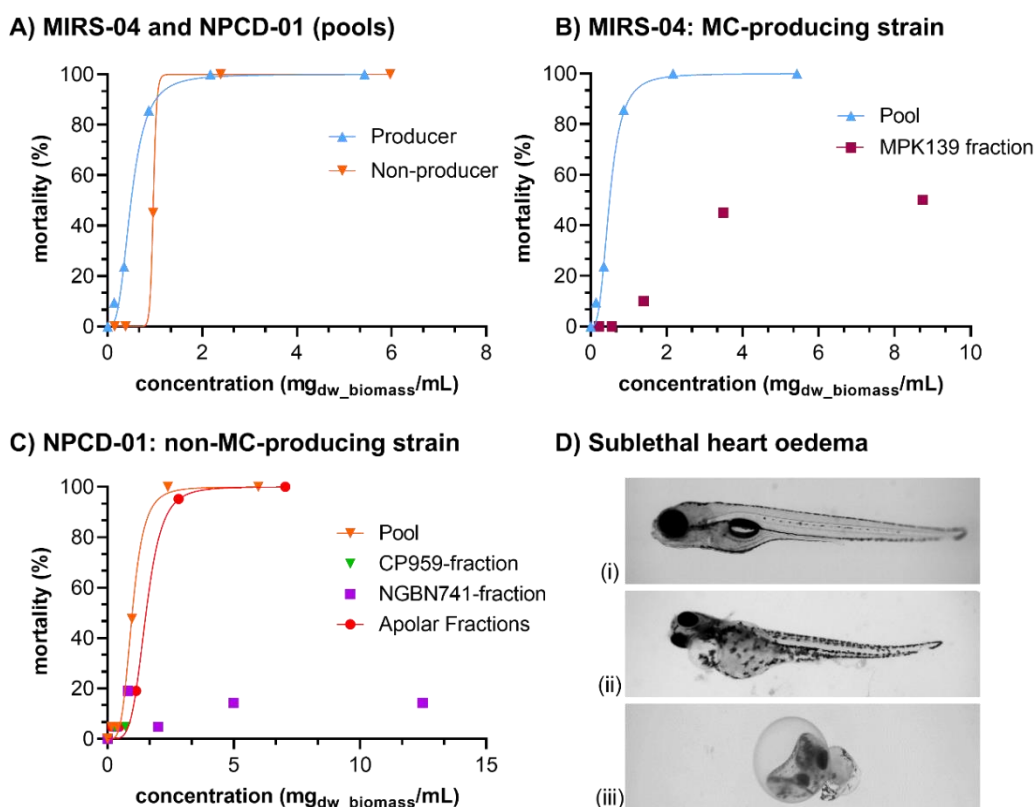
4.4.1 Sublethal morphological effects

Sublethal effects and morphological abnormalities (type and frequency) were monitored in zebrafish larvae exposed to pooled and individual HPLC fractions, derived from both MIRS-04 and NPCD-01 extracts. These sublethal endpoints included observations of malformation in the 120 hpf larvae, such as: oedema of the pericardial region; yolk deformations; malformation of the tail; general delay of development; malformation of the head; modified axis structure; yolk oedema; and no reaction to physical trigger.

Here both the NPCD-01 pooled fraction and the nostoginin BN741-containing fraction (fraction #3), tested separately, showed increased frequency of oedema of the pericardial region (Table 3, Figure 14D). The nostoginin BN741-fraction also prevented hatching in all affected larvae. The cyanopeptolin-959-containing fraction (fraction #1), which also showed no lethal effects, also triggered no observed sublethal effects in zebrafish embryos. For the microcystin-producing MIRS-04 strain, the pooled fraction gave rise no sublethal effects. However, the micropeptin K-139-containing fraction (fraction #7) derived from MIRS-04 caused oedema of the pericardial region.

Furthermore, morphological measures including body length, swim bladder and eye sizes, as well as changes in the heart rate (cardiological effects), were assessed for surviving larvae exposed to the pooled and individual fractions from both cyanobacterial strains. We measured such features aiming to have quantitative data on the meaningful morphological changes upon exposure, and measurements were performed in all surviving larvae, unless severely deformed. The results were statistically compared to corresponding controls (95% CI, Kruskal–Wallis one-way analysis of variance, Tukey's HSD test). Heart rates and body length were only affected in larvae exposed to MIRS-04 pooled fractions at 0.35 $mg_{dw\_biomass}$/mL (C4), which approximately matches the estimated LC_{50} range (0.34-0.52 $mg_{dw\_biomass}$/mL, Figure 3). At this concentration (C4) 16 out of 21 larvae survived, and a higher concentration could not be considered because mortality was too high, *i.e.*, two or fewer larvae survived. The micropeptin-K139-containing fraction from MIRS-04, did not show significant effects on these other morphological endpoints for larvae that survived and were not already affected by oedema of the pericardial region (Figure A19). Exposure of zebrafish embryos to the NPCD-01 pooled fraction at 0.96 $mg_{dw\_biomass}$/mL (C3) appeared to lower heart rates versus solvent control, though results were not statistically significant (p<0.01, Figure A20). The apolar fractions from NPCD-01 showed similar LC_{50} values as the pooled fractions and also showed a trend of lower heart rates exposed at 1.13 $mg_{dw\_biomass}$/mL exposure (C3), differing significantly from the controls (p>0.05), despite the pericardial area showing no significant variations (Figure A21). Heart rate and morphological measurements of larvae were not significantly affected when exposed to any concentration of the

cyanopeptolin-959-containing fraction (fraction #1) nor the nostoginin BN741-containing fraction (fraction #3) from NPCD-01 (p>0.05, Figures A22-23).

Roegner et al. (2019) exposed zebrafish embryos to different cyanobacterial metabolites (including known cyanotoxins), and reported that, of the 24 different metabolites tested, only cylindrospermopsin, [D-Asp^3]-MC-RR, L-BMAA, lyngbyatoxin-a, MC-LR and MC-RR resulted in significant embryo mortality after 5 days of exposure. However, similarly to our findings, the authors observed pericardial or yolk sac edema in embryos exposed to cyanopeptolin 1040 MB, microginin 690, anabaenopeptin B, anatoxin-a, MC-LA, and MC-HilR.

Overall, morphological assessments suggest that micropeptin K139 can contribute to lethal toxicity and oedema of the pericardial region; nostoginin BN741 also contributes to oedema of the pericardial region but not to acute mortality at the concentrations tested. As mentioned previously, many cyanopeptolin and microginin variants showed inhibitory activity against different serine proteases, and micropeptin K139 was proven to be a potent FVIIa-sTF and thrombin inhibitor (EC$_{50}$'s of 10.62 uM and 26.94 uM, respectively) (Anas et al. 2017). In zebrafish, Protease Activated Receptors (PARs) were shown to regulate the development and differentiation of the lymphatic system (Lei et al. 2021), and among the PAR upstream proteases, thrombin, plasmin, factor Xa, factor VIIa, trypsin, and matrix metalloproteases have been identified. We hypothesise that the cardiovascular effects observed in zebrafish (*i.e.*, oedema of the pericardial region), may arise through inhibition of the proteases, though further studies exploring *in vivo* inhibition of proteases would be required to test this hypothesis.

Lastly, the disruption of cardiac function during early organogenesis in fish larvae can subsequently alter cardiac morphology and function, initiating a cascade of adverse effects on cardiorespiratory and physiological performance (Incardona and Scholz 2016). The extent of the observed cardiological sublethal effects suggests that exposure to these concentrations of nostoginin BN741 and micropeptin K139 can potentially compromise the development and physiological performance of the fish and, as a consequence, its survival.

Table 3. Frequency and types of sublethal effects observed over 120h exposure of zebrafish larvae to MIRS-04 and NPCD-01 HPLC-fractions. The total number of larvae per replicate was n= 21 for the exposed groups, n=34 for negative control, and n= 37 for the solvent control.

NPCD-01 (non-producer)

Pooled fractions

Exposure group	Total exposed larvae (alive)	Total affected larvae	OPC	YD	MT	D	MH	A	YE	NR
5.97 mg/mL	0	0	-	-	-	-	-	-	-	-
2.39 mg/mL	0	0	-	-	-	-	-	-	-	-
0.96 mg/mL	12	1	1/1	0	1/1	1/1	1/1	0	1/1	0
0.38 mg/mL	21	2	2/1	1/2	0	0	1/2	0	1/2	0
0.15 mg/mL	22	3	3/3	1/3	1/3	1/3	1/3	0	1/3	1/3
SC	33	1	1/1	0	0	0	0	0	0	0
NC	31	2	1/2	1/2	0	0	0	0	0	0

Nostoginin BN741-containing fraction (#3)

Exposure group	Total exposed larvae (alive)	Total affected larvae	OPC	YD	MT	D	MH	A	YE	NR
12.48 mg/mL	21	3	2/3	0	0	1/3	1/3	0	1/3	0
4.99 mg/mL	21	3	3/3	0	0	0	0	0	0	0
1.99 mg/mL	21	1	1/1	0	0	0	0	0	0	0
0.80 mg/mL	21	4	4/4	1/4	0	1/4	0	0	0	0
0.32 mg/mL	21	1	0	0	0	1/1	0	0	0	0
SC	37	0	0	0	0	0	0	0	0	0
NC	34	1	0	0	0	1/1	0	0	0	0

Cyanopeptolin 959-containing fraction (#1)

Exposure group	Total exposed larvae (alive)	Total affected larvae	OPC	YD	MT	D	MH	A	YE	NR
11.27 mg/mL	21	0	0	0	0	0	0	0	0	0
4.51 mg/mL	21	0	0	0	0	0	0	0	0	0
1.80 mg/mL	21	0	0	0	0	0	0	0	0	0
0.72 mg/mL	21	1	1/1	0	0	0	0	0	0	0
0.29 mg/mL	20	0	0	0	0	0	0	0	0	0
SC	37	1	1/1	0	0	0	0	0	0	0
NC	34	0	0	0	0	0	0	0	0	0

Apolar fractions (#9+10)

Exposure group	Total exposed larvae (alive)	Total affected larvae	OPC	YD	MT	D	MH	A	YE	NR
7.04 mg/mL	0	0	0	0	0	0	0	0	0	0
2.81 mg/mL	1	0	0	0	0	0	0	0	0	0
1.13 mg/mL	17	2	2/2	0	0	1/2	2/2	0	0	0
0.45 mg/mL	20	0	0	0	0	0	0	0	0	0
0.18 mg/mL	20	0	0	0	0	0	0	0	0	0
SC	35	1	0	0	0	0	0	1/1	0	0
NC	32	1	1	0	1/1	1/1	1/1	0	1/1	0

MIRS-04 (MC-producer)

Pooled fractions

Exposure group	Total exposed larvae (alive)	Total affected larvae	OPC	YD	MT	D	MH	A	YE	NR
5.43 mg/mL	0	0	-	-	-	-	-	-	-	-
2.17 mg/mL	0	0	-	-	-	-	-	-	-	-
0.87 mg/mL	2	1	1/1	0	0	0	0	0	0	0
0.35 mg/mL	16	0	0	0	0	0	0	0	0	0
0.14 mg/mL	19	0	0	0	0	0	0	0	0	0
SC	32	5	5/5	1/5	0	4/5	1/5	0	0	0
NC	30	4	3/4	1/4	1/4	3/4	1/4	0	0	0

Micropeptin K139-containing fraction (#7)

Exposure group	Total exposed larvae (alive)	Total affected larvae	OPC	YD	MT	D	MH	A	YE	NR
8.74 mg/mL	10	4	3/4	2/4	0	0	0	0	0	3/4
3.49 mg/mL	11	2	1/2	0	2/2	0	0	1/2	0	1/2
1.40 mg/mL	18	2	2/2	0	0	0	0	0	1/2	0
0.56 mg/mL	20	1	1/1	1/1	0	0	0	0	1/1	0
0.22 mg/mL	20	1	1/1	0	0	0	0	0	1/1	0
SC	37	0	0	0	0	0	0	0	0	0
NC	34	0	0	0	0	0	0	0	0	0

OPC: oedema of the pericardial region; YD: yolk deformations; MT: malformation of the tail; D: general delay of development; MH: malformation of the head; A: modified axis structure; YE: yolk oedema; NR: no reaction to trigger; SC: solvent control; NC: negative control. Concentrations refer to mg of dry weight of biomass equivalents per mL.

Figure 15. Heartbeat(a), body length (b), eye size (c), and swim bladder size (d) measured of surviving larvae exposed to MIRS-04 pooled fractions (120h) at the second to lowest and lowest concentration levels tested (C4 and C5, respectively) and for the negative control (NC) and the solvent control (SC). Significant alterations between groups are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001).
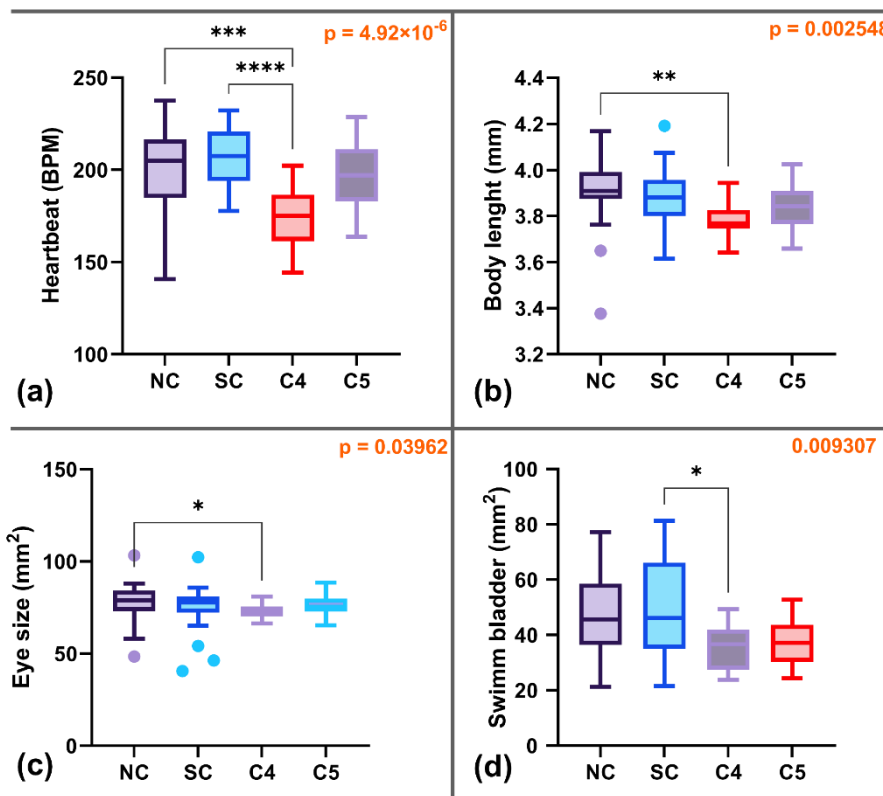
4.4.2 Behavioural effects

The behaviour tracking assay consisted of different phases: acclimation phase (20 minutes in light) (I); spontaneous swimming behaviour in light (20 minutes) (II); two alternated dark intervals of 10 minutes ("dark intervals", III and V); and two alternated light intervals of 10 minutes ("light intervals", IV and VI). The light-dark transition test in early life stages of zebrafish is frequently used to assess developmental neurotoxicity (Könemann, von Wyl, and vom Berg 2022), and each phase is illustrated in Figure 16. The peaks of increased locomotor activity normally occurring in unexposed zebrafish larvae can be noted in the dark intervals (III and V), known as light seeking behaviour.

In addition to the acute and sublethal effects observed, the exposure of zebrafish larvae to MIRS-04 pooled fraction at 0.35 mg$_{dw\_biomass}$/mL (C4) caused a significant decrease in locomotor activity during the dark stages, compared to that

recorded in larvae exposed at 0.14 mg$_{dw}$/mL (C5) and to controls (p<0.05, Figure 16). No significant changes in locomotor behaviour were observed for larvae exposed to micropeptin-containing fraction #7 (MIRS-04), NPCD-01's pooled fraction, nor nostoginin-containing fraction #3 from NPCD-01 (Figures A24-28). Note, that in some cases, the locomotor activity of exposed larvae differed significantly from the negative control in the dark phase but not from the solvent control. The solvent control, in turn, was then not significantly different from negative control (Figures A24-28).

Recent evidence indicates the neurotoxicological potential of microcystins on the nervous system of zebrafish (Martin, Bereman, and Marsden 2021; Qian et al. 2018; Yu et al. 2021b). Similar to our study, Qian et al. (2018) observed decreased locomotor activity in a concentration-dependent manner during the dark phase of the behaviour assay on zebrafish larvae exposed to a microcystin-producing strain of *Microcystis aeruginosa*. In addition, the levels of both acetylcholinesterase (AChE) and dopamine (DA) were reduced, followed by abnormal changes in the transcription of several genes associated with these neurotransmitters and the expression of nine marker genes for nervous system function and development. Yu et al. (2021) and Martin et al. (2021) went even further in the study of neurotoxicological effects of microcystins, using proteomic techniques to identify molecular changes of zebrafish exposed to the cyanotoxin. Both studies attested that zebrafish organisms are more susceptible to neurodegeneration and neurological disorders when exposed to microcystins. Our observations suggest that neither micropeptin K139 nor nostoginin BN741 affect zebrafish larvae behaviour, unlike the cited evidence suggests microcystins do; on the other hand, both peptides did induce sublethal cardiovascular effects in exposed larvae.

Figure 16. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplots representing the average distance moved in each phase of the behaviour assay - spontaneous, light intervals, and dark intervals - of larvae exposed to different concentrations of MIRS-04 pooled fractions. On boxplots, alterations between groups obtained by Kruskal–Wallis one-way analysis of variance, followed by Tukey's HSD multiple comparison test (95% CI) are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001). NC: Negative control; SC: Solvent control.
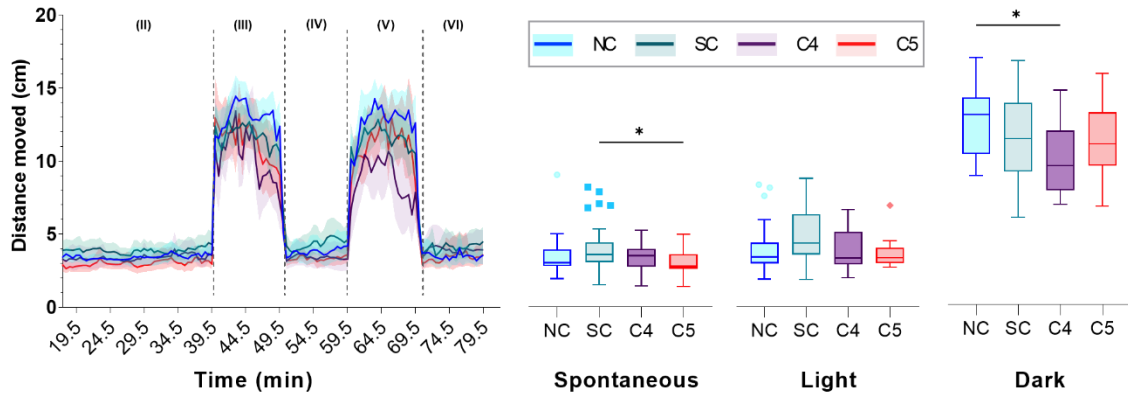
5 CONCLUSIONS AND OUTLOOK

A recent number of studies have pointed to toxicological risk of cyanobacterial metabolites beyond the WHO-recognized cyanotoxins and of toxin-free cyanobacterial extracts, in different levels of biological organisation and in organisms from different phyla. Our results further add to the existing body of literature highlighting that cyanobacterial metabolites other than microcystins can cause acute lethal and sublethal toxicological effects in aquatic organisms. We have confirmed that the NPCD-01 strain does not synthesise microcystins, yet inflicts comparable lethality in zebrafish larvae and microcrustaceans to that induced by the microcystin-producing MIRS-04 strain. Our results suggest that the toxicity of the non-producer was mostly prompted by lipophilic compounds present in the apolar fractions of the extract. However, the composition of the apolar fractions could not be entirely elucidated. In the freshwater microcrustacean *Thamnocephalus platyurus*, the apolar fractions marked an antagonistic effect, being more acutely toxic to the organisms when compared to the whole extract. We hypothesise that the toxicity observed in these fractions is presumably due to the ability of lipophilic compounds to more easily permeate across membranes (Berry et al. 2007; Roegner et al. 2019)

Individual cyanobacterial strains can co-produce an extensive range of metabolites, making it challenging to attribute toxicity to a particular compound or class of secondary metabolites. We demonstrated that micropeptin K139 can contribute to lethal toxicity (in both model organisms tested) and oedema of the pericardial region in fish, while nostoginin BN741 also contributes to oedema of the pericardial region and lack of hatching but not to acute mortality in zebrafish larvae. Based on reports for other micropeptin and microginin peptides, we hypothesise that the cardiovascular effects observed in zebrafish upon exposure to micropeptin K139 and nostoginin BN741 may be linked to protease inhibition. However, data on protease inhibition caused by cyanopeptides are, in general, are, thus far, based solely on fluorescence enzyme activity assay, and further studies on the *in vivo* inhibition of proteases are required to help understand the mode of action of such compounds. Despite adverse cardiovascular effects arising in zebrafish exposed to microginin-741, this cyanopeptide did not significantly contribute to the observed mortality in zebrafish or the fairy shrimp. Overall, the co-production of metabolites

with known toxins emphasised the need to consider multiple stressors during cyanobacterial blooms and further research on mixture toxicity and the environmental fate of these compounds is recommended.

In the context of cyanobacterial blooms, fish face not only multiple toxic metabolites but also additional anthropogenic pressures (*e.g.*, lower dissolved oxygen levels, presence of micropollutants) further compromising their performance and survival. Hence, we highlight the need to further study, and develop strategies to prevent cyanobacterial blooms, especially in tropical developing countries where these episodes are a long-lasting environmental and public health issue.

6 REFERENCES

AZEVEDO, S. M. F. O. et al. Human intoxication by microcystins during renal dialysis treatment in Caruaru - Brazil. **Toxicology**, v. 181–182, p. 441–446, 2002.

BAUMANN, H. I.; JU, F. Inter-annual stability of oligopeptide patterns of *Planktothrix rubescens* blooms and mass mortality of Daphnia in Lake Hallwilersee. v. 38, p. 350–359, 2008.

BERGEY, D. H.; HOLT, J. G. **Bergey's Manual of Determinative Bacteriology**. [s.l.] Williams & Wilkins, 1994.

BEVERSDORF, L. J. et al. Variable cyanobacterial toxin and metabolite profiles across six eutrophic lakes of differing physiochemical characteristics. **Toxins**, v. 9, n. 2, 2017.

BEVERSDORF, L. J. et al. Analysis of cyanobacterial metabolites in surface and raw drinking waters reveals more than microcystin. **Water Research**, v. 140, p. 280–290, 2018.

BLAGOJEVI, D. et al. Evaluation of cyanobacterial toxicity using different biotests and protein phosphatase inhibition assay. p. 49220–49231, 2021.

BLAGOJEVIĆ, D. et al. Evaluation of cyanobacterial toxicity using different biotests and protein phosphatase inhibition assay. **Environmental Science and Pollution Research**, v. 28, n. 35, p. 49220–49231, 2021.

BOBER, B.; BIALCZYK, J. Determination of the toxicity of the freshwater cyanobacterium *Woronichinia naegeliana* (Unger) Elenkin. **Journal of Applied Phycology**, v. 29, n. 3, p. 1355–1362, 3 jun. 2017.

BURJA, A. M. et al. Marine cyanobacteria - A prolific source of natural products. **Tetrahedron**, v. 57, n. 46, p. 9347–9377, 2001.

CALIJURI, M. C.; SANTOS, A. C. A.; ALVES, M. S. A. **Cianobactérias e cianotoxinas em águas continentais**. [s.l.] RiMa, 2006.

CARMICHAEL, W. W. Cyanobacteria secondary metabolites- the cyanotoxins. p. 445–459, 1992.

CARNEIRO, R. L. et al. Co-occurrence of microcystin and microginin congeners in Brazilian strains of *Microcystis sp*. **FEMS Microbiology Ecology**, v. 82, n. 3, p. 692–702, 2012.

CHORUS, I.; BARTRAM, J. **Toxic Cyanobacteria in Water : A guide to their public health consequences , monitoring and management**. London, UK: E & FN Spon, on behalf of the World Health Organisation, 1999.

CHORUS, I.; WELKER, M. **Toxic Cyanobacteria in Water - A Guide to Their Public Health Consequences, Monitoring and Management**. Second Edi ed. Geneva, CH: CRC Press, Boca Raton (FL), on behalf of the World Health Organization (WHO), 2021.

CORNEL, M. et al. Cyanopeptolins, new depsipeptides from the cyanobacterium *Microcystis sp.* pcc 7806. **The Journal of Antibiotics**, v. 46, n. 10, p. 1550–1556, 1993.

COSTA, M. et al. Identification of Cyanobacterial Strains with Potential for the Treatment of Obesity-Related Co-Morbidities by Bioactivity, Toxicity Evaluation and Metabolite Profiling. **Marine Drugs**, p. 1–16, 2019.

DI PAOLO, C. et al. Early life exposure to PCB126 results in delayed mortality and growth impairment in the zebrafish larvae. **Aquatic Toxicology**, v. 169, p. 168–178, 2015.

EGLI, C. M. et al. Inhibition of extracellular enzymes exposed to cyanopeptides. **Chimia**, v. 74, n. 3, p. 122–128, 2020.

ERNST, M. et al. Molnetenhancer: Enhanced molecular networks by integrating metabolome mining and annotation tools. **Metabolites**, v. 9, n. 7, 2019.

ESTERHUIZEN-LONDT, M. et al. Toxicity and Toxin Composition of *Microcystis aeruginosa* from Wangsong Reservoir. **Toxicology and Environmental Health Sciences**, v. 10, n. 3, p. 179–185, 2018.

FALCONER, I. R. An Overview of Problems Caused by Toxic Blue – Green Algae (Cyanobacteria) in Drinking and Recreational Water. p. 5–12, 1998.

FALTERMANN, S. et al. Molecular effects of the cyanobacterial toxin cyanopeptolin (CP1020) occurring in algal blooms: Global transcriptome analysis in zebrafish embryos. **Aquatic Toxicology**, v. 149, p. 33–39, abr. 2014.

FERNANDES, K. et al. Toxicity of cyanopeptides from two microcystis strains on larval development of *Astyanax altiparanae*. **Toxins**, v. 11, n. 4, p. 2–6, 2019.

FILATOVA, D. et al. Cyanobacteria and their secondary metabolites in three freshwater reservoirs in the United Kingdom. **Environmental Sciences Europe**, v. 33, n. 1, 2021.

FITZGERALD, J. A. et al. Emergence of consistent intra-individual locomotor patterns during zebrafish development. **Scientific Reports**, v. 9, n. 1, p. 1–14, 2019.

FREITAS, P. N. N. et al. Evaluation of the Toxicity of Microcyclamide Produced by *Microcystis aeruginosa* in *Danio rerio* Embryos. **Toxics**, p. 1–11, 2023.

GADEMANN, K. et al. Multiple toxin production in the cyanobacterium *Microcystis*: Isolation of the toxic protease inhibitor cyanopeptolin 1020. **Journal of Natural Products**, v. 73, n. 5, p. 980–984, 2010.

GENUÁRIO, D. B. et al. Cyanobacteria From Brazilian Extreme Environments. **Microbial Diversity in the Genomic Era**, p. 265–284, 2019.

GORHAM, P. R. et al. Isolation and culture of toxic strains of *Anabaena flos-aquae* (Lyngb.) de Bréb. **SIL Proceedings, 1922-2010**, v. 15, n. 2, p. 796–804, 1964.

HO, J. C.; MICHALAK, A. M.; PAHLEVAN, N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. **Nature**, v. 574, n. 7780, p. 667–670, 2019.

HOICZYK, E.; HANSEL, A. Cyanobacterial cell walls: News from an unusual prokaryotic envelope. **Journal of Bacteriology**, v. 182, n. 5, p. 1191–1199, 2000.

HUANG, I. S.; ZIMBA, P. V. Cyanobacterial bioactive metabolites—A review of their chemistry and biology. **Harmful Algae**, v. 83, n. February, p. 42–94, 2019.

HUISMAN, J. et al. Cyanobacterial blooms. **Nature Reviews Microbiology**, v. 16, n. 8, p. 471–483, 2018.

ISHIDA, K. et al. Microginins, zinc metalloproteases inhibitors cyanobacterium *Microcystis aeruginosa*. **Tetrahedron**, v. 56, n. 44, p. 8643–8656, 2000.

ISO. **ISO 7346-3:1996 - Water quality — Determination of the acute lethal toxicity of substances to a freshwater fish [Brachydanio rerio Hamilton-Buchanan (Teleostei, Cyprinidae)] — Part 3: Flow-through method**. [s.l.] International Standards, 1996.

JACINAVICIUS, F. R. et al. Effect of ultraviolet radiation on the metabolomic profiles of potentially toxic cyanobacteria. **FEMS Microbiology Ecology**, v. 97, n. 1, p. 1–16, 2021.

JANSSEN, E. M. L. Cyanobacterial peptides beyond microcystins – A review on co-occurrence, toxicity, and challenges for risk assessment. **Water Research**, v. 151, p. 488–499, 2019.

JONES, M. R. et al. CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria. **Water Research**, v. 196, p. 117017, 2021.

JONES, M. R.; JANSSEN, E. M. L. Quantification of Multi-class Cyanopeptides in Swiss Lakes with Automated Extraction, Enrichment and Analysis by Online-SPE HPLC-HRMS/MS. **Chimia**, v. 76, n. 1–2, p. 133, 2022.

KEIL, C. et al. Toxicity and microcystin content of extracts from a *Planktothrix* bloom and two laboratory strains. **Water Research**, v. 36, n. 8, p. 2133–2139, 2002.

KIMMEL, C. B. et al. Stages of embryonic development of the zebrafish. **Developmental Dynamics**, v. 203, n. 3, p. 253–310, 1995.

KINOSHITA, K. et al. Microginins screening in cyanobacteria by LC-MS. **Química Nova**. v. 13, p. 1385-1392, 2020.

KÖCHER, S. et al. From dolastatin 13 to cyanopeptolins, micropeptins, and lyngbyastatins: The chemical biology of Ahp-cyclodepsipeptides. **Natural Product Reports**, v. 37, n. 2, p. 163–174, 2020.

KOHLER, S. A.; PARKER, M. O.; FORD, A. T. High-throughput screening of psychotropic compounds: Impacts on swimming behaviours in artemia franciscana. **Toxics**, v. 9, n. 3, p. 1–20, 2021.

KORPINEN, S.; KARJALAINEN, M.; VIITASALO, M. Effects of cyanobacteria on survival and reproduction of the littoral crustacean *Gammarus zaddachi* (Amphipoda). **Hydrobiologia**, v. 559, n. 1, p. 285–295, 2006.

LAUGHINGHOUSE, H. D. et al. Biomonitoring genotoxicity and cytotoxicity of *Microcystis aeruginosa* (Chroococcales, Cyanobacteria) using the *Allium cepa* test. **Science of the Total Environment**, v. 432, p. 180–188, 2012.

LENZ, K. A.; MILLER, T. R.; MA, H. Anabaenopeptins and cyanopeptolins induce systemic toxicity effects in a model organism the nematode Caenorhabditis elegans. **Chemosphere**, v. 214, n. August 2014, p. 60–69, 2019.

LI, H. et al. Comparative toxicological effects of planktonic *Microcystis* and benthic *Oscillatoria* on zebrafish embryonic development : Implications for cyanobacteria risk assessment. **Environmental Pollution**, v. 274, p. 115852, 2021.

LOPES, G.; SILVA, M.; VASCONCELOS, V. **The Pharmacological Potential of Cyanobacteria**. [s.l.] Elsevier, 2022.

LYDON, C. A. et al. Identification of apparently neurotoxic metabolites from assemblages of marine filamentous cyanobacteria associated with the intoxication of captive bottlenose dolphins (Tursiops truncatus) in the Florida Keys. **Chemosphere**, v. 288, n. P1, p. 132423, 2022.

MALBROUCK, C.; KESTEMONT, P. Effects of Microcystins on fish. **Environmental Toxicology and Chemistry**, v. 25, n. 1, p. 72–86, 2006.

MARŠÁLEK, B.; BLÁHA, L. Comparison of 17 biotests for detection of cyanobacterial toxicity. **Environmental Toxicology**, v. 19, n. 4, p. 310–317, 2004.

MARTINS, J. et al. Peptide diversity in strains of the cyanobacterium *Microcystis aeruginosa* isolated from Portuguese water supplies. **Applied Microbiology and Biotechnology**, v. 82, n. 5, p. 951–961, 2009.

MAY, L. et al. Getting into hot water : Water quality in tropical lakes in relation to their utilisation. **IOP Conf. Series: Earth and Environmental Science**, v. 789, 2021.

MAZUR-MARZEC, H. et al. *Nodularia spumigena* peptides-accumulation and effect on aquatic invertebrates. **Toxins**, v. 7, n. 11, p. 4404–4420, 2015.

MEREL, S. et al. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. **Environment International**, v. 59, p. 303–327, 2013.

MERILUOTO, J.; SPOOF, L.; CODD, G. A. **Handbook of cyanobacterial monitoring and cyanotoxin analysis**. West Sussex - UK: John Wiley & Sons, 2018. v. 410

MONCHAMP, M. et al. Homogenization of lake cyanobacterial communities over a century of climate change and eutrophication. **Nature Ecology & Evolution**, v. 2, n. February, 2018.

NABOUT, J. C. et al. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. **Biodiversity and Conservation**, v. 22, n. 12, p. 2907–2918, 2013.

NATUMI, R.; DIEZIGER, C.; JANSSEN, E. M. L. Cyanobacterial Toxins and Cyanopeptide Transformation Kinetics by Singlet Oxygen and pH-Dependence in Sunlit Surface Waters. **Environmental Science and Technology**, v. 55, n. 22, p. 15196–15205, 2021.

NATUMI, R.; JANSSEN, E. M. L. Cyanopeptide Co-Production Dynamics beyond Mirocystins and Effects of Growth Stages and Nutrient Availability. **Environmental Science and Technology**, v. 54, n. 10, p. 6063–6072, 19 maio 2020.

NATUMI, R.; MARCOTULLIO, S.; JANSSEN, E. M. L. Phototransformation kinetics of cyanobacterial toxins and secondary metabolites in surface waters. **Environmental Sciences Europe**, v. 33, n. 1, 2021.

OBEREMM, A. et al. Effects of cyanobacterial toxins and aqueous crude extracts of cyanobacteria on the development of fish and amphibians. **Environmental Toxicology**, v. 14, n. 1, p. 77–88, 1999.

OBEREMM, A.; FASTNER, J.; STEINBERG, C. E. W. Effects of microcystin-LR and cyanobacterial crude extracts on embryo-larval development of zebrafish (*Danio rerio*). **Water Research**, v. 31, n. 11, p. 2918–2921, nov. 1997.

OECD. Test No. 236: Fish Embryo Acute Toxicity (FET) Test. n. July, p. 1–18, 2013.

PAERL, H. W.; HUISMAN, J. Climate: Blooms like it hot. **Science**, v. 320, n. 5872, p. 57–58, 2008.

PAIVA, F. C. R. et al. Identification, in vitro testing and molecular docking studies of microginins' mechanism of angiotensin-converting enzyme inhibition. **Molecules**, v. 22, n. 12, p. 1–10, 2017.

PAWLIK-SKOWROŃSKA, B.; BOWNIK, A. Synergistic toxicity of some cyanobacterial oligopeptides to physiological activities of *Daphnia magna* (Crustacea). **Toxicon**, v. 206, n. December 2021, p. 74–84,

2022.

PAWLIK-SKOWROŃSKA, B.; TOPOROWSKA, M.; MAZUR-MARZEC, H. Effects of secondary metabolites produced by different cyanobacterial populations on the freshwater zooplankters Brachionus calyciflorus and Daphnia pulex. **Environmental Science and Pollution Research**, v. 26, n. 12, p. 11793–11804, 2019.

PORTARIA DE CONSOLIDAÇÃO Nº 5, DE 28 DE SETEMBRO DE 2017. Brasil, Ministério da Saúde, 2017.

RAJA, R. et al. Recent developments in therapeutic applications of Cyanobacteria. **Critical Reviews in Microbiology**, v. 42, n. 3, p. 394–405, 2016.

RATNAYAKE, R. et al. Dolastatin 15 from a Marine Cyanobacterium Suppresses HIF-1 α Mediated Cancer Cell Viability and Vascularization. **ChemBioChem**, p. 2356–2366, 2020.

RESHEF, V.; CARMELI, S. Protease inhibitors from a water bloom of the cyanobacterium *Microcystis aeruginosa*. **Tetrahedron**, v. 57, p. 2885–2894, 2001.

RUTTKIES, C. et al. MetFrag relaunched: Incorporating strategies beyond in silico fragmentation. **Journal of Cheminformatics**, v. 8, n. 1, p. 1–16, 2016.

SANT'ANNA, C. L. et al. Review of toxic species of Cyanobacteria in Brazil. **Algological Studies**, v. 126, n. 1, p. 251–265, 2008.

SANT'ANNA, C. L. et al. Lista de Cyanobacteria do Estado de São Paulo. **Biota Neotropica**, v. 11, n. suppl 1, p. 455–495, 2011.

SANZ, M. et al. 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. **Marine Drugs**, v. 13, n. 6, p. 3892–3919, 2015.

SANZ, M.; SALINAS, R. K.; PINTO, E. Namalides B and C and Spumigins K-N from the Cultured Freshwater Cyanobacterium *Sphaerospermopsis torques-reginae*. **Journal of Natural Products**, v. 80, n. 9, p. 2492–2501, 2017.

SCHERER, M.; BEZOLD, D.; GADEMANN, K. Investigating the Toxicity of the Aeruginosin Chlorosulfopeptides by Chemical Synthesis. **Angewandte Chemie - International Edition**, v. 55, n. 32, p. 9427–9431, 2016.

SCHYMANSKI, E. L. et al. Identifying small molecules via high resolution mass spectrometry: Communicating confidence. **Environmental Science and Technology**, v. 48, n. 4, p. 2097–2098, 2014.

SIEBER, S. et al. Microviridin 1777: A Toxic Chymotrypsin Inhibitor Discovered by a Metabologenomic Approach. **Journal of Natural Products**, v. 83, n. 2, p. 438–446, 2020.

SILVA-STENICO, M. E. et al. Non-ribosomal peptides produced by Brazilian cyanobacterial isolates with antimicrobial activity. **Microbiological Research**, v. 166, n. 3, p. 161–175, 2011.

SILVA-STENICO, M. E. et al. Bioactive cyanopeptides produced by *Sphaerocavum brasiliense* strains (cyanobacteria). **Journal of the Brazilian Chemical Society**, v. 26, n. 10, p. 2088–2096, 2015.

STEWART, A. K. et al. Metabolomics-Guided Discovery of Microginin Peptides from Cultures of the Cyanobacterium *Microcystis aeruginosa*. **Journal of Natural Products**, v. 81, n. 2, p. 349–355, 2018.

STRANGMAN, W. K.; WRIGHT, J. L. C. Microginins 680, 646, and 612 - New chlorinated Ahoa-containing peptides from a strain of cultured *Microcystis aeruginosa*. **Tetrahedron Letters**, v. 57, n. 16, p. 1801–1803, 2016.

TARCZYNSKA, M. et al. Tests for the toxicity assessment of cyanobacterial bloom samples. **Environmental Toxicology**, v. 16, n. 5, p. 383–390, 2001.

VILAR, M. C. P.; DE ARAÚJO-CASTRO, C. M. V. A.; MOURA, A. N. Acute toxicity of *Microcystis spp.* (Cyanobacteria) bloom on *Moina minuta* (Cladocera) in a tropical reservoir, Northeastern Brazil. p. 93–98, 2014.

WALTER, J. M. et al. Ecogenomics and taxonomy of Cyanobacteria phylum. **Frontiers in Microbiology**, v. 8, n. NOV, 2017.

WELKER, M.; VON DÖHREN, H. Cyanobacterial peptides - Nature's own combinatorial biosynthesis. **FEMS Microbiology Reviews**, v. 30, n. 4, p. 530–563, 2006.

YU, Y. et al. Proteomic analysis of zebrafish brain damage induced by *Microcystis aeruginosa* bloom. **Science of the Total Environment**, v. 795, p. 148865, 2021.

ZERVOU, S. K. et al. New microginins from cyanobacteria of Greek freshwaters. **Chemosphere**, v. 248, p. 125961, 2020.

APPENDIX I

CyanoMetDB, a comprehensive database of cyanobacterial metabolites

CyanoMetDB, a comprehensive public database of secondary metabolites from cyanobacteria

Martin R. Jones[a], Ernani Pinto[b], Mariana A. Torres[c], Fabiane Dörr[c], Hanna Mazur-Marzec[d], Karolina Szubert[d], Luciana Tartaglione[e], Carmela Dell'Aversano[e], Christopher O. Miles[f], Daniel G. Beach[f], Pearse McCarron[f], Kaarina Sivonen[g], David P. Fewer[g], Jouni Jokela[g], Elisabeth M.-L. Janssen[a],*

[a] Department of Environmental Chemistry, Swiss Federal Institute of Aquatic Science and Technology (Eawag), 8600 Duebendorf, Switzerland
[b] Centre for Nuclear Energy in Agriculture, University of São Paulo, CEP 13418-260 Piracicaba, SP, Brazil
[c] School of Pharmaceutical Sciences, University of São Paulo, CEP 05508-900, São Paulo - SP, Brazil
[d] Division of Marine Biotechnology, University of Gdansk, Al. Marszałka Piłsudskiego 46, 81-378 Gdynia, Poland
[e] Department of Pharmacy, School of Medicine and Surgery, University of Napoli Federico II, Via D. Montesano 49, 80131 Napoli, Italy
[f] Biotoxin Metrology, National Research Council Canada, 1411 Oxford Street, Nova Scotia, Halifax B3H 3Z1, Canada
[g] Department of Microbiology, University of Helsinki, 00014 Helsinki, Finland

ARTICLE INFO

ABSTRACT

Harmful cyanobacterial blooms, which frequently contain toxic secondary metabolites, are reported in aquatic environments around the world. More than two thousand cyanobacterial secondary metabolites have been reported from diverse sources over the past fifty years. A comprehensive, publically-accessible database detailing these secondary metabolites would facilitate research into their occurrence, functions and toxicological risks. To address this need we created CyanoMetDB, a highly curated, flat-file, openly-accessible database of cyanobacterial secondary metabolites collated from 850 peer-reviewed articles published between 1967 and 2020. CyanoMetDB contains 2010 cyanobacterial metabolites and 99 structurally related compounds. This has nearly doubled the number of entries with complete literature metadata and structural composition information compared to previously available open access databases. The dataset includes microcytsins, cyanopeptolins, other depsipeptides, anabaenopeptins, microginins, aeruginosins, cyclamides, cryptophycins, saxitoxins, spumigins, microviridins, and anatoxins among other metabolite classes. A comprehensive database dedicated to cyanobacterial secondary metabolites facilitates: (1) the detection and dereplication of known cyanobacterial toxins and secondary metabolites; (2) the identification of novel natural products from cyanobacteria; (3) research on biosynthesis of cyanobacterial secondary metabolites, including substructure searches; and (4) the investigation of their abundance, persistence, and toxicity in natural environments.

1. Introduction

Cyanobacteria inhabit diverse freshwater, marine and terrestrial environments across the globe and can survive under extreme irradiation, temperature, pH, or salinity conditions. When cyanobacteria proliferate rapidly to form harmful blooms, they can produce high amounts of unique and bioactive secondary metabolites (Baumann and Juttner, 2008; Ferranti et al., 2013; Gkelis et al., 2015; Grabowska et al., 2014; Jancula et al., 2014; Kurmayer et al., 2011; Lopes et al., 2012; Mazur-Marzec et al., 2013). The early assumption that secondary metabolite production evolved in ecologically disadvantaged species experiencing high selection pressure was later underscored by molecular analysis of the genes that control their biosynthesis (Hartmann, 2008).

Secondary metabolites from cyanobacteria can fulfill diverse roles, including defense against other organisms (antibiotics, herbicides, fungicides etc.), as metal transporting agents; as facilitators of symbiosis, as photoprotectants, antioxidants, differentiation effectors or allelochemicals in signaling, among other functions (Demain A. L. and Fang A., 2001; Pereira Daniel and Giani, 2014; Sharif et al., 2008). The potent bioactivities of secondary metabo-

* Corresponding author.
E-mail address: elisabeth.janssen@eawag.ch (E.M.-L. Janssen).

Table 1

Publicly available suspect-lists with total number of entries, number of microcystins, available molecular formulae, primary references for the structure elucidation and structural code (e.g., SMILES code).

Open Access List	Entries	Microcystins	Mol. formulae & primary references	Structural codes (e.g., SMILES)
CyanoMetMass[1]	852	35	499	0
The Natural Products Atlas[2]	1006	27	1006	1006
Microcystins_Miles[3]	296	286	286	0
CyanoMetDB[4]	2010	310	2004	2009 (1691 isomeric SIMLES)

[1] LeManach et al., 2019; [2] van Santen et al., 2019, version 2019_12; Status of The Natural Products Atlas in April 2020 before updating with compounds from CyanoMetDB; [3] Bouaicha et al., 2019; [4] This study; CyanoMetDB contains 2010 secondary metabolites identified in cyanobacteria as well as 99 additional entries: 50 structurally related compounds that are semi-synthetic and synthetic; 5 metabolites that have been identified upon bioaccumulation in other organisms feeding on cyanobacteria, 3 common oxidation products, 341 metabolites that have been identified in other organisms and are structurally related to metabolites from cyanobacteria.

lites have long been recognized, making them lead compounds for the development of drugs (Harvey et al., 2015; Shen, 2015). Pharmaceutical research actively explores their use, e.g., to fight cancers, cardiac and autoimmune disorders or infectious diseases. A number of cyanobacterial secondary metabolites show antimicrobial or antifungal activity (Swain et al., 2017). Despite the recognition of their bioactivity, little is known about the potential human and ecotoxicological risks posed by exposure to cyanobacteria and their less-studied secondary metabolites. Epidemiological studies indicate human health effects and potential cancer development from acute and chronic exposure to cyanopeptides, and other work offers clues regarding the toxicity of less-studied cyanopeptides (Janssen, 2019; Liu et al., 2017; Svircev et al., 2017). Various countries have put forward drinking water guidelines for one toxic metabolite, microcystin-LR (Ibelings et al., 2014), for which the World Health Organization (WHO) proposed a threshold concentration of 1 μg L^{-1} for free and cell-bound exposure (WHO, 2004). Recent updates of these guidelines now include thresholds for a total microcystin content and short-term exposure thresholds of 12 μg L^{-1} for microcystins as well as threshold values for cylindrospermopsin, anatoxin-a, and saxitoxins with 3, 30, and 3 μg L^{-1} in drinking water, respectively (WHO, 2020a; b; c; d). Despite recent progress in establishing guidelines for these compounds, data on the occurrence, fate, transformation processes, and toxicity of most cyanobacterial secondary metabolites is lacking, and improved high-throughput analytical and effect-based methods are needed to overcome this information gap.

Two of the major challenges associated with studying cyanobacterial secondary metabolites are: 1) availability of analytical methods capable of detecting and identifying a broad range of compounds, and 2) a publically available list of known metabolites including chemical structure information. Research into analytical and toxicological methods for cyanobacterial metabolites relies on a comprehensive understanding of their structures. One obstacle that contributes to this is the lack of a bioinformatics platform that the cyanobacterial research community collectively supports. While information from commercial databases of secondary metabolites is only accessible to paying customers (e.g., Antibase, MarinLit, The Dictionary of Natural Products), several open access databases exist but are often limited in terms of the number of cyanobacterial metabolites or parameters listed (e.g., ALGAL-TOX List, NORINE database, Handbook of Marine Natural Products). Recently, a new database unified 411,621 entries of natural products from various organisms using stereochemistry-free InChI key information from 50 open and accessible databases, termed CO-CONUT: COlleCtion of Open Natural ProdUcTs (Sorokina and Steinbeck, 2020). Key open-access databases regarding (cyano)bacterial metabolites are listed in Table 1. The "Cyanomet mass" list by LeManach et al. (2019) contains 852 entries, of which 35 belong to the class of microcystins and nearly 500 compounds are listed with complete molecular formulae and literature references,

but no further structural information is given (Le Manach et al., 2019). In 2017, the Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis was published including a list of 248 microcystins and 10 nodularins (Spoof and Catherine, 2017). Today, the most comprehensive list of microcystins and nodularins was recently updated (2019) to include 286 microcystin and 10 nodularin variants with molecular formulae, references and a systematic naming implying the structural compositions but no structural codes (Bouaicha et al., 2019; Miles and Stirling, 2019). The Natural Products Atlas presents a repository maintained and actively updated by the Simon Fraser University in Canada and covers all microbially-derived natural products. It includes compounds published in peer-reviewed primary literature and contained 1006 cyanobacterial metabolites, including structural identifiers (isomeric SMILES codes), as of December 2019 (van Santen et al., 2019).

Here, we describe a database of cyanobacterial secondary metabolites, termed CyanoMetDB, containing 2010 unique entries. CyanoMetDB has been compiled from existing databases and in-house libraries, as well as through manual curation of available peer-reviewed primary literature published between 1967 and 2020. We present an overview of the structure of this database and explain the methodology used in its compilation and curation. We summarize areas of research that can benefit from this openly-accessible resource, including supporting compound identification using liquid chromatography coupled with mass spectrometry, analysis of biosynthetic variations, bioactivity, environmental behavior and, finally, we provide future perspectives on the database. Herein we supply the CyanoMetDB and literature metadata as (i) separate flat-files, and (ii) converted spreadsheet format files. The current and future versions of CyanoMetDB are available on Zenodo (Jones et al., 2021) and the NORMAN Suspect List Exchange (https://www.norman-network.com/nds/SLE/). We recommend citing the repository on Zenodo along with this article when CyanoMetDB content is used. This work was initiated at the 11th International Conference of Toxic Cyanobacteria (ICTC, 2019, Krakow, Poland) and we welcome participation of the wider research community in future editions of CyanoMetDB.

2. Materials and methods

2.1. Data sources and curation procedure

CyanoMetDB was established as a consolidation of multiple, disparate, sources of information pertaining to cyanobacterial secondary metabolites, including in-house libraries of the CyanoMetDB curation-team members and various open-access databases (Table 1). CyanoMetDB was then extended to include additional cyanobacterial secondary metabolites reported in the scientific literature. For each compound, the primary literature meta-

data was manually verified and where required, corrected. The sample type from which a compound was extracted and identified (e.g., genus/species/strain of cyanobacterium, or field sample), as well as whether nuclear magnetic resonance spectroscopy was used for its structure elucidation, were recorded. Furthermore, the 2D-chemical structure of each compound was manually drawn (ChemDraw, ChemDraw Professional, ACD/ChemSketch) based on the information provided in the primary references, from which structural identifiers were generated, including: simplified molecular input line entry system (SMILES) string, International Union of Pure and Applied Chemistry (IUPAC) International Chemical Identifier (InChI), and InChIkey and IUPAC name. In some instances, this led to the correction of structures originally reported in one of the consolidated data sources, e.g., aeruginosin 101 and aeruginosin 98C both contain a D-allo-Ile in their structure, but were previously misreported as L-allo-Ile derivatives in the primary literature (Ishida et al., 1999). Entries extracted from PubChem were carefully checked to verify the structure from the primary literature. Where discrepancies were observed between structures reported in the primary literature and those found in PubChem, we report the information from the primary literature in CyanoMetDB. In the above-mentioned example, aeruginosin 101 was misreported as its L-allo-Ile derivative whereas aeruginosin 98C was correctly reported in PubChem (accessed 08. September 2020). It is useful to highlight that some issues emerged, for example, with the representation of olefin stereochemistry when a compound's structure was drawn using an InChI code in ChemDraw. For this reason, we recommend the use of SMILES codes from CyanoMetDB for accurate representation of a compound's 2D structure.

After the initial compilation and expansion of CyanoMetDB, the accuracy of data in the database was confirmed through multiple rounds of database integrity checks. In each round, members of the CyanoMetDB curation team received sub-sets of CyanoMetDB compounds, whereupon they evaluated and if necessary, corrected the assignment of the primary (and secondary) literature sources and chemical structural descriptors (SMILES, InChI, InChIKey and IUPAC name).

2.2. CyanoMetDB structure

CyanoMetDB is presented as a flat-file database comprising the following core fields: compound identifier (key), compound name, compound class, molecular formula, molecular weight, monoisotopic mass, primary reference (in which the compound was first reported), SMILES string, InChI string, InChIKey and IUPAC name. Together, the SMILES, InChI, InChIKey and IUPAC name fields serve as textual identifiers for each compound (Heller et al., 2015).

For each entry in CyanoMetDB, the following additional fields were also optionally populated to clarify aspects of a compound's identification. The field "Nuclear magnetic resonance spectroscopy (NMR) used" indicates whether nuclear magnetic resonance spectroscopy was used to confirm a compound's structure, or (partial) relative stereochemistry. The field "secondary reference", if populated, provides a citation to an article in which: 1) a compound was first reported in a cyanobacterial species, after having first been reported (in the primary reference) in a non-cyanobacterial species; or 2) the structural annotation information reported in the primary reference has since been clarified or refined. The "genus", "species" and "strain" fields, collectively, provide an overview of the sample type in which a compound was identified. The field "field sample" includes information for instances where a compound was identified in samples from bloom material comprising one, or multiple, (potentially undefined) cyanobacterial species. Finally, the "Notes" field provides further comments of interest concerning a compound's origin and structural annotation. Table 2 provides a detailed description and illustrative examples for each data field in CyanoMetDB. For a subset of 1097 compounds, we also provide the structural representation as a building block string with abbreviated monomers (e.g., three-letter amino acid codes).

For all entries in CyanoMetDB we provide, as a minimum, the canonical SMILES code describing the connectivity of each atom in the compound's planar structure. An isomeric SMILES code is provided when a compound's stereochemistry was clearly reported in the primary or secondary reference(s), or where evidence from related compounds or known biosynthetic pathways provides evi-
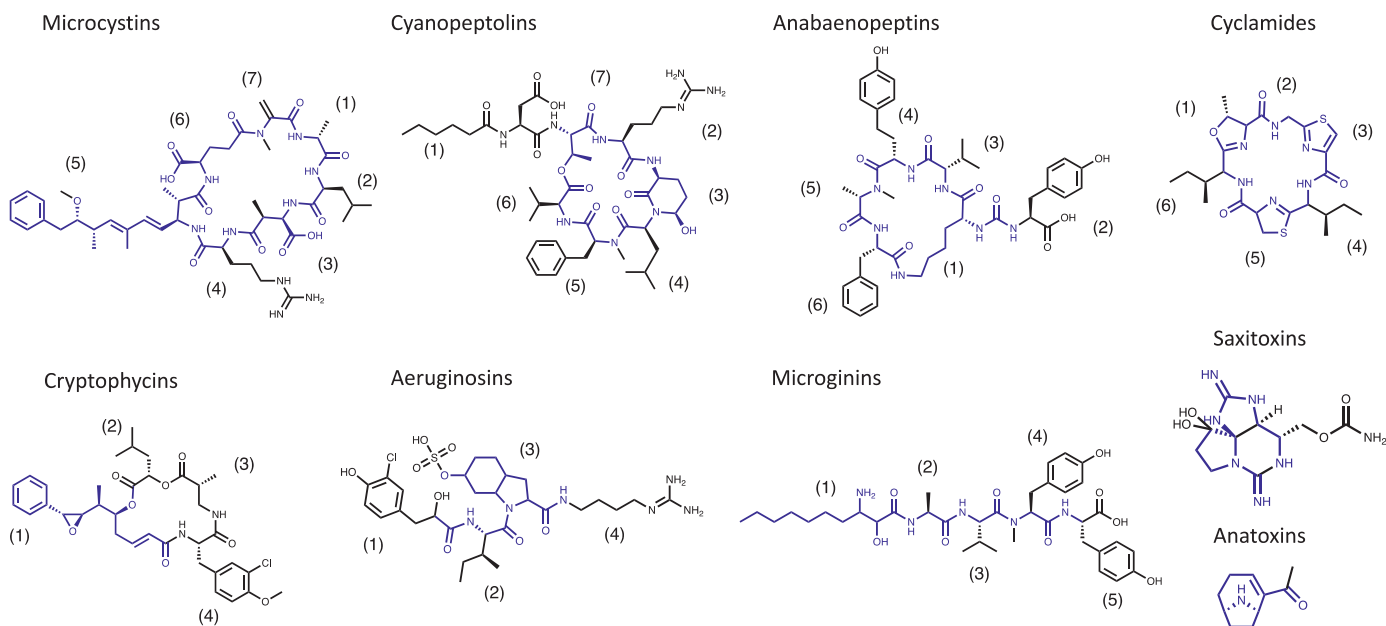
Fig. 1. Representative compounds for nine families of cyanobacterial secondary metabolites: microcystin-LR, cyanopeptolin A, anabaenopeptin A, aerucyclamide A, cryptophycin 1, aeruginosin 98A, microginin 713, saxitoxin, and anatoxin-a. Core-structures that are shared among most variants in the respective class are marked in blue. Numbers in brackets indicate subunit residue numbers. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

dence for probable stereochemical configurations. For example, in all microcystins characterized to date by multiple and complimentary analytical techniques, the monomers identified in positions 2 and 4 of the cyclic structure have always been found to be L-amino acids, while in positions 1, 3 and 6, D-amino acids are always reported. Consequently, the microcystins in CyanoMetDB identified solely by LC–MS have been assumed to retain this stereochemistry. The "Notes" column provides additional details as to whether uncertainty remains regarding some aspect of a compound's structure.

Compounds in CyanoMetDB are classified into major metabolite classes based on conserved molecular substructures, as outlined in previous studies (Fig. 1; Welker and von Döhren, 2006). For example: microcystins are heptapeptides with a characteristic Adda moiety (i.e., (2S,3S,4E,6E,8S,9S)-3-amino-9–methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid) or derivatives thereof; cyanopeptolins contain a β-lactone ring and a characteristic Ahp (3-amino-6–hydroxy-2-piperidone) moiety; anabaenopeptins contain a ureido bond connecting the primary amine of lysine with the primary amine of the neighboring amino acid, and lysine's ε-amine with the carboxyl group of the C-terminal amino acid to form a urea moiety and five-membered peptide ring; aeruginosins contain a Choi (2-carboxy-6-hydroxyoctahydroindole) moiety and its 3–4 substituents; and microginins contain the characteristic Ahda moiety (3-amino-2-hydroxydecanoic acid, or -octanoic acid) (Fig. 1). We conducted substructure searches to verify the class assignments and to identify other structural motifs (ChemAxon JChem Base software 20.4.0, 2020).

3. Results and discussion

To date, CyanoMetDB includes 2010 cyanobacterial secondary metabolites with complete literature and structural information, and some additional entries: 50 semi-synthetic and synthetic entries presenting related compounds that have only been chemically derived or where cyanobacterial metabolites have been chemically modified; 41 metabolites that are structurally related to metabolites from cyanobacteria but so far, have only been identified in other organisms, and; 5 metabolites that have been identified in other organisms that were feeding on cyanobacteria.

The earliest entry in CyanoMetDB was published in 1967 for the chromophore phycoerythrobilin, followed by more than 115 additional compounds by the end of 1990. There was a four-fold increase in the number of reported cyanobacterial metabolites between 1990 and 2000. This rapid increase in the number of reported secondary metabolites during the 1990s was, in part, likely associated with the realization that microcystins pose significant hepatotoxic risks to humans. This led to MC-LR being included in the WHO water quality guidelines (WHO, 2004), prompting significant research on cyanobacteria. Nearly one thousand cyanobacterial compounds were identified by the year 2010, and a further one thousand compounds have been reported in the subsequent decade. The increasing incidences of cyanobacterial blooms and availability of advanced analytical instrumentation (e.g., high resolution mass spectrometry (HRMS) and high-field NMR spectroscopy with cryogenic probes), have contributed to discoveries in recent years (Fig. 2A inset).

3.1. Publication trends

Publication trends in the field suggest that the discovery of cyanobacterial metabolites has not yet reached a plateau (Fig. 2A). The question arises: how close are we to identifying the majority of cyanobacterial secondary metabolites? Also, how many of the newly-described compounds are chemical variants of known families and how many describe new families? These are also overarching questions in the general domain of natural product research. Pye and coworkers recently surveyed the chemical space of natural products and concluded that new discoveries mostly relate structurally to previously published compounds and that the "range of scaffolds readily accessible from nature is limited" (Pye et al., 2017). This limitation does not mean that new discoveries are expected to be exhausted, but rather that most new secondary metabolites are likely to share structural similarities to previous reported ones. In CyanoMetDB, 27% of the entries have been identified by mass spectrometry and such data acquisition or pro-
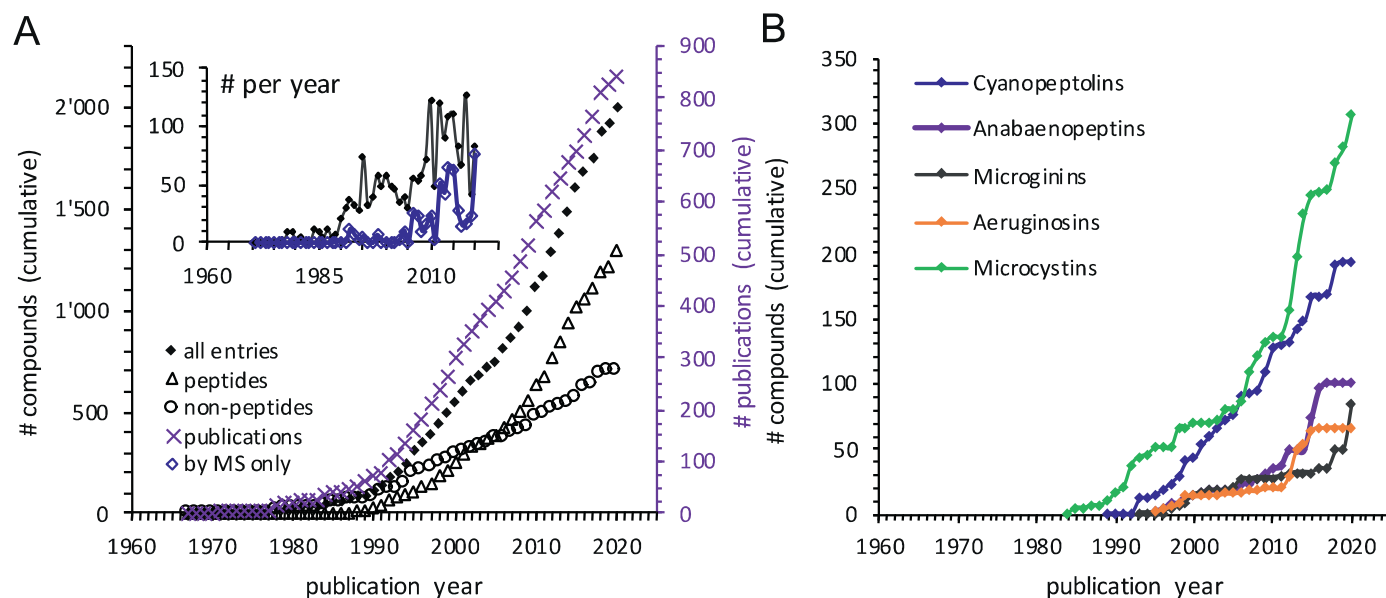
Fig. 2. (A) Cumulative number of: identified secondary metabolites from cyanobacteria (diamonds); peptide-based compounds (open triangles); non-peptide compounds (open circles), and; number of new publications (purple crosses, secondary y-axis). The inset shows the number of new compounds published each year (non-cumulative, black diamonds) and those compounds for which only mass spectrometry (MS) has been used for identification (blue diamonds). (B) Cumulative numbers of major cyanopeptide classes of: cyanopeptolins (blue); anabaenopeptins (purple); microginins (black); aeruginosins (orange), and; microcystins (green), between 1967 and 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

cessing methods are generally optimized for known classes of compound such as cyanopeptides (e.g., microcystins) or low molecular weight molecules (e.g., anatoxins). This increases the probability of identifying additional variants of these classes, rather than compounds with structural differences that could require different LC–MS settings. The discovery of new variants in a known class is still highly relevant because the toxicity is dependent on chemical structure. For example, among 18 microcystin congeners, the IC_{50} values for enzyme inhibition of serine/threonine-protein phosphatases ranged over six orders of magnitude (Altaner et al., 2020). Another reason may be that scientific results are limited by the environments that are predominantly explored, while less focus has thus-far been paid to, for example, cyanobacteria from terrestrial or extreme environments, or that live symbiotically with other organisms (Kaasalainen et al., 2012). Discoveries of natural products from other bacteria started in the 1940s (Pye et al., 2017), whereas cyanobacterial metabolite discoveries began only in the late 1960s. The overall trend is especially driven by discoveries of peptide-based metabolites, and common cyanopeptide classes make up more than one third of all peptidic metabolites in CyanoMetDB (top 5 contributing classes in Fig. 2B). To date, the database contains 2010 compounds including 310 microcystins, 193 cyanopeptolins (also called micropeptins), 211 other depsipeptides, 101 anabaenopeptins, 85 microginins, 67 aeruginosins, 64 cylamides, 38 cryptophycins, 38 saxitoxins, 26 spumigins, 25 microviridins, 16 nodularins, 11 anatoxins, and 5 cylindrospermopsins. Within each class, compounds show high structural similarity, supporting the previous observations that the contribution of novel structural scaffolds is typically low for natural products, including within the pool of cyanobacterial metabolites (Pye et al., 2017).

3.2. Chemical space

CyanoMetDB shows that cyanobacterial secondary metabolites cover a wide range of molecular weights, between 118 and 2708 Da. Of these, 59% are cyclic compounds and 69% are peptides. The distribution of molecular weights of the 2010 compounds is shown in Fig. 3A and 3B, which demonstrate that compounds with at least one peptide bond account for most of the compounds with molecular weights of 900 Da and higher.

Particularly among the peptides, many compounds are present in the 1000–1100 Da range. These peptides can be classified to a large extent into common peptide classes including microcystins,

cyanopeptolins and other cyclic depsipeptides that cover the majority of compounds with molecular weights above 900 Da (Fig. 4). The distribution based on metabolite classes shows that microcystins and cyanopeptolins, in particular, contribute to the high abundance of known metabolites between 1000 and 1100 Da. For the other non-peptide metabolites, there is a particularly high frequency of metabolites with masses between 350 and 500 Da (Fig. 3B), dominated by linear non-peptides (Fig. 4). The molecular weight distribution of more than 28,000 marine natural products has previously been shown to also center around 350 Da and the chemical diversity aligned with biological diversity of the producing species (Blunt et al., 2018). These authors reported that compounds with higher molecular weights that diverged from this trend were predominantly produced by cyanobacteria as well as echinodermata, dinophyta and tracheophyta (mangroves).

Non-peptide secondary metabolites make up only 41% of entries in CyanoMetDB, but this might be an underrepresentation. In general, the non-peptide-based metabolites from cyanobacteria are more difficult to classify because they lack unifying structural features. The structural information (SMILES codes) in the database allows for substructure searching to identify common molecular motifs. For example, of the non-peptide compounds, 15% contain an ester bond and 11% a pyrrolidine ring. Most compounds show unsaturation with 64% carrying at least one aromatic ring, 30% having 1–2 aromatic rings, and 6% having 3–6 aromatic rings. Halogen atoms are present in 35% of the non-peptide compounds (30% Cl, and 5% Br) and 12% contain sulfur. Isomeric compounds, i.e., compounds with the same molecular formulae but different atom connectivity, are common among the cyanobacterial secondary metabolites. For 270 unique molecular formulae, we identified between 2 and 11 isomeric compounds, affecting 706 compounds in the database. Data in Fig. 3C show the distribution of isomeric compounds across molecular weights. Microcystins show the highest number of isomeric compounds (5 to 11) for molecular weights around 1000 Da. The database contains 179 molecular formulae with 2 isomeric compounds, 56 formulae with 3 isomeric compounds and 20 formulae with 4 isomeric compounds. The presence of isomeric compounds makes unequivocal identification of individual compounds challenging for mass spectrometry-based analytical methods and highlights the importance of MS/MS data interpretation and availability of authentic standards for retention time matching.
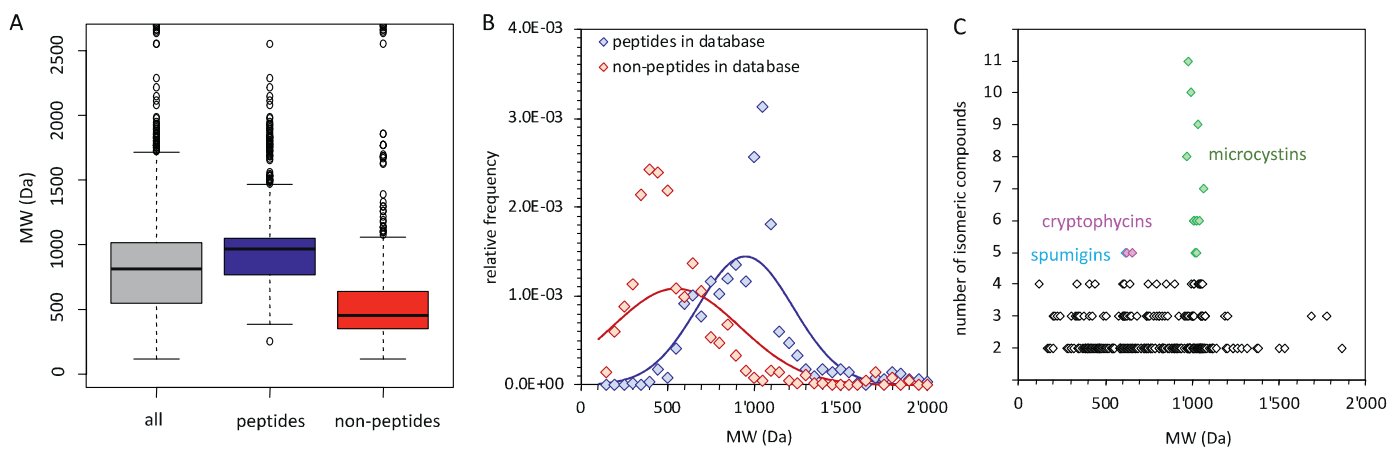
Fig. 3. Molecular weight distribution of cyanobacterial secondary metabolites: (A) Box plot of molecular weights for all metabolites (gray), metabolites containing at least one peptide bond (blue) and all other non-peptide-bond-containing metabolites (red); (B) Relative frequencies (%) of molecular weights for these categories, showing the distribution of the data (50-Da bins) as diamonds and the fitted normal distribution density functions as solid lines; (C) Number of isomeric compounds based on identical molecular formulae. Compound classes were assigned for those with more than 4 isomeric compounds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 4. Distribution of molecular weights for all cyanobacterial secondary metabolites reported in CyanoMetDB, ranging from 100 to more than 2000 Da with 50-Da bin size, showing the contribution of prominent metabolite classes and other linear and cyclic peptides and non-peptides, and the relative contribution to all database entries in the pie chart.
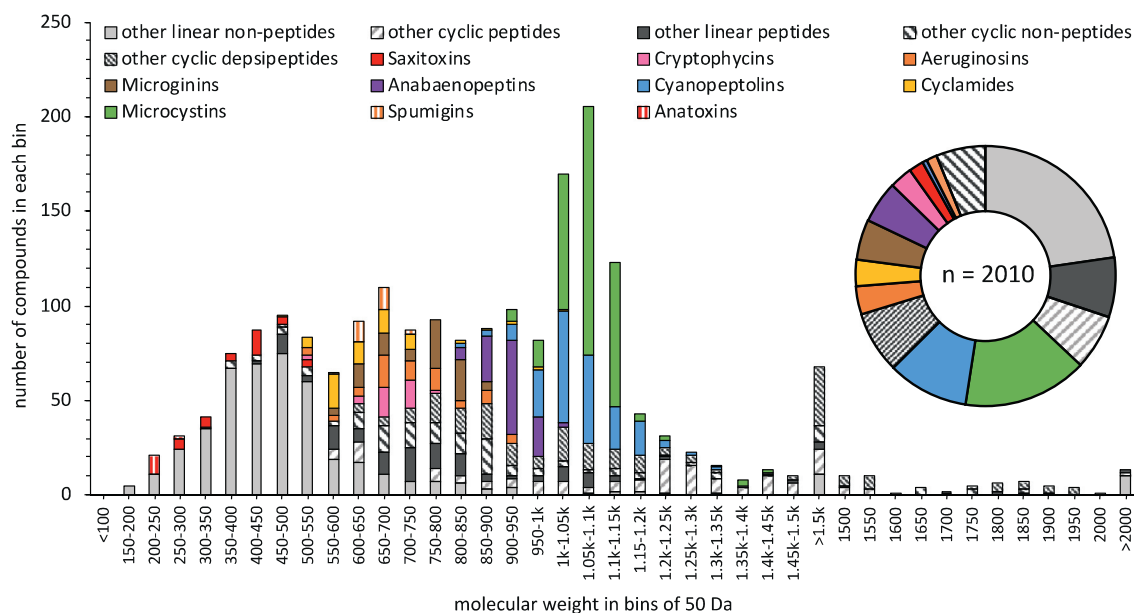
Secondary metabolites from cyanobacteria have mostly been discovered by "top-down" approaches from extraction of biomass and analyses that were guided by chemical motifs (e.g., peptides, molecular weight groups, common mass spectrometric product ions) or non-targeted searches by MS. Within CyanoMetDB, 12% of compounds were first identified from field samples and the remaining compounds were identified from laboratory-grown cultures. In total, more than 50 different cyanobacterial genera were used, considering the genera listed in the associated publications that first identified the structure. Note that the taxonomic classification has subsequently changed for some genera, and additional changes may be introduced in accordance with future knowledge on taxonomy of cyanobacteria. Dominant genera were: *Moorea/Lyngbya, Microcystis, Nostoc, Anabaena/Dolichospermum, Oscillatoria/Planktothrix, Nodularia, Scytonema, Fischerella*, and *Symploca*, in decreasing order of total entries. These genera are not necessarily the main producers of the respective metabolites but were predominantly used for initial structure elucidation.

Compounds from CyanoMetDB that were not previously listed in the Natural Products Atlas but that fulfilrequirements for inclusion have now been added to this online repository of bacterial secondary metabolites; The NP Atlas version 2020_08 (van Santen et al., 2019). Comparing the chemical space of these 1640 natural products from cyanobacteria against more than 28,000 natural products from any bacterium, the NP Atlas demonstrates a relatively even spread of cyanobacterial metabolites supporting their high structural diversity (Fig. 5). The spherical plot uses the molecular composition to position each compound based on C:H ratios (radial value in xy-plane), C:O ratios (angle with z-axis), and C:N ratios (distance from origin) and nodes illustrate the relationship among clusters regarding their structural similarities (van Santen et al., 2019). Here, cyanobacterial secondary metabolites are linked by 239 out of 6800 nodes in The Natural Products Atlas repository.

3.3. Implications for cyanobacterial research

CyanoMetDB holds significant potential as a tool for supporting diverse areas of cyanobacterial research, for instance: aiding the identification of known and novel cyanobacterial metabo-
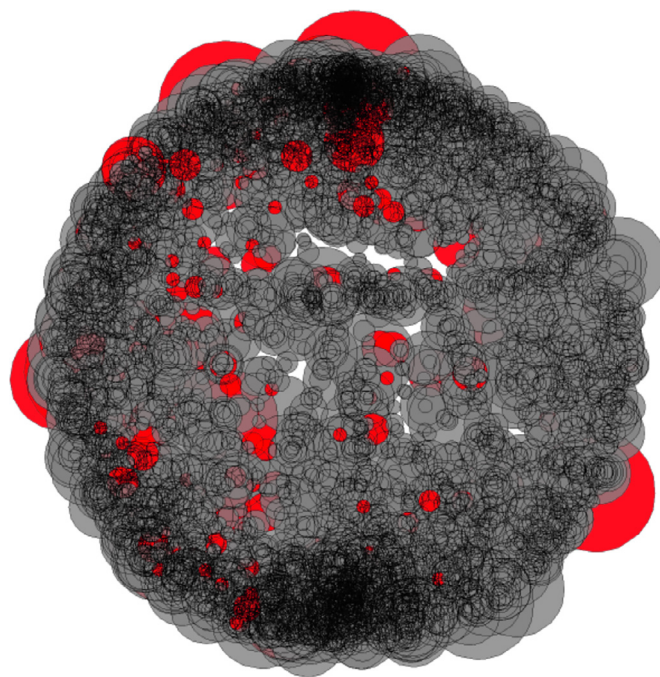
Fig. 5. Global view, highlighting positions of a preliminary set of 1640 secondary metabolites of CyanoMetDB from cyanobacteria (red) among the chemical space of more than 28,000 compounds in the Natural Products Atlas covering 239 out of 6800 nodes that indicate structural similarity (diameter scales with the number of compounds that belong to the node). Data correspond to the first addition of the CyanoMetDB content in July 2020. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

lites; exploring the impact of biosynthetic pathways on cyanobacterial metabolite profiles and dynamics; as a framework around which information may be collated regarding structural characteristics and biological activity of cyanobacterial metabolites; understanding the environmental occurrence, fate and transformation of cyanobacterial secondary metabolites. Below, we explore some of these opportunities in further detail.

3.4. Identification by liquid chromatography-mass spectrometry (LC–MS)

Just as mass spectrometry-based analytical methods have played a significant role in the rapid discovery of many of the cyanobacterial metabolites included in CyanoMetDB, modern LC-MS(/MS) methods are also one of the primary applications of the database. Comprehensive screening of culture- or field-derived samples for cyanobacterial secondary metabolites presents a significant analytical challenge, in part due to the large number of chemically diverse compounds that exist and due to the presence of many isomeric compounds. Moreover, commercially available chemical standards do not exist for the vast majority of these compounds. Together, these factors render targeted triple-quadrupole-based LC–MS/MS methods less effective for analysis of a wider range cyanobacterial metabolites. Instead, LC-high resolution MS (LC–HRMS) methods that are able to efficiently separate and selectively detect ionizable chemicals in complex matrices, with or without fragmentation, hold great potential. A popular strategy by which to analyze compounds by LC–HRMS is termed 'suspect-screening'. This approach involves searching full-scan HRMS data for exact m/z values of interest, which is usually done using commercial or open-source software (e.g., Thermo Trace Finder and Compound Discoverer, or Skyline, respectively) (Gunthardt et al., 2020; Natumi and Janssen, 2020). Detected compounds of interest require further confirmation, ideally through retention time matching against a chemical standard and with detection of characteristic MS/MS fragmentation spectra, applying standardized criteria to specify the confidence of compound identification (Schymanski et al., 2014). Relatively comprehensive full-scan and tandem HRMS data can be acquired using either data-dependent or data-independent HRMS/MS acquisition strategies, which are suitable for suspect-screening and confirmation of abundant cyanobacterial secondary metabolites. CyanoMetDB will greatly expand the power of such LC–HRMS-based suspect-screening workflows, allowing known metabolites to be detected more routinely and thereby avoid "rediscoveries" of known compounds (i.e., dereplication). In the case of data-dependent acquisition, the m/z values in CyanoMetDB can be used as an inclusion list to preferentially trigger MS/MS fragmentation of precursors detected in full-scan with the m/z values of interest.

In addition to suspect-screening, one can also interrogate LC–HRMS datasets in more holistic ways, referred to as untargeted- or non-target analysis. Non-target screening is primarily done using commercial or open-source metabolomics software packages (Hohrenk et al., 2020; Schymanski et al., 2015; Walsh et al., 2019) or with dedicated custom tools (e.g., Global Natural Product Social Molecular Networking, GNPS) (Wang et al., 2016). These workflows typically build upon suspect-screening strategies, starting with the identification of expected compounds and then extending the search space to detect compounds outside the realm of known compounds (as defined by a list or database of known compounds). Here, novel compounds are revealed through similarities in MS/MS data, mass defect analysis, searching for retention time or molecular formulae ranges of interest and by assessing statistical differences between sample groups within a dataset. For example, untargeted workflows have been reported for microcystins that can detect new analogues based on characteristic product ions that are sensitively detected for most known microcystins (Ortiz et al., 2017; Roy-Lachapelle et al., 2019). The comprehensive structural and formula data within CyanoMetDB will enable development of such untargeted workflows and improve the detection of known and novel cyanobacterial secondary metabolites in cultured and environmental samples. Moreover, this information can also be used in combination with *in silico* MS/MS fragmentation tools (e.g., Mass Frontier, mMass, MetFrag) to predict possible structures for MS/MS product ions, i.e., to aid compound annotation (Natumi and Janssen, 2020; Niedermeyer and Strohalm, 2012; Ruttkies et al., 2016; Sheldon et al., 2009). You can now find CyanoMetDB among the databases available in MetFrag (https://msbi.ipb-halle.de/MetFragBeta/) to generate predicted MS/MS data. Recent developments in the area of machine learning-guided MS/MS data interpretation hold great potential in enhancing the accuracy of *in silico*-based compound identification efforts, in particular through their capacity to predict both the types and intensities of product ions likely to be generated from a given structure, e.g., Competitive Fragmentation Model, CFM-ID (Djoumbou-Feunang et al., 2019). Naturally, the effectiveness of such approaches hinge upon the availability and quality of the MS/MS data, which is not always the case in environmental samples with low abundant metabolites or in the presence of complex matrices.

3.5. Biosynthetic analysis

Discovery of secondary metabolites based on genetic information is a promising "bottom-up" approach that can reveal additional, unknown compounds (Moosmann et al., 2018). Recent advances in microbial genomics have greatly improved our understanding of the biochemical mechanisms responsible for the biosynthesis of cyanobacterial natural products (Dittmann et al., 2015). The majority of these natural products are synthesized through secondary metabolic pathways in coordinated enzyme cascades (Dittmann et al., 2015). Approximately 75% of the natural products included in the CyanoMetDB dataset can be assigned to structural families for which biosynthetic pathways have been reported from one or more representatives. Many families of secondary cyanobacterial metabolites exhibit extensive chemical variation but share a structural core that defines the family (Fig. 1). Typically, the biosynthesis of compounds in one family shares a set of conserved enzymes for the synthesis of the defining structural core, but also a set of accessory tailoring enzymes that are not universally conserved (Dittmann et al., 2013, 2015). The biosynthetic logic underlying the biosynthesis of most common cyanobacterial toxins, including microcystins (Tillett et al., 2000), nodularins (Moffitt and Neilan, 2004), saxitoxins (Kellmann et al., 2009), cylindrospermopsins (Mihali et al., 2008) and anatoxins (Mejean et al., 2009) are now well understood. However, the basis for the biosynthesis of specific chemical variants is incomplete in many cases. Numerous molecular ecology methods have been developed to characterize the types of toxin producers in blooms, based on the biosynthetic gene clusters (Dittmann et al., 2013). The compilation of CyanoMetDB has shown that the chemical variation of the major secondary metabolites is more extensive than previously thought. A more complete understanding of the biosynthetic basis for the chemical variation of secondary metabolites from cyanobacteria is necessary to ensure the unbiased detection of their biosynthetic pathways directly from environmental samples.

3.6. Bioactivity and structure

With the fast-growing number of new secondary metabolites from cyanobacteria and the growing awareness of their (sometimes beneficial, and sometimes detrimental) biological activities, the availability of a comprehensive database is essential. Cyanobacterial metabolites present cytotoxic, dermatotoxic, hepatotoxic, neurotoxic, enzyme inhibiting, antimicrobial, antifungal, antiprotozoal, and anti-inflammatory activities that can also be exploited by the pharmaceutical industry to develop new drugs that are potentially beneficial to humans (Huang and Zimba, 2019; Kini et al., 2020; Singh et al., 2011). Future discoveries are likely to share structural

similarities to previously discovered metabolites but may, however, exhibit differing potencies and thus are important to identify. Any modifications to the structure, such as replacement of amino acids by other residues, substitution by methylation, halogenation or oxidation and changes in configuration, can significantly affect the ability of cyanobacterial metabolites to evoke a biological, or toxic, response. For example, in the case of microcystins, biological activity toward protein phosphatases is underpinned by their cyclic structure and the Adda–D-Glu-region, which interacts with the catalytic unit of these enzymes (McLellan and Manderville, 2017). The modification of the less conserved positions 2 and 4 (Fig. 1) also plays a significant role in microcystin bioactivity (Bouaicha et al., 2019; Fontanillo and Kohn, 2018). The selective activity of cyanopeptolins against serine proteases depends on the residue in the Ahp-neighboring position. The Arg–Ahp-containing cyanopeptolin variants are mainly active against trypsin, whereas cyanopeptolins with hydrophobic amino acids (e.g. Phe, Tyr, Ile) show potent activity against chymotrypsin (Yamaki et al., 2005). In the case of cryptophycins, which are cyclic depsipeptides composed of four subunits, their effect on microtubule dynamics is determined by the intact 16-membered macrolide structure, reactive epoxide ring in unit 1, methyl groups in units 1 and 3, *O*-methyl group and chloro–substituent in unit 2, and isobutyl group in unit 4 (units 1–4 in Fig. 1 correspond to units A–D in Golakoti et al., 1996). The interaction of a metabolite with cellular molecules that causes an observable adverse outcome (i.e., toxicodynamics) depends critically on the structure of the metabolite. *In silico* studies, termed virtual screening (target-based or ligand-based) use structural information of compounds deposited in databases to overcome problems related to the limited availability of chemical standards (i.e., pure compounds) and lack of information on their physico-chemical properties (Kirchweger and Rollinger, 2018). In the case of new drug development from cyanopeptides, virtual screening could also help reduce the cost and increase the success rate of the process. Cheminformatics can also assist the discovery of as-yet unknown cyanobacterial metabolites in so-called target-fishing (Brzuzan et al., 2020; Liang et al., 2018; Zhu et al., 2015). Modern techniques of machine learning can assist in predicting the effect of an unknown metabolite after training the model with known compounds, as recently demonstrated for microcystins (Altaner et al., 2020). CyanoMetDB presents a database including structural information for each metabolite that is deposited in a format immediately accessible to software tools. The deposited structures in CyanoMetDB can also be used as templates in these approaches and to design new chemical entities with desirable traits and ligand–target interactions.

3.7. Environmental behavior

As thousands of secondary metabolites have been identified from cyanobacteria, we now face the questions: how persistent are they? How stable are they in surface waters? How do their concentrations change during bloom events? Can they reach water treatment plant intakes and does sufficient abatement occur during treatment? Answers to these questions are needed to quantify the exposure side of the risk assessment equation and to prioritize cyanobacterial metabolites for toxicity testing, monitoring of surface waters, and evaluating removal in engineered water treatment systems. Physico-chemical properties, reactivity with oxidants in surface waters and during water treatment, and biotransformation mechanisms are relevant to predict the behavior of secondary metabolites. With known chemical structures, models can be developed to predict physico-chemical properties from quantitative structure–activity relationships (QSARs). For example, water–octanol partitioning coefficients were predicted for 45 polar plant toxins with three different models (KOWWIN, ACD/Percepta,

and Chemicalize) and results were in close agreement with empirically derived values (Schonsee and Bucheli, 2020). Good correlations, i.e., QSARs, have also been observed among structurally-related organic micropollutants and their reaction with oxidants used in (advanced) water treatment (e.g., O_3, ClO_3, $HOCl$, $HFeO_4^-$) (Lee and von Gunten, 2012). The use of training sets enables calibration of such models and improves their predictive power for properties of unknown compounds. Machine learning can be combined with such modeling to train QSARs, as recently demonstrated for the prediction of pK_a values of more than 6000 organic molecules (Mansouri et al., 2019). Also, biotransformation mechanisms can in part be predicted for compounds with known structures. For example, EnviPath presents an open access tool that uses known transformation rules to propose degradation pathways for organic molecules (Wickert et al., 2016). Such predictions can also assist in the search for transformation products of known metabolites. The structural information in CyanoMetDB allows substructure searching for moieties with known reactivity in biotic and abiotic transformation processes, and the transformation products can be included in QSAR models. QSAR models may be particularly useful within compound classes that share a core structure, but where empirical parameters are only available for some congeners. Such models can be explored to prioritize secondary metabolites from cyanobacteria for further research and monitoring according to their expected toxicity, persistence and mobility in the environment.

4. Conclusion

Cyanobacterial metabolites have now been studied for over 60 years. During this time, thousands of chemical structures have been reported across hundreds of primary research articles, which have only in part been included in curated lists by individual research groups. In this work, with a view to promoting effective analysis and interchange of information about cyanobacterial metabolites, we have manually collated and evaluated these disparate resources (including 850 primary research articles, as of December 2020) to generate a single, unified, database of known cyanobacterial secondary metabolites. Termed CyanoMetDB, this database contains 2010 individual records, each corresponding to a unique cyanobacterial secondary metabolite and associated chemical descriptors: SMILES string, molecular weight, monoisotopic mass, molecular formula, etc. We made CyanoMetDB an open-access resource to ensure that it is accessible to a broad audience, whilst also fulfilling the goal of supporting the analysis and identification of cyanobacterial secondary metabolites in the future. Herein, we supply CyanoMetDB and literature metadata as separate flat-files and converted spreadsheet format files. To this end, the current and future versions of CyanoMetDB are and will be available on Zenodo (Jones et al., 2021) and the NORMAN Suspect List Exchange (https://www.norman-network.com/nds/SLE/). We recommend citingcrecommiting the repository on Zenodo along with this article when CyanoMetDB content is used.

Moving forward, our aim is to continue the work presented herein by: 1) adding missing and newly reported cyanobacterial secondary metabolites to CyanoMetDB as they are published, and 2) using CyanoMetDB as a framework for connecting and collating various data sources associated with cyanobacterial metabolites, e.g., their tandem mass spectrometry product ion spectra, toxicity data, etc. We anticipate that the continued curation of CyanoMetDB will enrich existing and establish new collaborative research efforts, enhance the frequency with which compound annotations are assigned, and aid communication, comparison and interpretation of results.

The approach used to build the CyanoMetDB dataset could also be applied to build related datasets containing toxins and sec-

ondary metabolites from other organisms, e.g., those produced by marine microalgae. Such efforts would be highly beneficial to facilitate monitoring of marine microalgal toxins in the food chain for research and monitoring purposes. Indeed, some toxin classes, such as saxitoxins, were first identified in marine dinoflagellates before their discovery in cyanobacterial samples. Thus, datasets of cyanobacterial and marine microalgal metabolites will partially overlap.

Declaration of Competing Interest

None

Acknowledgments

Supplementary materials

Supplementary material associated with this article can be found in the online version, at doi:10.1016/j.watres.2021.117017.

Appendix

Table 2
Description of the individual input fields of CyanoMetDB.

Input field	Description	Example
Compound identifier	Continuous numbering of compounds in database (CyanoMetDB_xxxx)	CyanoMetDB_0816
Compound name	Most common name for the compound	Micropeptin 90
	Version of the Compound name compatible with chemoinformatics tools (e.g. symbols replaced)	Micropeptin 90
Class of compound	Class the compound belongs to, based on structural similarity; common building block characteristic for this class; classes are *not* selected based on biosynthetic pathways	Cyanopeptolin
Alternative class names	Alternative name(s) for this class	Micropeptin; Anabaenopeptilide
Molecular formula (neutral)	Molecular formula of the neutral species	$C_{42}H_{59}N_9O_{15}S$
Molecular weight	Molecular weight (g/mol)	962.04
Monoisotopic mass	Exact mass obtained using the principal isotopes for the neutral molecule	961.38513
SMILES (isomeric)	SMILES codes with stereoinformation	C[C@H]1[C@@H](C(N[C@@H](C(N[C@@H]2CC[C@@H](N(C2=O)[C@@H](C(N([C@@H](C(NC(C(O1)=O)C(C)C)=O)CC3=CC=C(C=C3)O)C)=O)CC4=CC=CC=C4)O)=O)CCC/N=C(N)\N)=O)NC(C(COS(=O)(O)=O)O)=O
SMILES (canonical)	SMILES codes without stereoinformation	CC1C(NC(C(O)COS(=O)(O)=O)=O)C(NC(CCC/N=C(N)\N)C(NC2CCC(O)N(C(CC3=CC=CC=C3)C(N(C)C(CC4=CC=C(O)C=C4)C(NC(C(C)C)C(O1)=O)=O)=O)C2=O)=O)=O
InChI	IUPAC textual identifier for chemical substances	InChI=1S/C42H59N9O15S/c1–22(2)33–41(61)66–23(3)34(49–37(57)31(53)21–65–67(62,63)64)38(58)46–27(11–8–18–45–42(43)44)35(55)47–28–16–17–32(54)51(39(28)59)30(20–24–9–6–5–7–10–24)40(60)50(4)29(36(56)48–33)19–25–12–14–26(52)15–13–25/h5–7,9–10,12–15,22–23,27–34,52–54H,8,11,16–21H2,1–4H3,(H,46,58)(H,47,55)(H,48,56)(H,49,57)(H4,43,44,45)(H,62,63,64)/t23-,27+,28+,29+,30+,31?,32-,33?,34-/m0/s1
InChI Key	Hashed version of the full IUPAC Chemical Identifier (InChI)	XYLRPTCPMWMQJO-DTZDLPBWSA-N
IUPAC name	International Union of Pure and Applied Chemistry nomenclature naming scheme	3-(((2R,5R,11S,12S,15R,18R,21S)−2-benzyl-15-(3-((diaminomethylene)amino)propyl)−21-hydroxy-5-(4-hydroxybenzyl)−8-isopropyl-4,11-dimethyl-3,6,9,13,16,22-hexaoxo-10-oxa-1,4,7,14,17-pentaazabicyclo[16.3.1]docosan-12-yl)amino)−2-hydroxy-3-oxopropyl hydrogen sulfate
Pubchem CID	PubChem compound identification number	45,269,517
Ref_ID	Continuous numbering of references in database	Ref_42
DOI	Digital Object Identifier of the primary literature for structure identification of the compound	10.1016/0040–4039(95)00,547-P

(continued on next page)

Table 2 (continued)

Input field	Description	Example
Url	Weblink to literature	https://linkinghub.elsevier.com/retrieve/pii/004040399500547P
Reference text	Title; Journal; Vol,; Issue; pages; year; publication type; DOI; author1; authors2; and additional authors	Micropeptin 90, a plasmin and trypsin inhibitor from the blue-green alga *Microcystis aeruginosa* (NIES-90); Tetrahedron Letters; 36; 20; 3535–3538; 1995; journal article; 10.1016/0040–4039(95)00,547-P; Ishida, Keishi; Murakami, Masahiro; Matsuda, Hisashi; Yamaguchi, Katsumi
Building block string	Short version of each subunit for a multi-subunit compound	[(Arg)–(Ahp)–(Phe)–(N-Me-Tyr)–(Val)–(O-Thr)] –SuGA with Ahp = 3-amino–6–hydroxy-2-oxo-1-piperidine; N-MeTyr = *N*-methyltyrosine; SuGA = glyceric acid hydrogen sulfate
Nuclear magnetic resonance spectroscopy (NMR) used	«yes» indicates that the linked literature used NMR spectroscopy to elucidate the structure of the compound; «no» indicates that no NMR data was provided in the listed references	Yes
Genus	Genus of specimen used for extraction and structure elucidation as in listed reference; this does *not* show a list of possible producing genera only that used in the listed primary literature. When field samples were used, it is referred to as "*field* sample"	*Microcystis*
Species	Species of specimen used for extraction and structure elucidation as in listed reference; this does *not* show a list of possible producing species only that used in the listed primary literature	*aeruginosa*
Strain	Identification of strain used	NIES-90
Field sample	Information on field sample if this was used	
Notes	Any additional comment on the entry	stereochemistry not completely resolved

References

Altaner, S., Jaeger, S., Fotler, R., Zemskov, I., Wittmann, V., Schreiber, F., Dietrich, D.R., 2020. Machine learning prediction of cyanobacterial toxin (microcystin) toxicodynamics in humans. ALTEX - Alt. Anim. Exp. 37 (3), 24–36.

Baumann, H.I., Juttner, F., 2008. Inter-annual stability of oligopeptide patterns of *Planktothrix rubescens* blooms and mass mortality of *Daphnia* in Lake Hallwilersee. Limnologica 38 (3–4), 350–359.

Blunt, J.W., Carroll, A.R., Copp, B.R., Davis, R.A., Keyzers, R.A., Prinsep, M.R., 2018. Marine natural products. Nat. Prod. Rep. 35 (1), 8–53.

Bouaicha, N., Miles, C.O., Beach, D.G., Labidi, Z., Djabri, A., Benayache, N.Y., Nguyen-Quang, T., 2019. Structural diversity, characterization and toxicology of microcystins. Toxins (Basel) 11 (12), 714.

Brzuzan, P., Mazur-Marzec, H., Florczyk, M., Fidor, A., Konkel, R., Woźny, M., 2020. Luciferase reporter assay for small-molecule inhibitors of MIR92b-3p function: screening cyanopeptolins produced by *Nostoc* from the Baltic Sea. Toxicol. In Vitro 68, 104951.

Demain, A.L., Fang, A., 2001. The natural functions of secondary metabolites. In: Fichter, A. (Ed.), History of Modern Biotechnology I. Springer, Berlin, Germany, pp. 1–39.

Dittmann, E., Fewer, D.P., Neilan, B.A., 2013. Cyanobacterial toxins: biosynthetic routes and evolutionary roots. FEMS Microbiol. Rev. 37 (1), 23–43.

Dittmann, E., Gugger, M., Sivonen, K., Fewer, D.P., 2015. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. Trends Microbiol. 23 (10), 642–652.

Djoumbou-Feunang, Y., Pon, A., Karu, N., Zheng, J.M., Li, C., Arndt, D., Gautam, M., Allen, F., Wishart, D.S., 2019. CFM-ID 3.0: significantly improved ESI-MS/MS prediction and compound identification. Metabolites 9 (4), 72.

Ferranti, P., Fabbrocino, S., Chiaravalle, E., Bruno, M., Basile, A., Serpe, L., Gallo, P., 2013. Profiling microcystin contamination in a water reservoir by MALDI-TOF and liquid chromatography coupled to Q/TOF tandem mass spectrometry. Food Res. Int. 54 (1), 1321–1330.

Fontanillo, M., Kohn, M., 2018. Microcystins: synthesis and structure–activity relationship studies toward PP1 and PP2A. Bioorg. Med. Chem. 26 (6), 1118–1126.

Gkelis, S., Lanaras, T., Sivonen, K., 2015. Cyanobacterial toxic and bioactive peptides in freshwater bodies of Greece: concentrations, occurrence patterns, and implications for human health. Mar. Drugs 13 (10), 6319–6335.

Golakoti, T., Ogino, J., Heltzel, C.E., LeHusebo, T., Jensen, C.M., Larsen, L.K., Patterson, G.M.L., Moore, R.E., Mooberry, S.L., Corbett, T.H., Valeriote, F.A, 1996. Structure determination, conformational analysis, chemical stability studies, and antitumor evaluation of the cryptophycins. Isolation of 18 new analogs from *Nostoc* sp. strain CSV 224 (vol 117, pg 12030, 1995). J. Am. Chem. Soc. 118 (13) 3323-3323.

Grabowska, M., Kobos, J., Torunska-Sitarz, A., Mazur-Marzec, H., 2014. Non–ribosomal peptides produced by *Planktothrix agardhii* from Siemianowka Dam Reservoir SDR (northeast Poland). Arch. Microbiol. 196 (10), 697–707.

Gunthardt, B.F., Schonsee, C.D., Hollender, J., Hungerbuhler, K., Scheringer, M., Bucheli, T.D., 2020. Is there anybody else out there?"—First insights from a suspect screening for phytotoxins in surface water. Chimia (Aarau) 74 (3), 129–135.

Hartmann, T., 2008. The lost origin of chemical ecology in the late 19th century. Proc. Natl. Acad. Sci. U.S.A. 105 (12), 4541–4546.

Harvey, A.L., Edrada-Ebel, R., Quinn, R.J., 2015. The re-emergence of natural products for drug discovery in the genomics era. Nat. Rev. Drug Discov. 14 (2), 111–129.

Heller, S.R., McNaught, A., Pletnev, I., Stein, S., Tchekhovskoi, D., 2015. InChI, the IUPAC international chemical identifier. J. Cheminf. 7, 23.

Hohrenk, L.L., Itzel, F., Baetz, N., Tuerk, J., Vosough, M., Schmidt, T.C., 2020. Comparison of software tools for liquid chromatography–high-resolution mass spectrometry data processing in nontarget screening of environmental samples. Anal. Chem. 92 (2), 1898–1907.

Huang, I.S., Zimba, P.V., 2019. Cyanobacterial bioactive metabolites—A review of their chemistry and biology. Harmful Algae 86 138-138.

Ibelings, B.W., Backer, L.C., Kardinaal, W.E.A., Chorus, I, 2014. Current approaches to cyanotoxin risk assessment and risk management around the globe. Harmful Algae 40, 63–74.

Ishida, K., Okita, Y., Matsuda, H., Okino, T., Murakami, M., 1999. Aeruginosins, protease inhibitors from the cyanobacterium *Microcystis aeruginosa*. Tetrahedron 55 (36), 10971–10988.

Jancula, D., Strakova, L., Sadilek, J., Marsalek, B., Babica, P., 2014. Survey of cyanobacterial toxins in Czech water reservoirs—The first observation of neurotoxic saxitoxins. Environ. Sci. Pollut. Res. 21 (13), 8006–8015.

Janssen, E.M.L, 2019. Cyanobacterial peptides beyond microcystins—A review on co-occurrence, toxicity, and challenges for risk assessment. Water Res. 151, 488–499.

Jones, M.R., Pinto, E., Torres, M., Dörr, F., Mazur-Marzec, H., Szubert, K., Tartaglione, L., Dell'Aversano, C., Beach, D.G., McCarron, P., Miles, C.O., Sivonen, K., Fewer, D.P., Jokela, J. and Janssen, E.M.-L. (2021). S75 | CyanoMetDB | Comprehensive database of secondary metabolites from cyanobacteria [Data set]. Zenodo. https://zenodo.org/record/4551528

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. U.S.A. 109 (15), 5886–5891.

Kellmann, R., Mihali, T.K., Neilan, B.A., 2009. Identification of a saxitoxin biosynthesis gene with a history of frequent horizontal gene transfers. J. Mol. Evol. 68 (3) 292-292.

Kini, S., Divyashree, M., Mani, M.K., Mamatha, B.S., 2020. Algae and cyanobacteria as a source of novel bioactive compounds for biomedical applications. In: Singh, P.K., Kumar, A., Singh, V.K., Shrivastava, A.K. (Eds.), Advances in Cyanobacterial Biology. Academic Press, London, UK, pp. 173–194.

Kirchweger, B., Rollinger, J., 2018. Virtual screening for the discovery of active principles. In: Cechinel Filho, V. (Ed.), Natural Products: Research and Development, Challenges and Perspectives. Springer, Cham, Switzerland, pp. 333–364.

Kurmayer, R., Schober, E., Tonk, L., Visser, P.M., Christiansen, G., 2011. Spatial divergence in the proportions of genes encoding toxic peptide synthesis among populations of the cyanobacterium *Planktothrix* in European lakes. FEMS Microbiol. Lett. 317 (2), 127–137.

Le Manach, S., Duval, C., Marie, A., Djediat, C., Catherine, A., Edery, M., Bernard, C., Marie, B, 2019. Global metabolomic characterizations of *Microcystis* spp. highlights clonal diversity in natural bloom-forming populations and expands metabolite structural diversity. Front. Microbiol. 10, 791.

Lee, Y., von Gunten, U., 2012. Quantitative structure–activity relationships (QSARs) for the transformation of organic micropollutants during oxidative water treatment. Water Res. 46 (19), 6177–6195.

Liang, T.T., Zhao, Q., He, S., Mu, F.Z., Deng, W., Han, B.N., 2018. Modeling analysis of potential target of dolastatin 16 by computational virtual screening. Chem. Pharm. Bull. 66 (6), 602–607.

Liu, W.Y., Wang, L.Q., Yang, X.H., Zeng, H., Zhang, R.P., Pu, C.W., Zheng, C.F., Tan, Y., Luo, Y., Feng, X.B., Tian, Y.Q., Xiao, G.S., Wang, J., Huang, Y.J., Luo, J.H., Feng, L., Wang, F., Yuan, C.Y., Yao, Y.A., Qiu, Z.Q., Chen, J.A., Wu, L.P., Nong, Q.Q., Lin, H., Shu, W.Q., 2017. Environmental microcystin exposure increases liver injury risk induced by hepatitis B virus combined with aflatoxin: a cross-sectional study in southwest China. Environ. Sci. Technol. 51 (11), 6367–6378.

Lopes, V.R., Ramos, V., Martins, A., Sousa, M., Welker, M., Antunes, A., Vasconcelos, V.M., 2012. Phylogenetic, chemical and morphological diversity of cyanobacteria from Portuguese temperate estuaries. Mar. Environ. Res. 73, 7–16.

Mansouri, K., Cariello, N.F., Korotcov, A., Tkachenko, V., Grulke, C.M., Sprankle, C.S., Allen, D., Casey, W.M., Kleinstreuer, N.C., Williams, A.J., 2019. Open-source QSAR models for pK$_a$ prediction using multiple machine learning approaches. J. Cheminf. 11 (1), 60.

Mazur-Marzec, H., Kaczkowska, M.J., Blaszczyk, A., Akcaalan, R., Spoof, L., Meriluoto, J., 2013. Diversity of peptides produced by Nodularia spumigena from various geographical regions. Mar. Drugs 11, 1–19.

McLellan, N.L., Manderville, R.A., 2017. Toxic mechanisms of microcystins in mammals. Toxicol. Res.-U.K. 6 (4), 391–405.

Mejean, A., Mann, S., Maldiney, T., Vassiliadis, G., Lequin, O., Ploux, O., 2009. Evidence that biosynthesis of the neurotoxic alkaloids anatoxin-a and homoanatoxin-in-a in the cyanobacterium *Oscillatoria* PCC 6506 occurs on a modular polyketide synthase initiated by L-proline. J. Am. Chem. Soc. 131 (22), 7512–7513.

Mihali, T.K., Kellmann, R., Muenchhoff, J., Barrow, K.D., Neilan, B.A., 2008. Characterization of the gene cluster responsible for cylindrospermopsin biosynthesis. Appl. Environ. Microb. 74 (3), 716–722.

Miles, C.O. and Stirling, D. (2019) Toxin Mass List, version 16. Available online: https://www.researchgate.net/publication/337258461_Toxin_mass_list_COM_v160_microcystin_and_nodularin_lists_and_mass_calculators_for_mass_spectrometry_of_microcystins_nodularins_saxitoxins_and_anatoxins. Accessed on 14 April 2020.

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. Microb. 70 (11), 6353–6362.

Moosmann, P., Ueoka, R., Gugger, M., Piel, J., 2018. Aranazoles: extensively chlorinated nonribosomal peptide-polyketide hybrids from the cyanobacterium *Fischerella* sp. PCC 9339. Org. Lett. 20 (17), 5238–5241.

Natumi, R., Janssen, E.M.L, 2020. Cyanopeptide co-production dynamics beyond mirocystins and effects of growth stages and nutrient availability. Environ. Sci. Technol. 54 (10), 6063–6072.

Niedermeyer, T.H.J., Strohalm, M, 2012. mMass as a software tool for the annotation of cyclic peptide tandem mass spectra. PLoS ONE 7 (9), e44913.

Ortiz, X., Korenkova, E., Jobst, K.J., MacPherson, K.A., Reiner, E.J., 2017. A high throughput targeted and non-targeted method for the analysis of microcystins and anatoxin-a using online solid phase extraction coupled to liquid chromatography–quadruple time-of-flight high resolution mass spectrometry. Anal. Bioanal. Chem. 409 (21), 4959–4969.

Pereira Daniel, A., Giani, A., 2014. Cell density-dependent oligopeptide production in cyanobacterial strains. FEMS Microbiol. Ecol. 88 (1), 175–183.

Pye, C.R., Bertin, M.J., Lokey, R.S., Gerwick, W.H., Linington, R.G., 2017. Retrospective analysis of natural products provides insights for future discovery trends. Proc. Natl. Acad. Sci. U.S.A. 114 (22), 5601–5606.

Roy-Lachapelle, A., Solliec, M., Sauvé, S., Gagnon, C., 2019. A data-independent methodology for the structural characterization of microcystins and anabaenopeptins leading to the identification of four new congeners. Toxins (Basel) 11 (11), 619.

Ruttkies, C., Schymanski, E.L., Wolf, S., Hollender, J., Neumann, S., 2016. MetFrag relaunched: incorporating strategies beyond *in silico* fragmentation. J. Cheminf. 8.

Schonsee, C.D., Bucheli, T.D., 2020. Experimental determination of octanol–water partition coefficients of selected natural toxins. J. Chem. Eng. Data 65 (4), 1946–1953.

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. 48 (4), 2097–2098.

Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N.S., Bletsou, A., Zwiener, C., Ibanez, M., Portoles, T., de Boer, R., Reid, M.J., Onghena, M., Kunkel, U., Schulz, W., Guillon, A., Noyon, N., Leroy, G., Bados, P., Bogialli, S., Stipanicev, D., Rostkowski, P., Hollender, J., 2015. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal. Bioanal. Chem. 407 (21), 6237–6255.

Sharif, D.I., Gallon, J., Smith, C.J., Dudley, E., 2008. Quorum sensing in cyanobacteria: *n*-octanoyl-homoserine lactone release and response, by the epilithic colonial cyanobacterium *Gloeothece* PCC6909. ISME J. 2 (12), 1171–1182.

Sheldon, M.T., Mistrik, R., Croley, T.R., 2009. Determination of ion structures in structurally related compounds using precursor ion fingerprinting. J. Am. Soc. Mass Spectrom. 20 (3), 370–376.

Shen, B., 2015. A new golden age of natural products drug discovery. Cell 163 (6), 1297–1300.

Singh, R.K., Tiwari, S.P., Rai, A.K., Mohapatra, T.M., 2011. Cyanobacteria: an emerging source for drug discovery. J. Antibiot. 64 (6), 401–412.

Sorokina, M., Steinbeck, C., 2020. Review on natural products databases: where to find data in 2020. J. Cheminf. 12 (1), 20.

Spoof, L., Catherine, A., 2017. Tables of microcystins and nodularins. In: Meriluoto, J., Spoof, L., Codd, G.A. (Eds.), Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis. John Wiley & Sons, Chichester, UK, pp. 526–537.

Svircev, Z., Drobac, D., Tokodi, N., Mijovic, B., Codd, G.A., Meriluoto, J., 2017. Toxicology of microcystins with reference to cases of human intoxications and epidemiological investigations of exposures to cyanobacteria and cyanotoxins. Arch. Toxicol. 91 (2), 621–650.

Swain, S.S., Paidesetty, S.K., Padhy, R.N., 2017. Antibacterial, antifungal and antimycobacterial compounds from cyanobacteria. Biomed. Pharmacother. 90, 760–776.

Tillett, D., Dittmann, E., Erhard, M., von Döhren, H., Borner, T., Neilan, B.A., 2000. Structural organization of microcystin biosynthesis in *Microcystis aeruginosa* PCC7806: an integrated peptide–polyketide synthetase system. Chem. Biol. 7 (10), 753–764.

van Santen, J.A., Jacob, G., Singh, A.L., Aniebok, V., Balunas, M.J., Bunsko, D., Neto, F.C., Castano-Espriu, L., Chang, C., Clark, T.N., Little, J.L.C., Delgadillo, D.A., Dorrestein, P.C., Duncan, K.R., Egan, J.M., Galey, M.M., Haeckl, F.P.J., Hua, A., Hughes, A.H., Iskakova, D., Khadilkar, A., Lee, J.H., Lee, S., LeGrow, N., Liu, D.Y., Macho, J.M., McCaughey, C.S., Medema, M.H., Neupane, R.P., O'Donnell, T.J., Paula, J.S., Sanchez, L.M., Shaikh, A.F., Soldatou, S., Terlouw, B.R., Tran, T.A., Valentine, M., van der Hooft, J.J.J., Vo, D.A., Wang, M.X., Wilson, D., Zink, K.E., Linington, R.G, 2019. The Natural Products Atlas: an open access knowledge base for microbial natural products discovery. ACS Cent. Sci. 5 (11), 1824–1833.

Walsh, J.P., Renaud, J.B., Hoogstra, S., McMullin, D.R., Ibrahim, A., Visagie, C.M., Tanney, J.B., Yeung, K.K.C., Sumarah, M.W, 2019. Diagnostic fragmentation filtering for the discovery of new chaetoglobosins and cytochalasins. Rapid Commun. Mass Spectrom. 33 (1), 133–139.

Wang, M.X., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., Porto, C., Bouslimani, A., Melnik, A.V., Meehan, M.J., Liu, W.T., Criisemann, M., Boudreau, P.D., Esquenazi, E., Sandoval-Calderon, M., Kersten, R.D., Pace, L.A., Quinn, R.A., Duncan, K.R., Hsu, C.C., Floros, D.J., Gavilan, R.G., Kleigrewe, K., Northen, T., Dutton, R.J., Parrot, D., Carlson, E.E., Aigle, B., Michelsen, C.F., Jelsbak, L., Sohlenkamp, C., Pevzner, P., Edlund, A., McLean, J., Piel, J., Murphy, B.T., Gerwick, L., Liaw, C.C., Yang, Y.L., Humpf, H.U., Maansson, M., Keyzers, R.A., Sims, A.C., Johnson, A.R., Sidebottom, A.M., Sedio, B.E., Klitgaard, A., Larson, C.B., Boya, C.A., Torres-Mendoza, A., Gonzalez, D.J., Silva, D.B., Marques, L.M., Demarque, D.P., Pociute, E., O'Neill, E.C., Briand, E., Helfrich, E.J.N., Granatosky, E.A., Glukhov, E., Ryffel, F., Houson, H., Mohimani, H., Kharbush, J.J., Zeng, Y., Vorholt, J.A., Kurita, K.L., Charusanti, P., McPhail, K.L., Nielsen, K.F., Vuong, L., Elfeki, M., Traxler, M.F., Engene, N., Koyama, N., Vining, O.B., Baric, R., Silva, R.R., Mascuch, S.J., Tomasi, S., Jenkins, S., Macherla, V., Hoffman, T., Agarwal, V., Williams, P.G., Dai, J.Q., Neupane, R., Gurr, J., Rodriguez, A.M.C., Lamsa, A., Zhang, C., Dorrestein, K., Duggan, B.M., Almaliti, J., Allard, P.M., Phapale, P., Nothias, L.F., Alexandrovr, T., Litaudon, M., Wolfender, J.L., Kyle, J.E., Metz, T.O., Peryea, T., Nguyen, D.T., VanLeer, D., Shinn, P., Jadhav, A., Muller, R., Waters, K.M., Shi, W.Y., Liu, X.T., Zhang, L.X., Knight, R., Jensen, P.R., Palsson, B.O., Pogliano, K., Linington, R.G., Gutierrez, M., Lopes, N.P., Gerwick, W.H., Moore, B.S., Dorrestein, P.C., Bandeira, N, 2016. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. 34 (8), 828–837.

Welker, M., von Döhren, H., 2006. Cyanobacterial peptides—Nature's own combinatorial biosynthesis. FEMS Microbiol. Rev. 30 (4), 530–563.

WHO (2004) Guidelines for Drinking Water Quality. World Health Organization, Geneva. Available online: https://www.who.int/water_sanitation_health/dwq/GDWQ2004web.pdf?ua=1. Assessed on 14 April 2020.

WHO (2020a) Background document for development of WHO Guidelines for drinking-water quality and Guidelines for safe recreational water environments. Cyanobacterial toxins: Anatoxin-A and Analogues. World Health Organization, Geneva. Available online: https://apps.who.int/iris/handle/10665/338060. Assessed on 15. Sep. 2020.

WHO (2020b) Background document for development of WHO Guidelines for drinking-water quality and Guidelines for safe recreational water environments. Cyanobacterial toxins: Cylindrospermopsins. World Health Organization, Geneva. Available online: https://apps.who.int/iris/handle/10665/338063. Assessed on 15. Sep. 2020.

WHO (2020c) Background document for development of WHO Guidelines for drinking-water quality and Guidelines for safe recreational water environments. Cyanobacterial toxins: Microcystins. World Health Organization, Geneva. Available online: https://apps.who.int/iris/handle/10665/338066. Assessed on 15. Sep. 2020.

WHO (2020d) Background document for development of WHO Guidelines for drinking-water quality and Guidelines for safe recreational water environments. Cyanobacterial toxins: Saxitoxins. World Health Organization, Geneva. Available online: https://apps.who.int/iris/handle/10665/338069. Assessed on 15. Sep. 2020.

Wickert, J., Lorsbach, T., Gutlein, M., Schmid, E., Latino, D., Kramer, S., Fenner, K., 2016. enviPath–The environmental contaminant biotransformation pathway resource. Nucleic Acids Res. 44 (D1), 502–508.

Yamaki, H., Sitachitta, N., Sano, T., Kaya, K., 2005. Two new chymotrypsin inhibitors isolated from the cyanobacterium *Microcystis aeruginosa* NIES-88. J. Nat. Prod. 68 (1), 14–18.

Zhu, Z., Zhang, L., Shi, G.Q., 2015. Proteasome as a molecular target of microcystin-LR. Toxins (Basel) 7 (6), 2221–2231.

APPENDIX II

Table A1. ASM-1 culture medium composition (GORHAM et al., 1964).

Stock solution	Components	Concentration in the stock solution (mg/L)	Concentration in the culture medium (mg/L)
A	$NaNO_3$	8.5	170
	$MgSO_4.7H_2O$	2.45	49
	$MgCl_2.6H_2O$	2.05	41
	$CaCl_2.2H_2O$	1.45	29
B	KH_2PO_4	8.7	43.5
	$Na_2HPO_4.12H_2O$	7.06	35.3
	H_3BO_3	28.4	2.84
	$MnCl_2.4H_2O$	13.9	1.39
C	$FeCl_2.6H_2O$	9.14	0.914
	$ZnCl_2$	3.35	0.335
	$CoCl_2.6H_2O$	0.19	0.019
	$CuCl_2.2H_2O$	0.014	0.0014
D	Titriplex® III (EDTA)	18.6	7.44

Table A2. Molecular networks generated on the Global Natural Products Social Molecular Networking (GNPS) platform (available on http://gnps.ucsd.edu)

Sample	GNPS link
Molecular network of NPCD-01 fractions #9 and # 10 (RT 22 to 26 min)	https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=b824403b3b774a27acda0f57a033ed1f
MolNetEnhancer analysis of NPCD-01 fractions #9 and # 10 (RT 22 to 26 min)	https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=8bf6ebc2f98144b382d1acff5f2f5c95
Molecular network of MIRS-04 fractions #9 and # 10 (22 – 26 min)	https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=dfed4df6fa354cdb9bf8c5a3c406de33
MolNetEnhancer analysis of MIRS-04 fractions #9 and # 10 (RT 22 to 26 min)	https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=f06f839360cc4fa9a76824cbab9e0e98

Table A3. Concentration (mg/L) of cyanopeptides as MC-LR equivalents in the FET test exposure solutions at 0h in the experiments with the pooled and single fractions of both MIRS-04 and NPCD-01 strains.

Cyanopeptide	Dilution of FET exposure solution	Cyanopeptide concentration (mg/L, MC-LR equivalents)	
Microcystis panniformis (MIRS-04)			
		Pooled fractions	Fraction #7 (RT 16 to 18 min)
Micropeptin K139	C1	24.716	18.324
	C2	9.886	7.329
	C3	3.955	2.932
	C4	1.582	1.173
	C5	0.633	0.469
Microcystin-LR	C1	6.182	0.00340
	C2	2.473	0.00144
	C3	0.989	0.00077
	C4	0.396	0.00061
	C5	0.158	0.00060
Cyanopeptolin 972	C1	3.769	0.105
	C2	1.508	0.050
	C3	0.603	0.027
	C4	0.241	0.002
	C5	0.081	0.001
Cyanopeptolin 958	C1	2.906	0.004
	C2	1.162	0.002
	C3	0.465	0.001
	C4	0.186	0.001
	C5	0.070	0.001
[D-Asp^3]MC-LR	C1	0.847	0.001
	C2	0.339	0.001
	C3	0.135	<LQ
	C4	0.047	<LQ
	C5	0.027	<LQ
Cyanopeptolin 1014	C1	0.764	0.004
	C2	0.254	0.002
	C3	0.076	0.001
	C4	0.036	0.001
	C5	0.005	0.001
[D-Asp^3,ADMAdda^5]MC-LHar	C1	0.749	0.001
	C2	0.311	0.001
	C3	0.110	0.001
	C4	0.044	<LQ
	C5	0.026	<LQ
MC-LY	C1	0.377	<LQ
	C2	0.151	<LQ
	C3	0.065	<LQ
	C4	0.035	<LQ
	C5	0.006	<LQ

LOQ 0.0027 mg/L; LOD 0.0083 mg/L.

Table A3. Concentration (mg/L) of cyanopeptides as MC-LR equivalents in the FET test exposure solutions at 0h in the experiments with the pooled and single fractions of both MIRS-04 and NPCD-01 strains (continued).

Cyanopeptide	Dilution	Cyanopeptide concentration (mg/L, MC-LR equivalents)		
		Pooled fractions	Fraction #1 (RT 4 to 6 min)	Fraction #3 (RT 8 to10 min)
		Microcystis aeruginosa(NPCD-01)		
Nostoginin BN741	C1	2.316	0.012	2.910
	C2	0.927	0.005	1.164
	C3	0.371	0.001	0.466
	C4	0.148	<LQ	0.186
	C5	0.060	<LQ	0.076
Microginin FR6	C1	0.283	0.003	0.165
	C2	0.113	0.002	0.059
	C3	0.043	0.001	0.031
	C4	0.007	<LQ	0.005
	C5	0.003	<LQ	0.002
Cyanopeptolin 959	C1	0.201	0.089	0.031
	C2	0.088	0.052	0.005
	C3	0.042	0.034	0.002
	C4	0.027	0.010	0.001
	C5	0.003	0.005	0.001
Microginin SD755	C1	0.155	<LQ	0.051
	C2	0.052	<LQ	0.027
	C3	0.026	<LQ	0.002
	C4	0.001	<LQ	0.001
	C5	0.002	<LQ	0.001

LOQ 0.0027 mg/L; LOD 0.0083 mg/L.
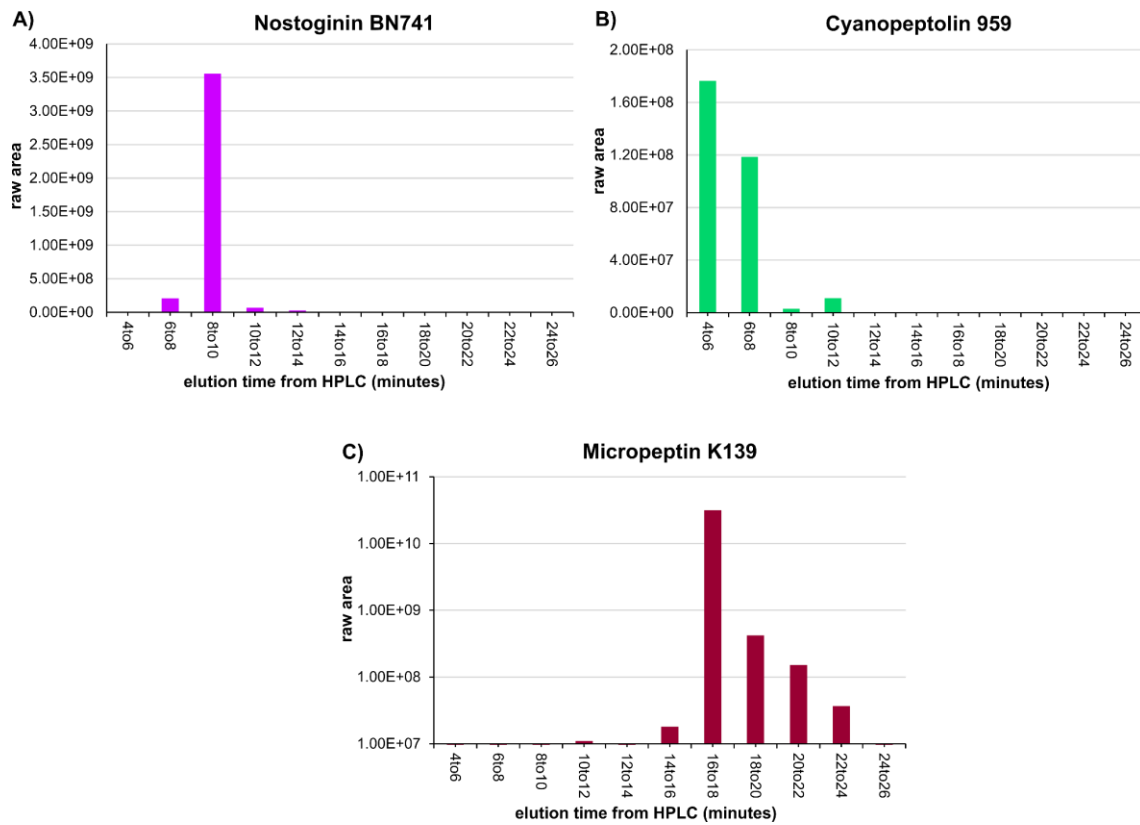
Figure A1. Raw area distribution of nostoginin BN741 (A) and cyanopeptolin-959 (B) over the different fractions from the NPCD-01 extract; and raw area distribution of micropeptin K139 over the fractions of the MIRS-01 strain (C).

Micropeptin K139
m/z 987.5488 | RT 12.21 | HCD 30 | [Arg-Ahp-Ile-N-Me-Tyr-Ile-O-Thr-Asp-HA]
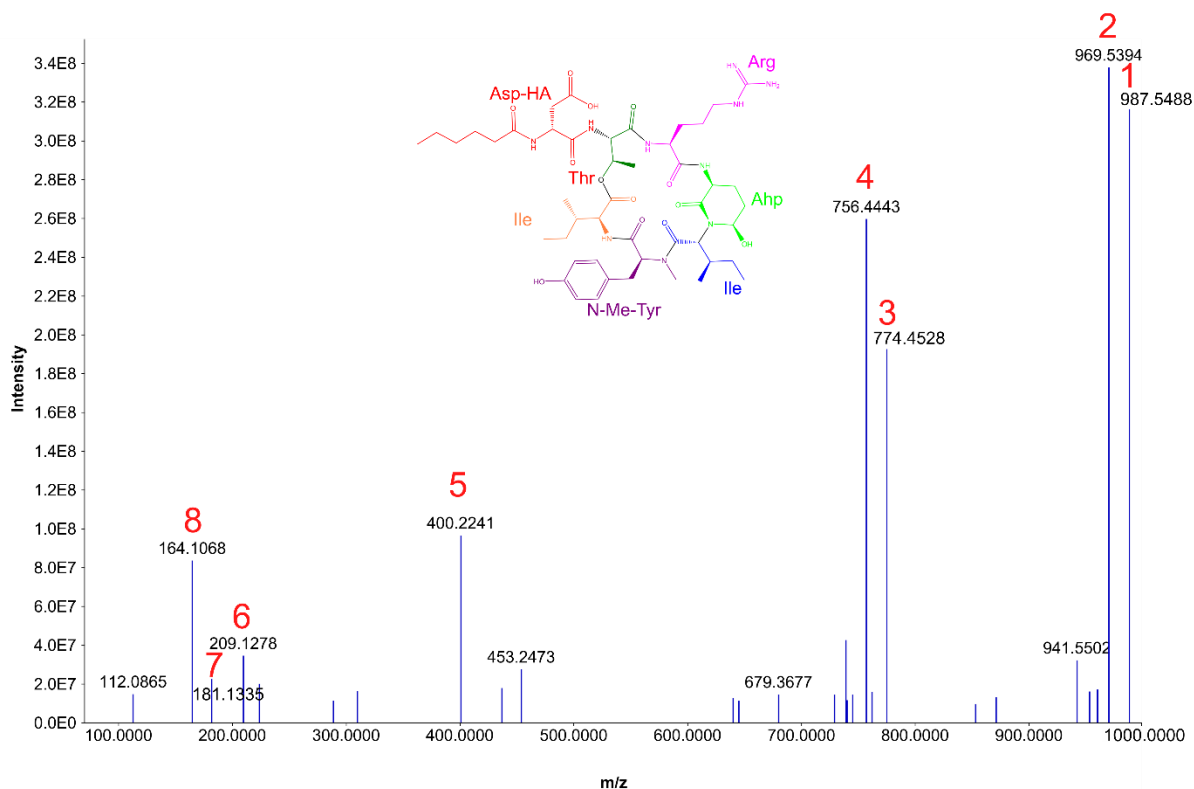
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
1	987.5510	987.5488	M+H
2	969.5404	969.5394	[M - H_2O]+
3	774.4509	774.4528	[M - Asp–HA] + H+
4	756.4405	756.4443	[M - Asp–HA - H_2O] + H+
5	400.2205	400.2241	[M - Arg–Ahp–Ile–Me–Tyr] +H +
6	209.1274	209.1278	[Ile–Ahp - H_2O] + H+
7	181.1336	181.1335	[Ile–Ahp - H_2O - CO] + H+
8	164.1070	164.1068	MeHTyr immonium ion

Figure A2. MS/MS annotation at HCD30 for the *m/z* 987.5488 and RT 12.21, matching with micropeptin K139 (*m/z* 987.5510, $C_{47}H_{74}N_{10}O_{13}$, confidence level 2b based on MS^2 annotation) from the MIRS-04 strain.

MC-LR

m/z 995.5544 | RT 9.91 | HCD 30 | cyclo[D-Ala-Leu-D-bMe-Asp-Arg-Adda-D-Glu-Mdha]
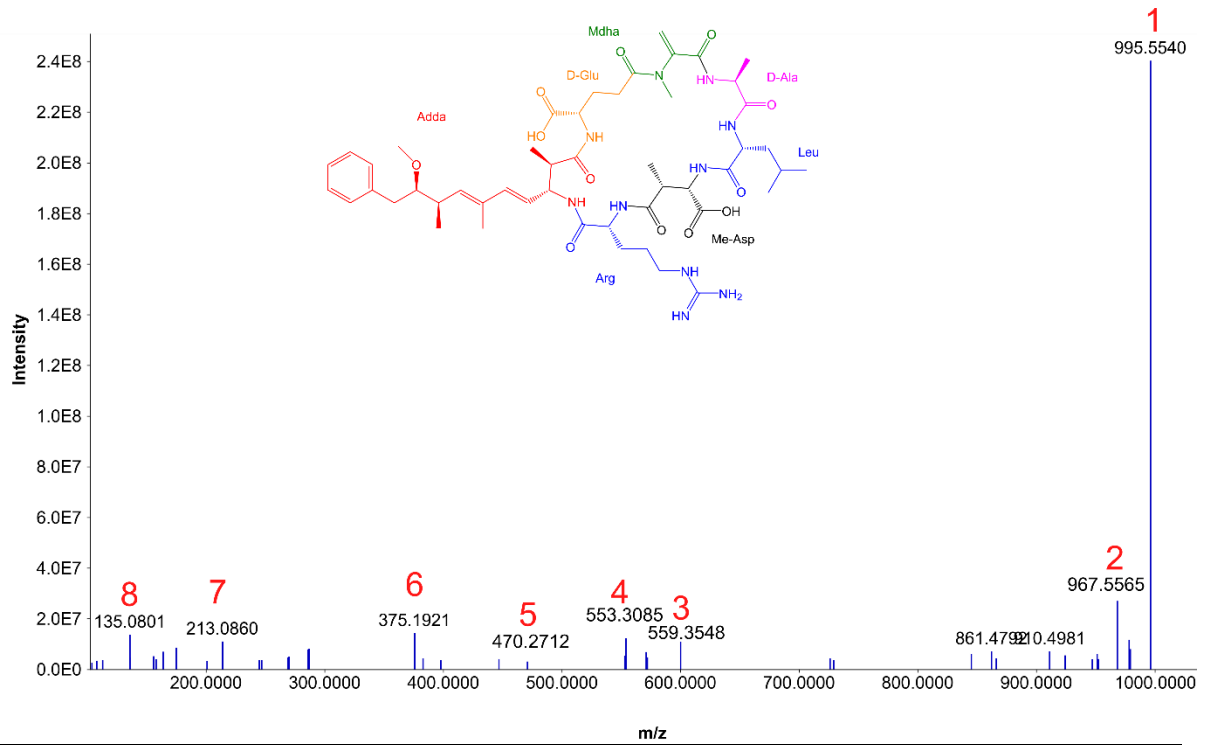
Fragment number	Calculated *m/z*	Peak *m/z* MIRS-04 pooled fractions	Peak *m/z* MC-LR analytical standard (0.5mg/L)	Fragment
1	995.5560	995.5540	995.5533	M+H
2	967.5611	967.5565	967.5590	[M - CO]+ H+
3	599.3552	599.3548	599.3566	[M - Glu–Mdha–Ala–Arg] + H+
4	553.3093	553.3085	553.3093	[M - Adda–Glu] + H+
5	470.2727	470.2712	470.2680	[M - Adda–Glu–Mdha]+H+
6	375.1914	375.1921	375.1924	[$C_{11}H_{15}O$ (Adda fragment)–Glu–Mdha]+
7	213.0875	213.0860	213.0872	[Glu–Mdha] + H+
8	135.0804	135.0801	135.0804	Adda fragment

Figure A3. MS/MS annotation at HCD30 for *m/z* 995.5544 and RT 9.91, matching with microcystin-LR (*m/z* 995.5560, $C_{49}H_{74}N_{10}O_{12}$, confidence level 1 confirmed by reference standard) from the MIRS-04 extract and analytical standard.

Cyanopeptolin 972
m/z 973.5345 | RT 8.59 | HCD 30 | [Arg-Ahp-Leu/Ile-N-Me-Tyr-Val-O-Thr-Asp-HA]
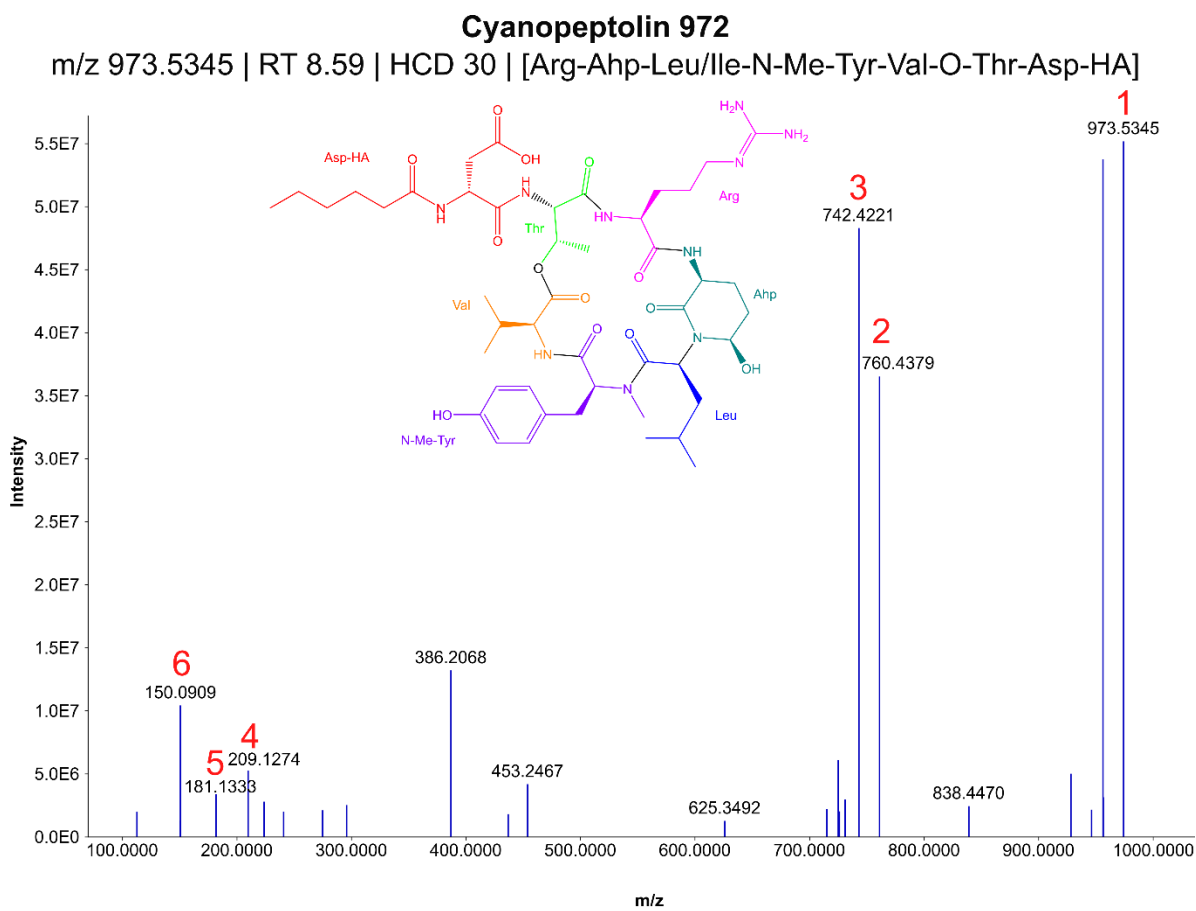
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
1	973.5353	973.5345	M+H
2	760.4358	760.4379	[M - Asp-HA] + H+
3	742.4252	742.4221	[M - Asp-HA - H_2O] + H+
4	209.1274	209.1274	[Leu/Ile–Ahp - H_2O] + H+
5	181.1336	181.1333	[Leu/Ile–Ahp - H_2O - CO] + H+
6	150.0909	150.0909	Tyr fragment

Figure A4. MS/MS annotation at HCD30 for *m/z* 973.5345 and RT 8.59, matching with cyanopeptolin 872 (*m/z* 973.5353, $C_{46}H_{72}N_{10}O_{13}$, confidence level 2b based on MS^2 annotation) from the MIRS-04 extract.

Cyanopeptolin 958
m/z 959.5169 | RT 8.82 | HCD 30 | [Arg-Ahp-Leu-N-Me-Tyr-Val-O-Thr-Glu-But]
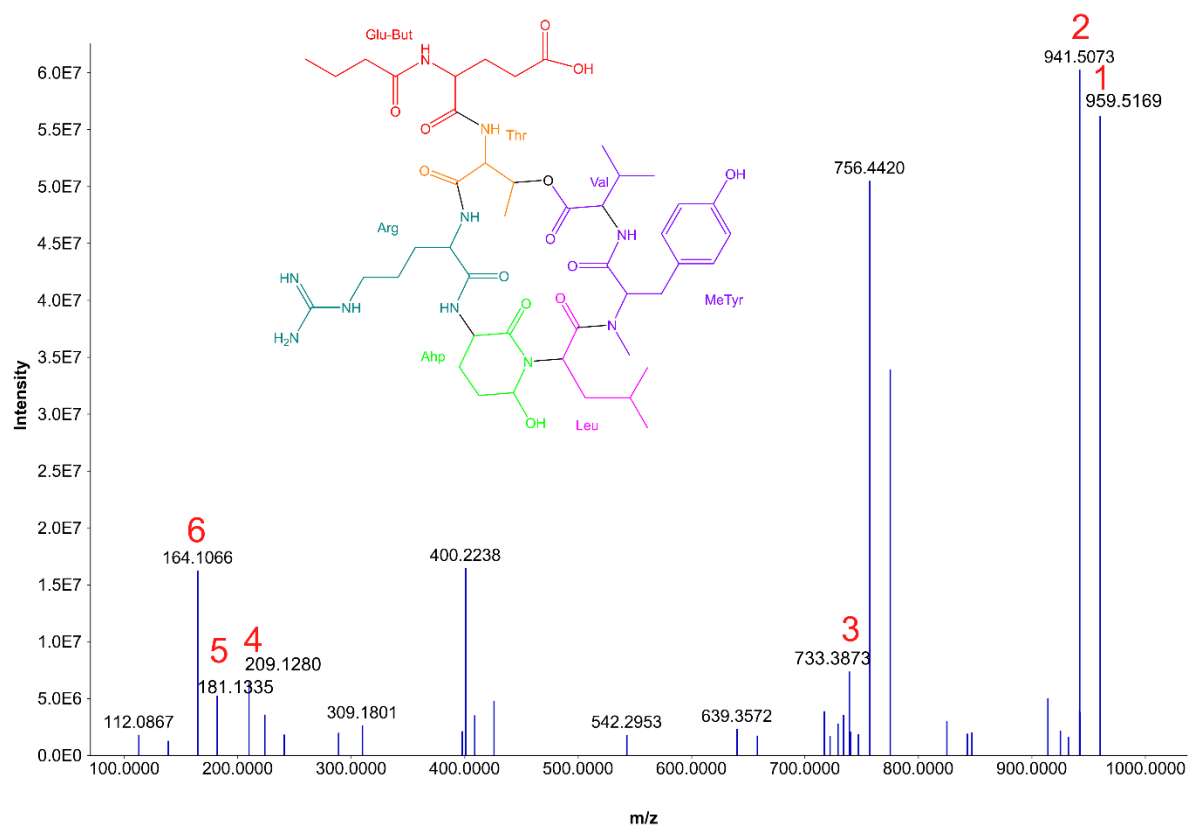
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
1	959.5197	959.5169	M+H
2	941.5096	941.5073	[M - H_2O]+
3	733.3885	733.3873	[M - Leu–Ahp]+
4	209.1274	209.1280	[Leu/Ile–Ahp - H_2O] + H+
5	181.1336	181.1335	[Leu/Ile –Ahp - H_2O - CO] + H+
6	164.1070	164.1066	MeHTyr immonium ion

Figure A5. MS/MS annotation at HCD30 for *m/z* 959.5169 and RT 8.82, matching with cyanopeptolin 958 (*m/z* 959.5197, $C_{45}H_{70}N_{10}O_{13}$, confidence level 2b based on MS^2 annotation) from the MIRS-04 extract.
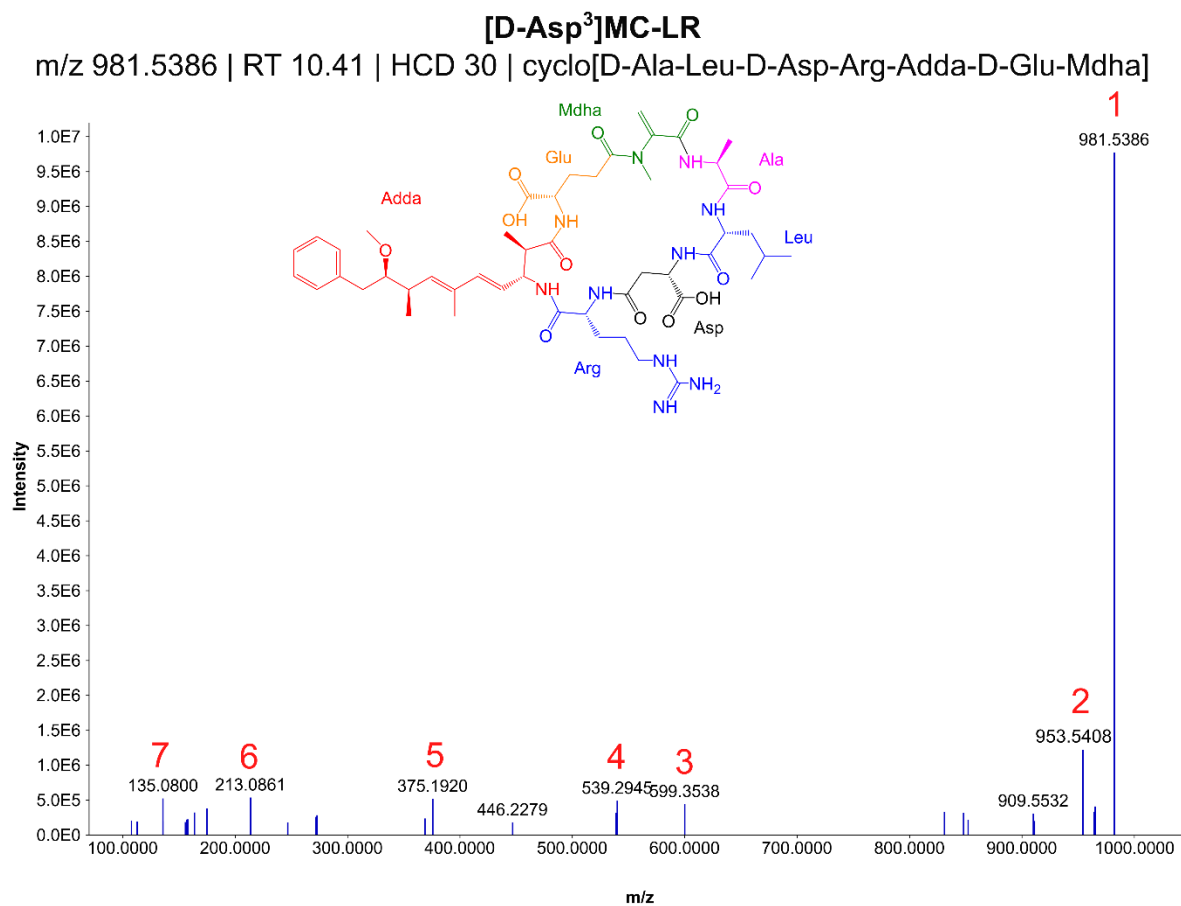
[D-Asp³]MC-LR

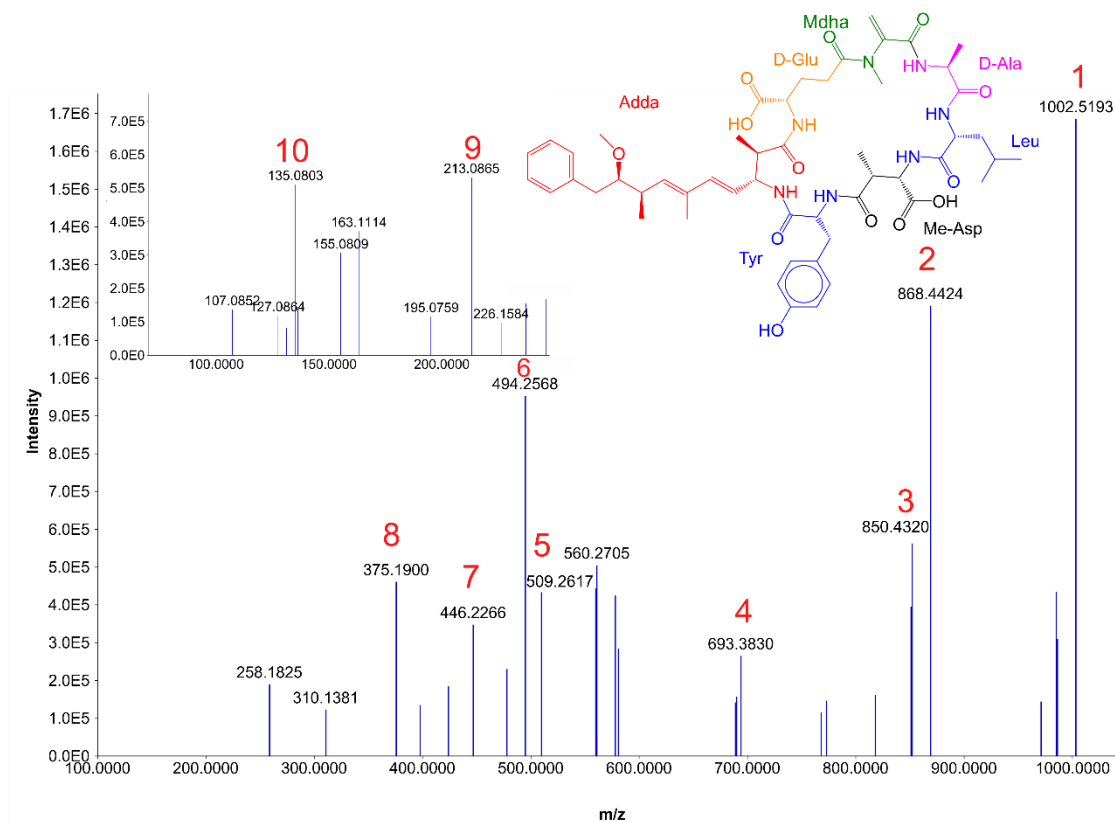m/z 981.5386 | RT 10.41 | HCD 30 | cyclo[D-Ala-Leu-D-Asp-Arg-Adda-D-Glu-Mdha]

Fragment number	Calculated m/z	Peak m/z MIRS-04 pooled fractions	Peak m/z MC-LR analytical standard (0.5mg/L)	Fragment
1	981.5404	981.5386	981.5408	[M + H]+
2	953.5455	953.5408	953.5413	M+ H -CO
3	599.3557	599.3538	599.3574	[M - D-Ala - Leu - D-Asp -Mdha]
4	539.2942	539.2945	539.2950	[M - Adda - D-Glu - Mdha]
5	375.1914	375.1920	375.1925	[$C_{11}H_{15}O$ (Adda fragment)–Glu–Mdha]+
6	213.0875	213.0861	213.0869	[Glu–Mdha] + H+
7	135.0804	135.0800	135.0804	Adda fragment

Figure A6. MS/MS annotation at HCD30 for m/z 981.5386 and RT 10.41, matching with [D-Asp³]microcystin-LR (m/z 981.5404, $C_{48}H_{72}N_{10}O_{12}$, confidence level 1 confirmed by reference standard) from the MIRS-04 extract and analytical standard.

Cyanopeptolin 1014
m/z 1015.5784 | RT 16.82 | HCD 30 | [Arg-Ahp-Leu-N-Me-Tyr-Val-O-Thr-Glu-OA]
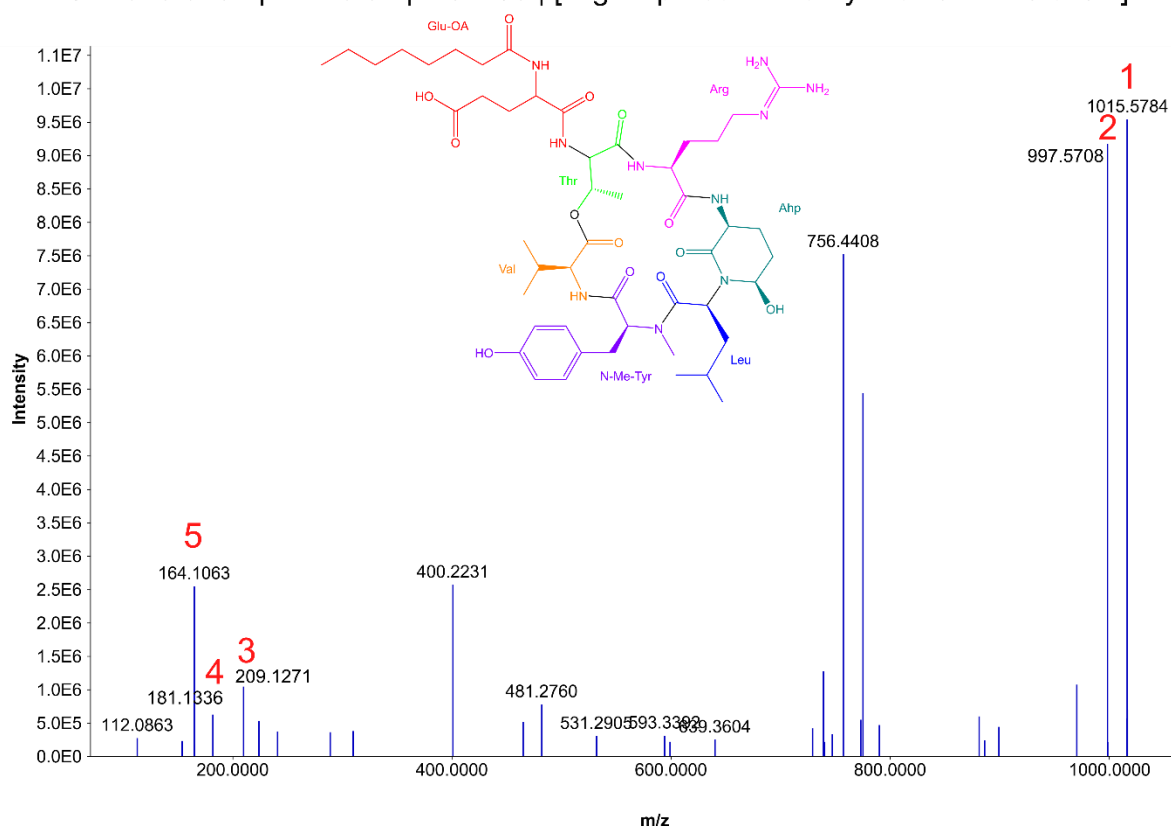
Fragment number	Calculated m/z	Peak m/z	Fragment
1	1015.5823	1015.5784	M+H
2	997.5717	997.5708	[M - H_2O]+
3	209.1274	209.1271	[Leu/Ile–Ahp - H_2O] + H+
4	181.1336	181.1336	[Leu/Ile –Ahp - H_2O - CO] + H+
5	164.1070	164.1063	MeHTyr immonium ion

Figure A7. MS/MS annotation at HCD30 for m/z 1015.5784 and RT 16.82, matching with cyanopeptolin 1014 (m/z 1015.5823, $C_{49}H_{78}N_{10}O_{13}$, confidence level 2b based on MS^2 annotation) from the MIRS-04 extract.

[D-Asp³,ADMAdda⁵]MC-LHar

m/z 1023.5504 | RT 16.61| HCD 30 | cyclo[D-Ala-Leu-D-Asp-Har-ADMAdda-D-Glu-Mdha]
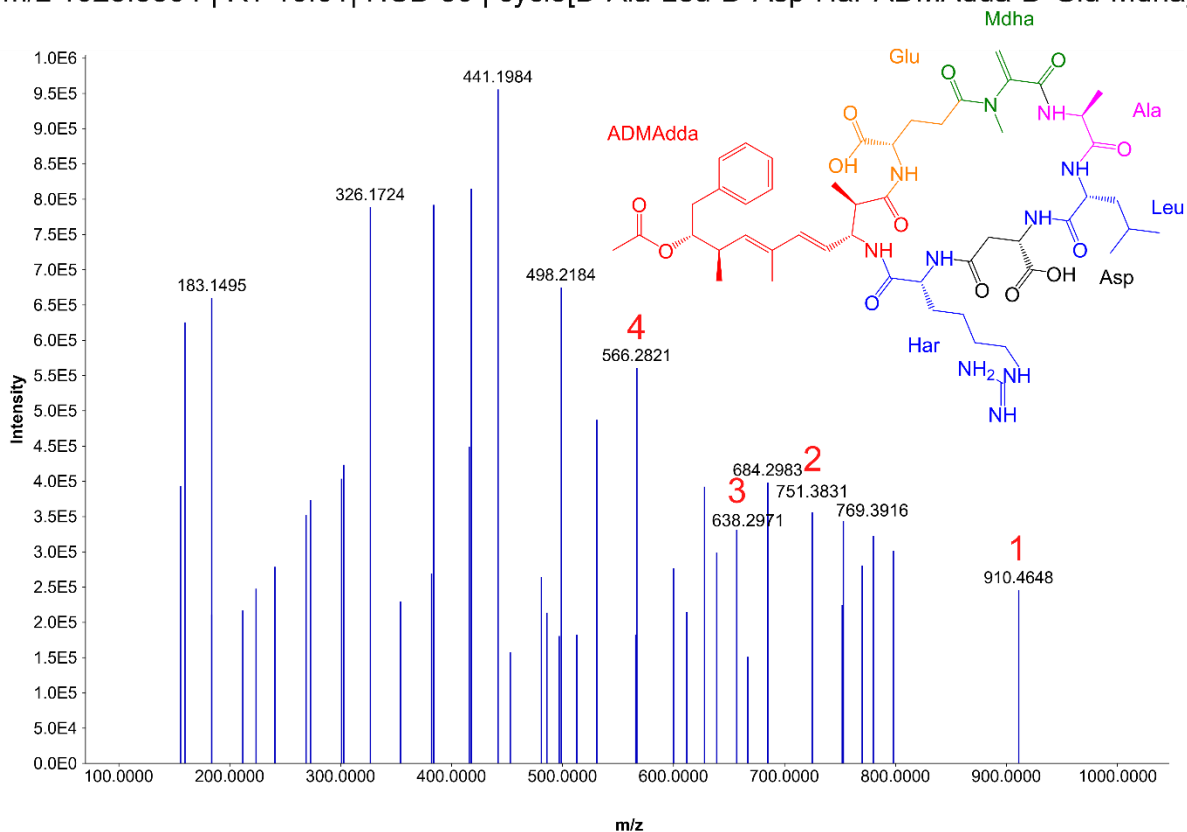
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
*	1023.5509	1023.5504	M+H
1	910.4674	910.4648	[M- Leu]
2	751.3865	751.3831	[M - Part of ADMAdda ($C_{17}H_{21}O_2$) - NH_2]
3	638.2952	638.2971	[M - Har - Asp - Leu]
4	566.2813	566.2821	[M - ADMAdda - Glu]

Figure A8. MS/MS annotation at HCD30 for *m/z* 1023.5504 and RT 16.61, matching with [D-Asp³,ADMAdda⁵]MC-LHar (*m/z* 1023.5509, $C_{50}H_{74}N_{10}O_{13}$, confidence level 3 as only few fragments could be annotated) from the MIRS-04 extract. (*) precursor ion peak not shown in the spectrum but visible with HCD15..

MC-LY
m/z 1002.5193 | RT 15.50 | HCD 15&30 | cyclo[D-Ala-Leu-D-MeAsp-Tyr-Adda-D-Glu-Mdha]

Fragment number	Calculated *m/z*	Peak *m/z* MIRS-04 pooled fractions	Peak *m/z* MC-LR analytical standard (0.5mg/L)	Fragment
1	1002.5183	1002.5193	1002.5165	[M + H]+
2	868.4456	868.4424	868.4462	[M - Adda fragment] + H+
3	850.4351	850.4320	850.4350	[M - Adda fragment - H_2O] +
4	693.3863	693.3830	693.3828	[M - Leu - D-MeAsp - Tyr - NH_3] +
5	509.2652	509.2617	509.2639	[M - Ala - Leu - D-MeAsp - Tyr - NH_3] +
6	494.2615	494.2568	494.2604	[M - Adda - Glu - Mdha - NH_2] + H+
7	446.2291	446.2266	446.2287	[M- Leu - D-MeAsp - Tyr - NH_3 - $C_9H_{11}O$ (Part of Adda)] +
8	375.1914	375.1900	375.1925	[$C_{11}H_{15}O$ (Adda fragment)–Glu–Mdha]+
9	213.0875	213.0865	213.0869	[Glu–Mdha] + H+
10	135.0804	135.0803	135.0804	Adda fragment

Figure A9. MS/MS annotation at HCD 15 and HCD30 (see inset) for *m/z* 1002.5193 and RT 15.50, matching with microcystin-LY (*m/z* 1002.5183, $C_{50}H_{74}N_{10}O_{13}$, confidence level 1 confirmed by reference standard) from the MIRS-04 extract and analytical standard.

Nostoginin BN741
m/z 742.4386 | RT 7.47 | HCD 30 | [Ahoa-Val-NMeIle-NMeTyr-Tyr]
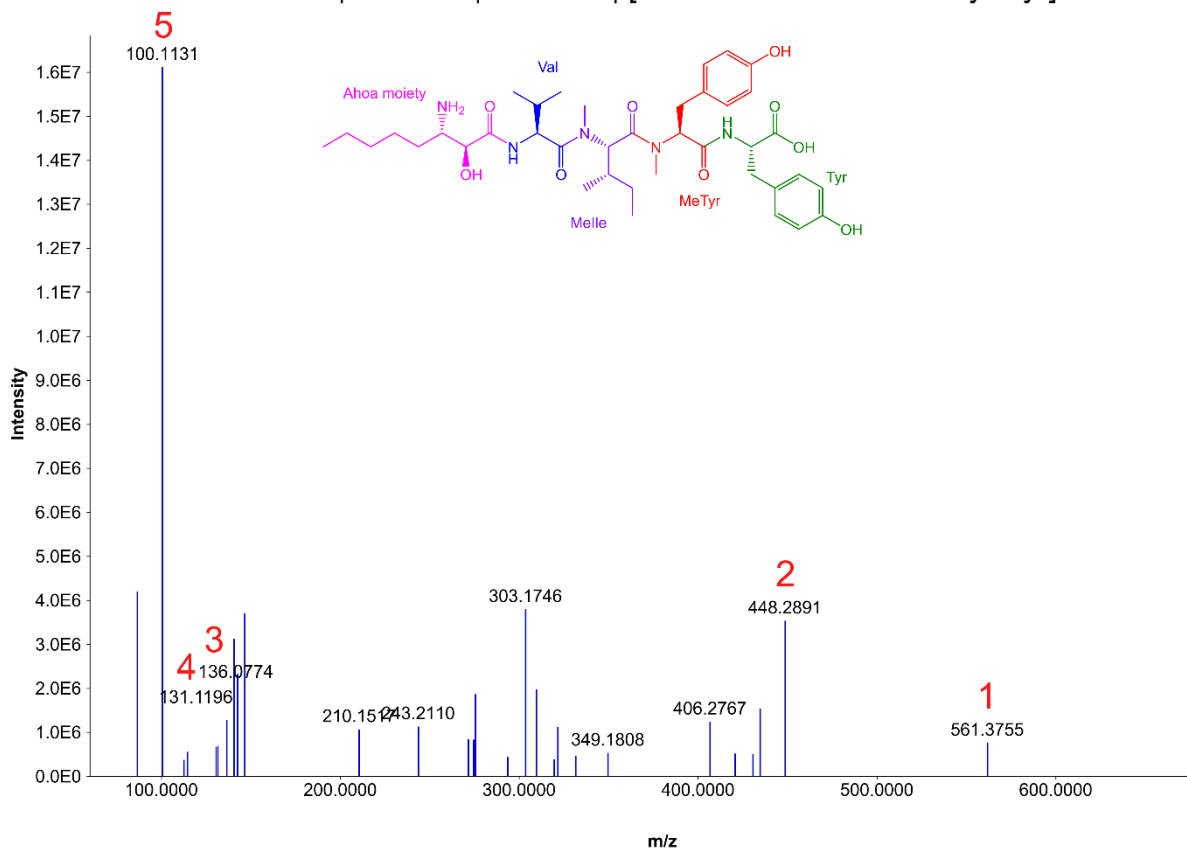
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
*	742.4373	742.4386	M+H
1	561.3652	561.3755	[M-Tyr]+
2	448.2811	448.2891	[Part of Ahoa ($C_2H_2O_2$)–Val–MeLeu–MeTyr]+
3	136.0762	136.0774	Tyr immonium ion
4	131.1184	131.1196	Valine residue
5	100.1126	100.1131	Ahoa fragment or MeIle immonium ion

Figure A10. MS/MS annotation at HCD30 for *m/z* 742.4386 and RT 7.47, matching with microginin variant Nostoginin BN741 (*m/z* 742.4373, $C_{39}H_{59}N_5O_9$, confidence level 2b based on MS^2 annotation) from the NPCD-01 extract. (*) precursor ion peak not shown in the spectrum but visible with HCD15.

Microginin FR6
m/z 740.4224 | RT 6.00 | HCD 30 | [MeAhda-Val-Pro-Tyr-Tyr]
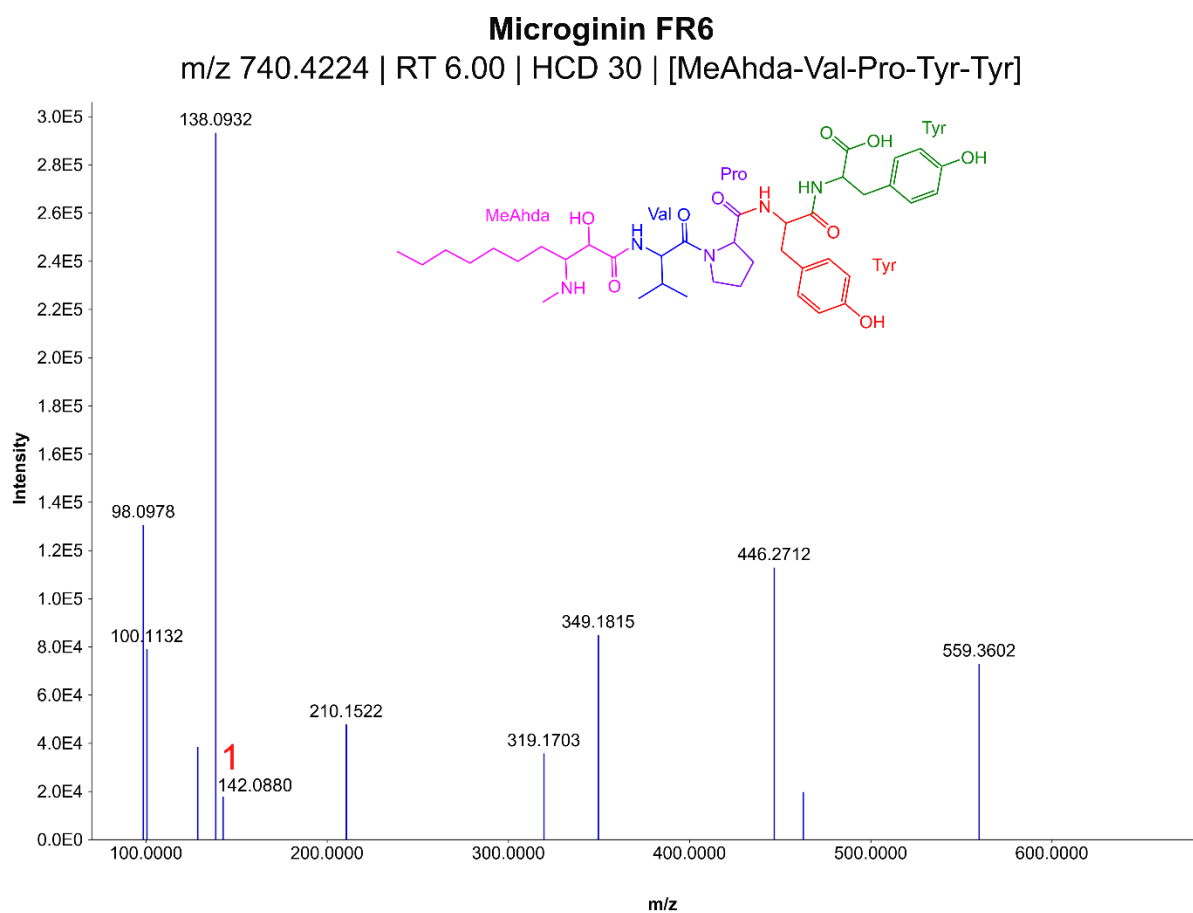
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
*	740.4229	740.4224	M+H
1	142.0882	142.0880	[MeAhda] + H+

Figure A11. MS/MS annotation at HCD30 for *m/z* 740.4224 and RT 6.00, matching with microginin FR6 (*m/z* 740.4229, $C_{39}H_{57}N_5O_9$, level 3 as only few fragments could be annotated) from the NPCD-01 extract. (*) precursor ion peak not shown in the spectrum but visible with HCD15.

Cyanopeptolin 959

m/z 960.4119 | RT 9.30 | HCD 30 & 45| [Arg-Ahp-Phe-NMe-Tyr-Ile-Thr]-SuGA
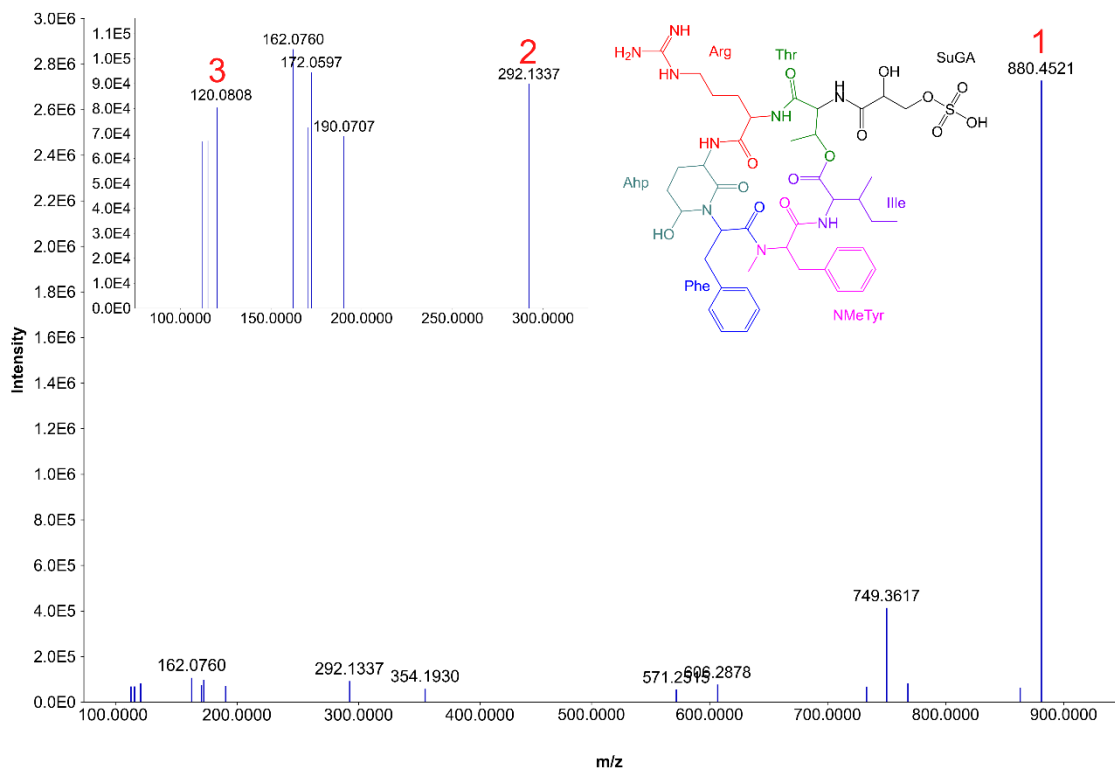
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
*	960.4131	960.4119	M+H
1	880.4563	880.4521	[M+H+] - SO_3
2	292.1333	292.1337	Phe–MeTyr - H_2O
3	120.0808	120.0808	Phe immonium ion

Figure A12. MS/MS annotation at HCD 30 and HCD45 (inset) for *m/z* 960.4119 and RT 9.30, matching with cyanopeptolin 959 (*m/z* 960.4131, $C_{43}H_{61}N_9O_{14}S$, level 2b based on MS^2 annotation) from the NPCD-01 extract. (*) precursor ion peak not shown in the spectrum but visible with HCD15.

Microginin SD755
m/z 756.4528 | RT 9.24 | HCD 30 | MeAhoa-Val-Melle-MeTyr-Tyr
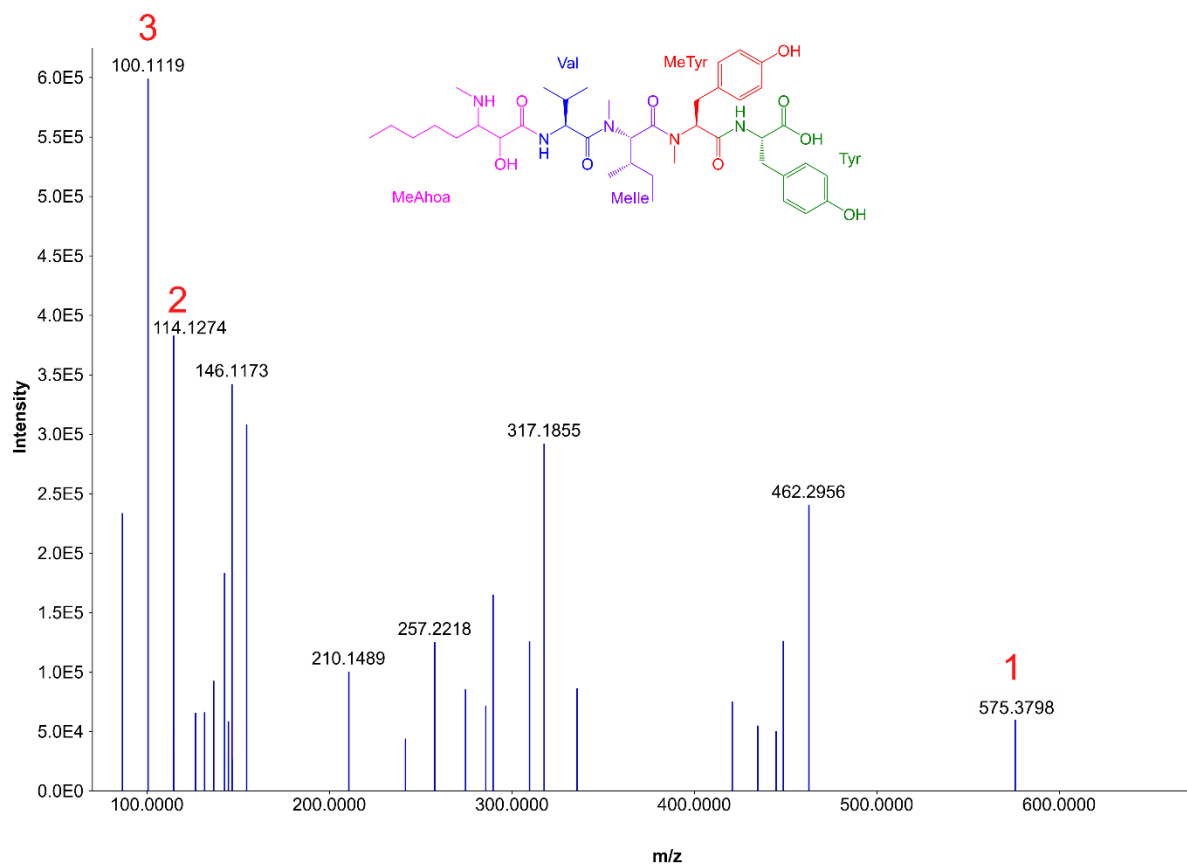
Fragment number	Calculated *m/z*	Peak *m/z*	Fragment
*	756.4542	756.4528	M+H
1	575.3809	575.3798	[M-Tyr]+
2	114.1278	114.1274	MeAhoa fragment
3	100.1120	100.1119	MeLeu immonium

Figure A13. MS/MS annotation at HCD30 for *m/z* 756.4528 and RT 9.24, matching with microginin SD755 (*m/z* 756.4542, $C_{40}H_{61}N_5O_9$, confidence level 2b based on MS^2 annotation) from the NPCD-01 extract. (*) precursor ion peak not shown in the spectrum but visible with HCD15.

Glyceryl palmitate
m/z 331.2843 | RT 8.13 | HCD 15
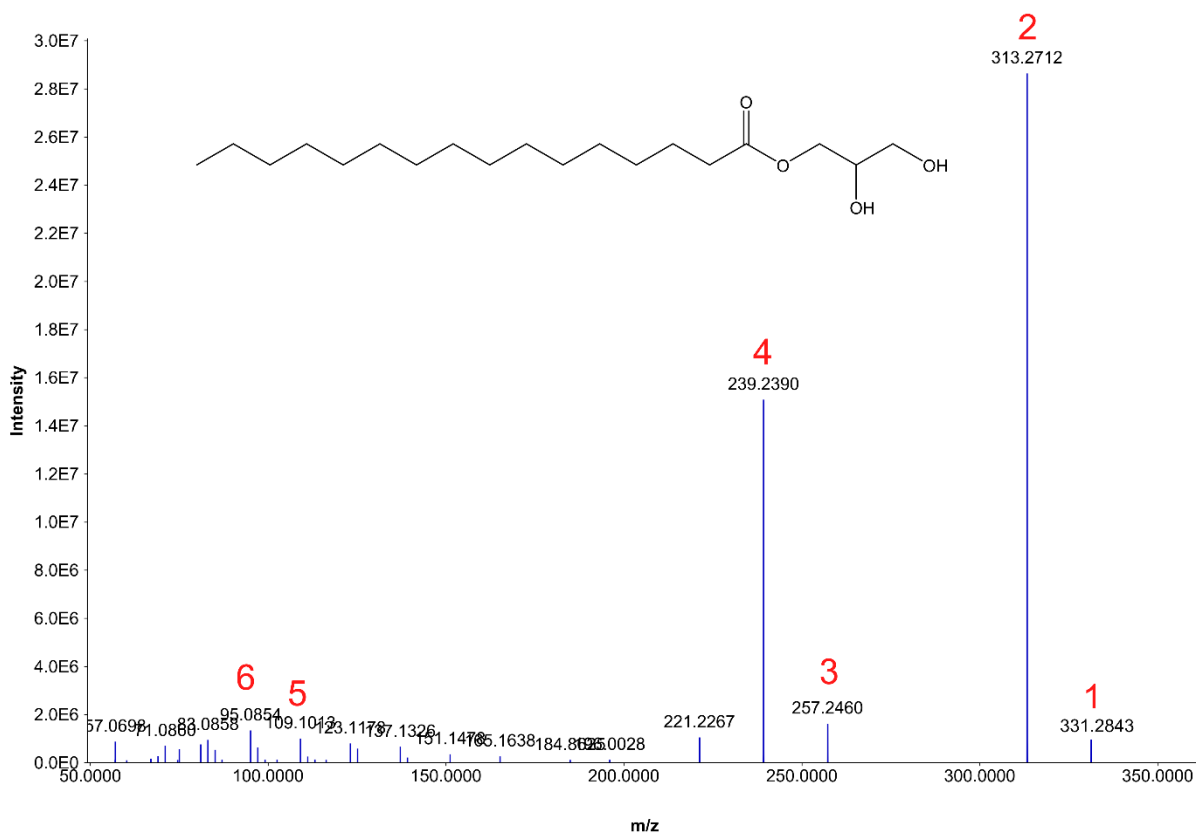
Fragment number	Calculated *m/z*	Peak *m/z*
1	331.2843	331.2843
2	313.2743	313.2712
3	257.2445	257.2460
4	239.2373	239.2390
5	109.1008	109.1009
6	95.0851	95.0851

Figure A14. MS/MS annotation at HCD15 for *m/z* 331.2843 and RT 8.13, detected in the HPLC fractions #9 and # 10 (22 – 26 min) of the NPCD-01 extract, matching with glyceryl palmitate (*m/z* 331.2843, $C_{19}H_{38}O_4$, confidence level 2a based on MS^2 annotation. Compound and spectra information were obtained in the MassBank of North America (MoNA ID CCMSLIB00000849055).
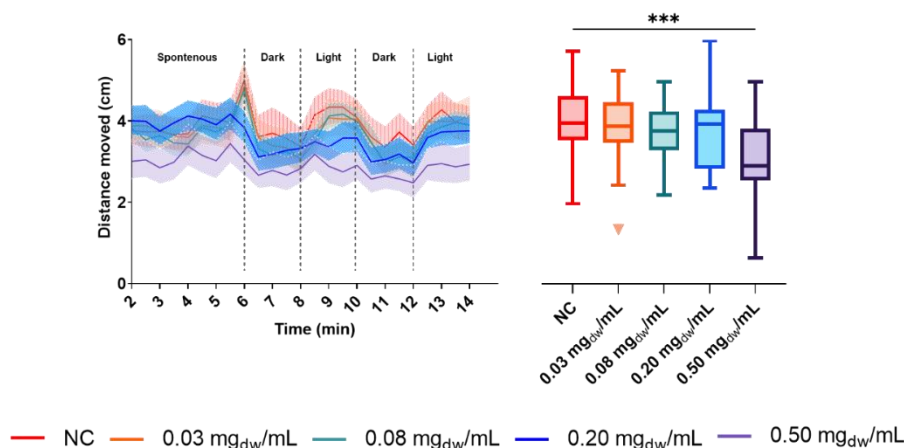
Figure A15. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to the different concentrations of the pooled fractions subtracting fraction #7 from the MIRS-04 strain, in terms of dry weight biomass. Significant alterations between groups are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001). NC: Negative control:
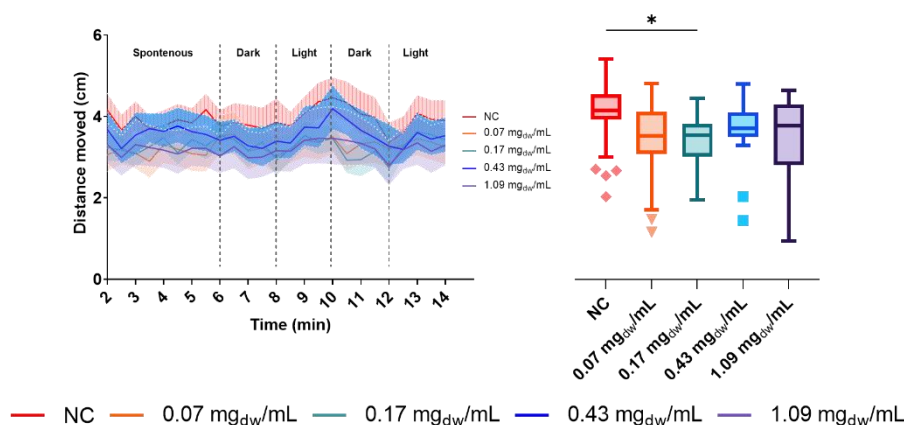
Figure A16. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplot representing the average distance moved of *T. platyurus* instar larvae exposed to the different concentrations of the apolar fractions #9+10 from the NPCD strain, in terms of dry weight biomass. Significant alterations between groups are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001). NC: Negative control:
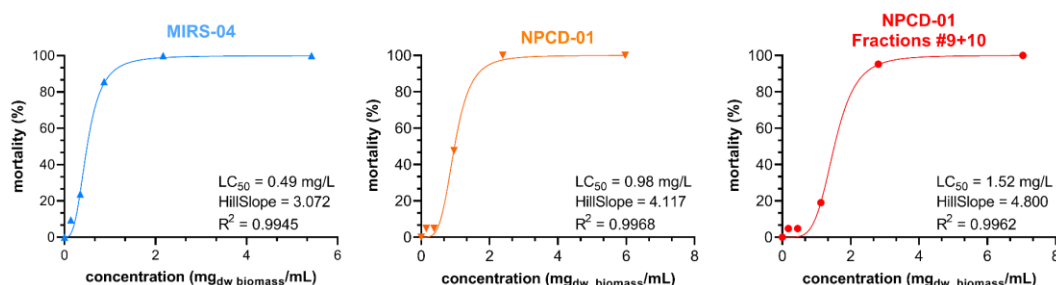
Figure A17. Mortality curves after 120h exposure of zebrafish larvae to MIRS-04 pooled fractions (a), NPCD-01 pooled fractions (b), and NPCD-01's fractions #9+10 (c). Fitted lines and LC_{50} values were obtained using a non-linear regression model, and the slope factor (HillSlope) and goodness of fit (R^2) of each model are presented in the plots (*p*<0.05).
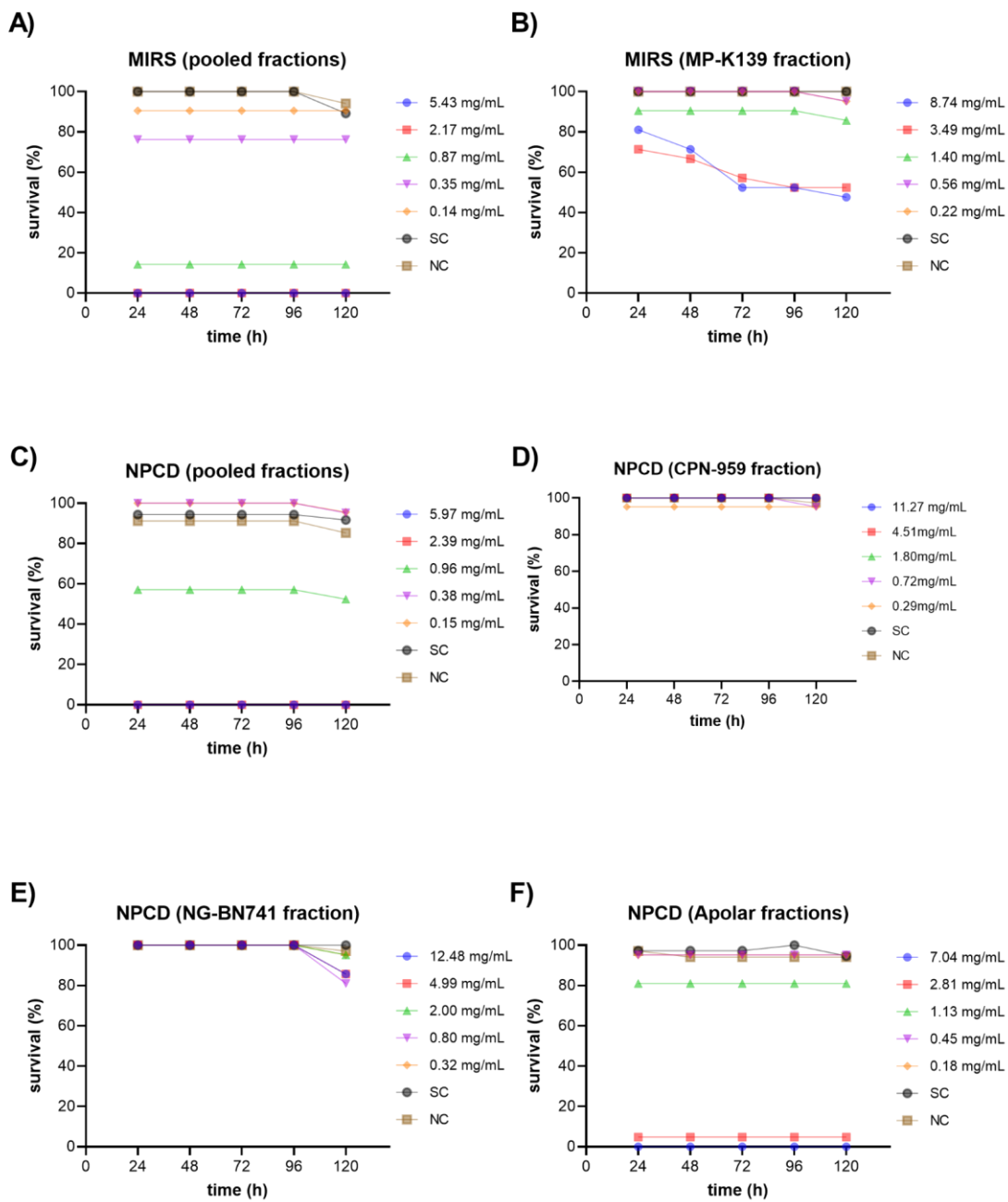
Figure A18. Survival (%) over 120 hour exposure of zebrafish larvae to (A) MIRS-04 pooled fractions; (B) MIRS-04 fraction #7 containing predominantly micropeptin K-139; (C) NPCD-01 pooled fractions; (D) NPCD-01 fraction #1 containing predominantly cyanopeptolin 959; (E) NPCD-01 fraction #3 containing predominantly nostoginin BN741; and (F) NPCD-01 apolar fractions #9+10.
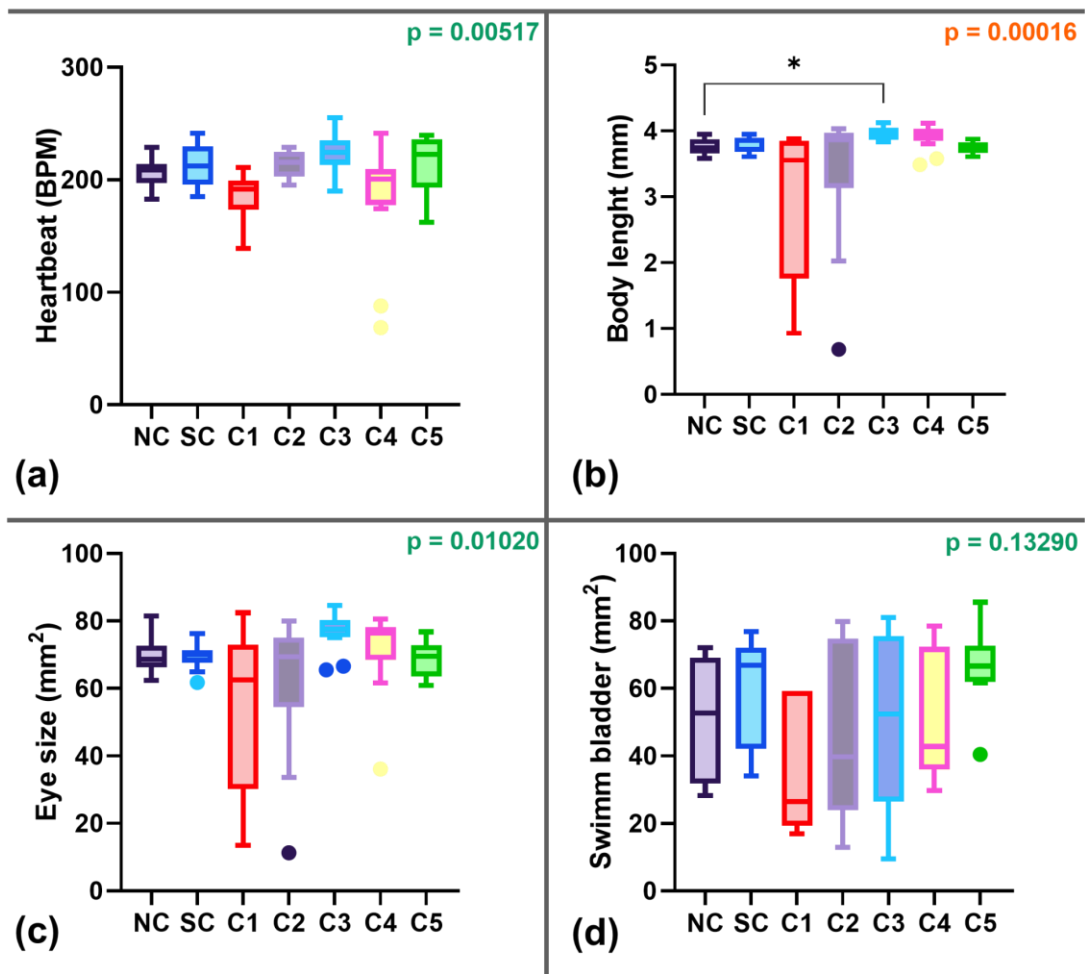
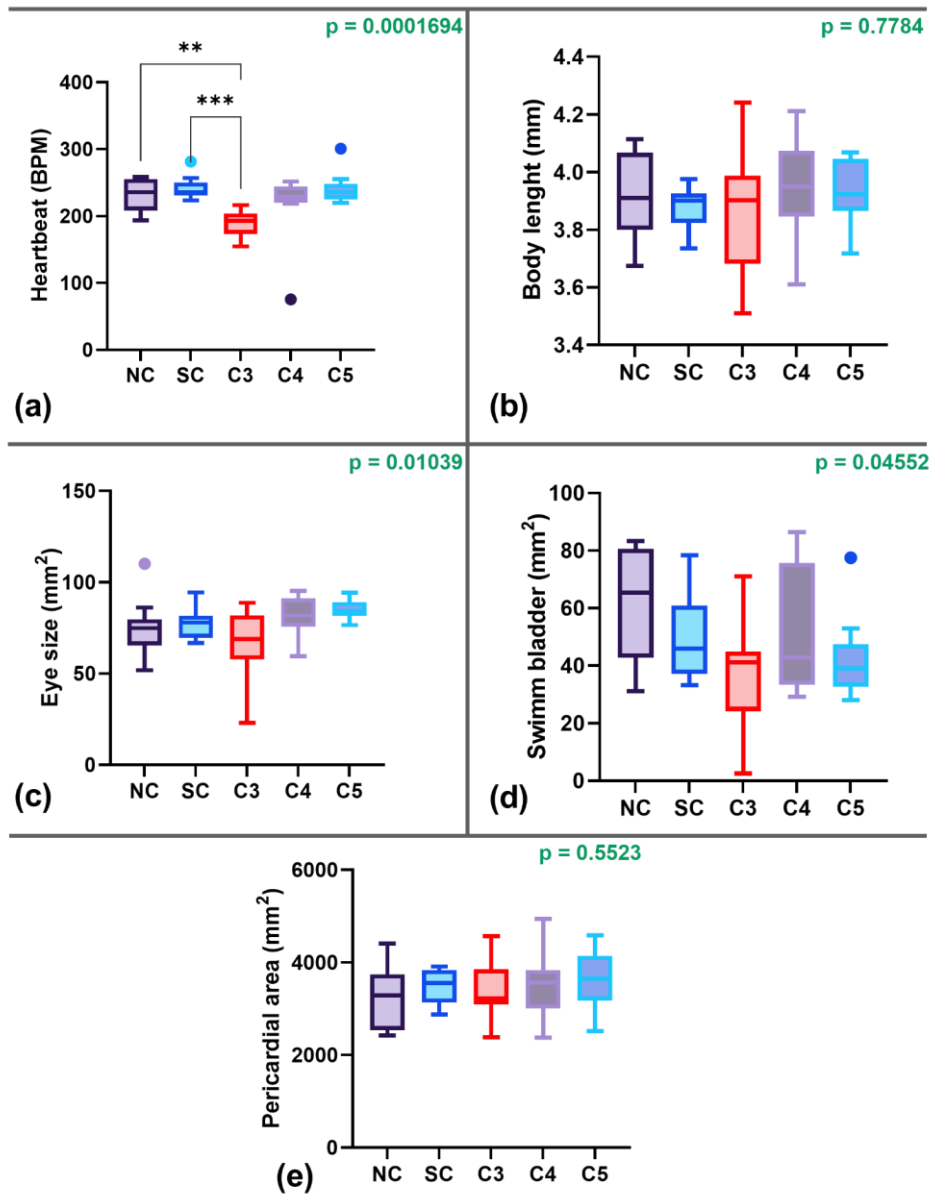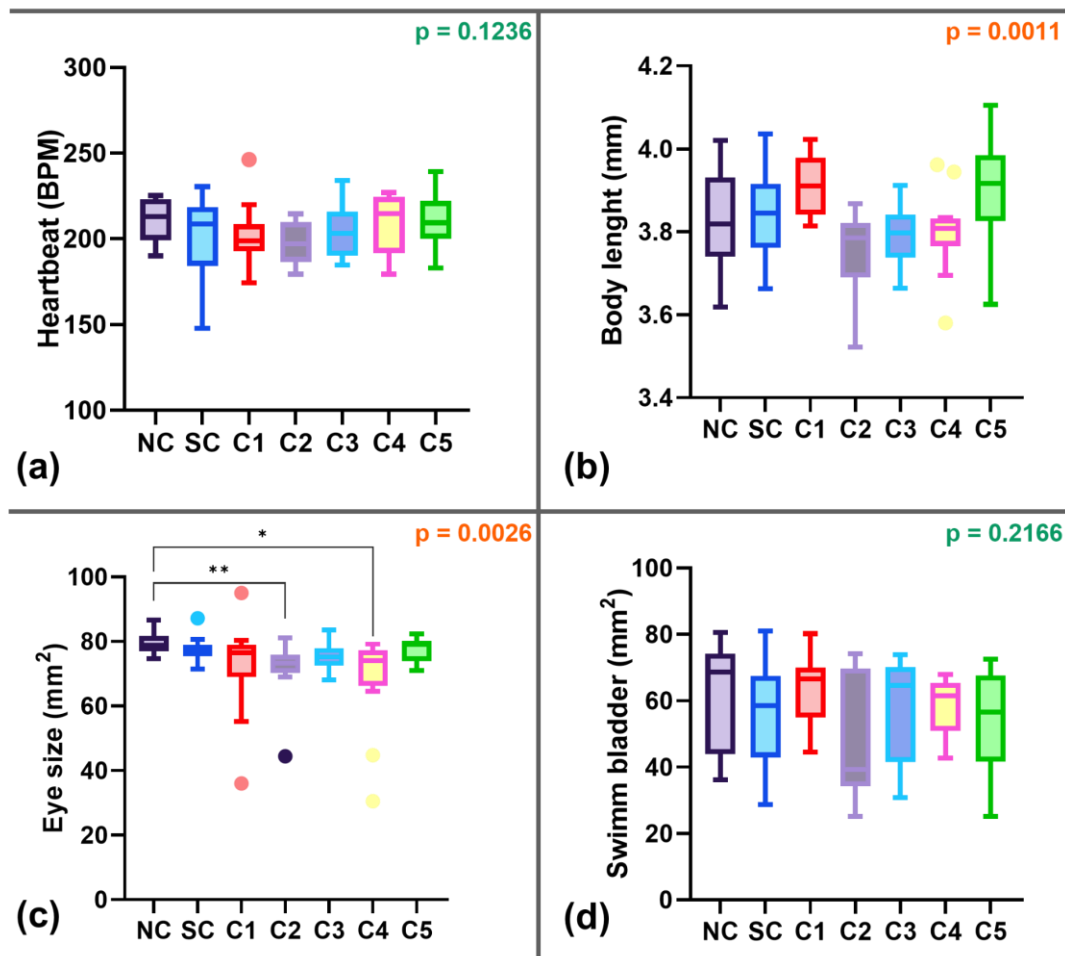Significant alterations between groups are indicated by asterisks (*p < 0.05); NC: Negative control; SC: Solvent control.

Figure A19. Heartbeat (a), body length (b), eye size (c), and swim bladder size (d) of larvae exposed (120 hour) to MIRS-04 micropeptin-K139-containing fraction (fraction #7).
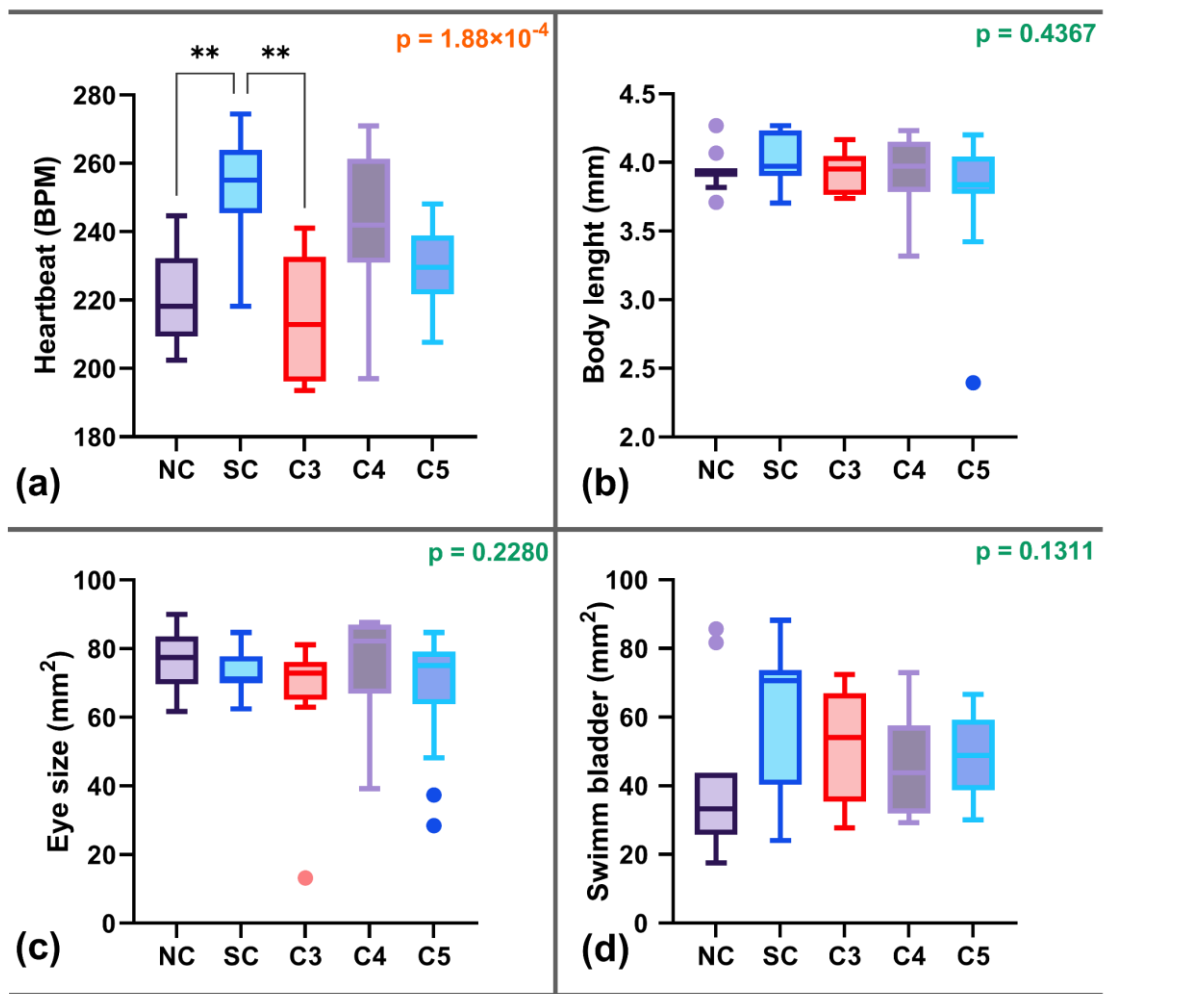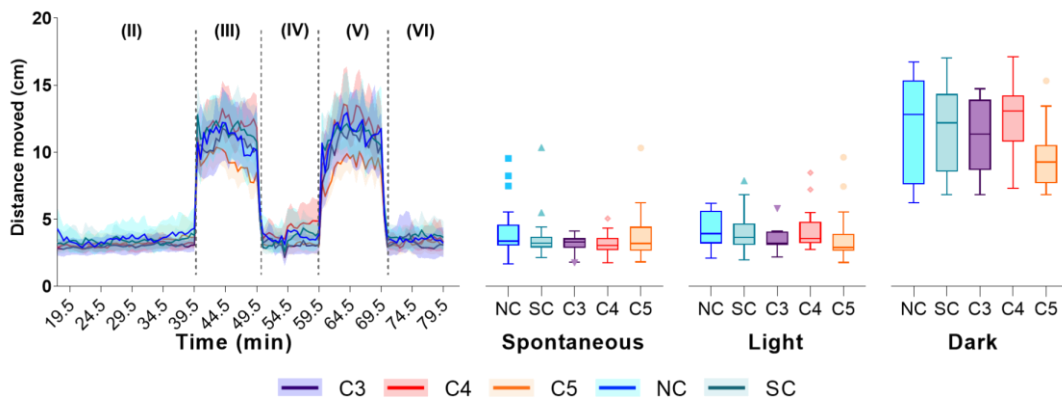
Significant alterations between groups are indicated by asterisks (*p < 0.05; **p < 0.01), NC: Negative control: SC: Solvent control.

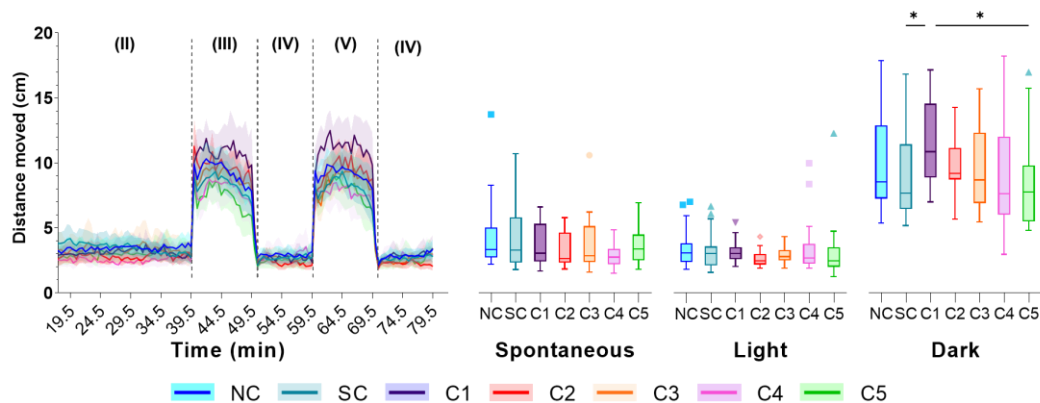**Figure A20.** Heartbeat (a), body length (b), eye size (c), and swim bladder size (d) of larvae exposed (120 hours) to NPCD-01 pooled fractions.

Significant alterations between groups are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001), NC: Negative control: SC: Solvent control.

Figure A21. Heartbeat (a), body length (b), eye size (c), swim bladder size (d), and pericardial area (e) of larvae exposed (120 hours) to NPCD-01 apolar fractions #9+10.
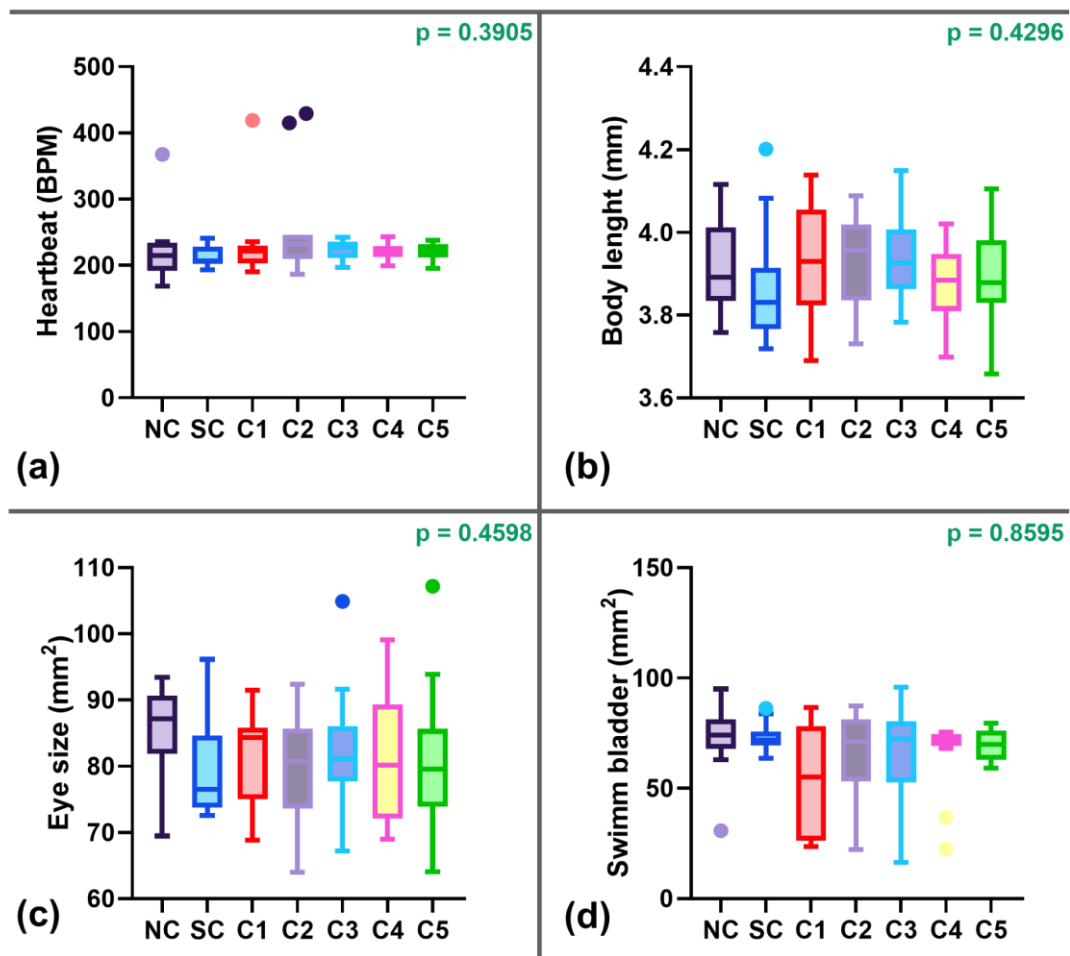
NC: Negative control: SC: Solvent control.

Figure A22. Heartbeat (a), body length (b), eye size (c), and swim bladder size (d) of larvae exposed (120 hours) to NPCD-01 cyanopeptolin-959-containing fraction (fraction #1).
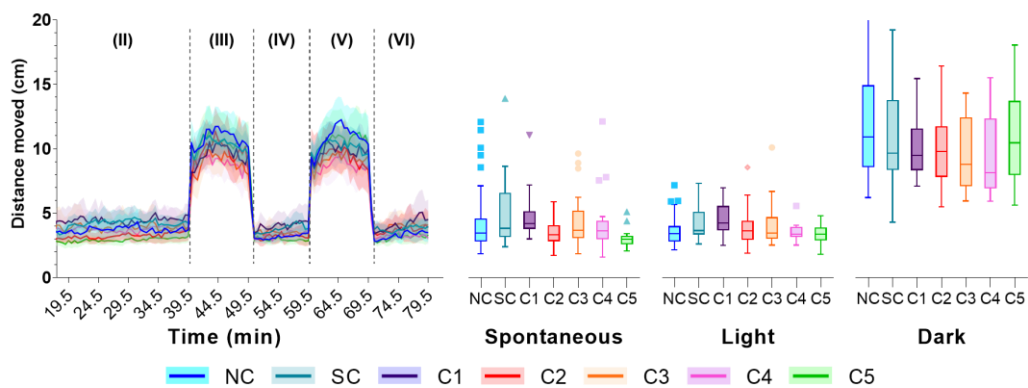
Significant alterations between groups are indicated by asterisks (*p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001), NC: Negative control: SC: Solvent control.

Figure A23. Heartbeat (a), body length (b), eye size (c), and swim bladder size (d) of larvae exposed (120 hours) to NPCD-01 nostoginin BN741-containing fraction (fraction #3).
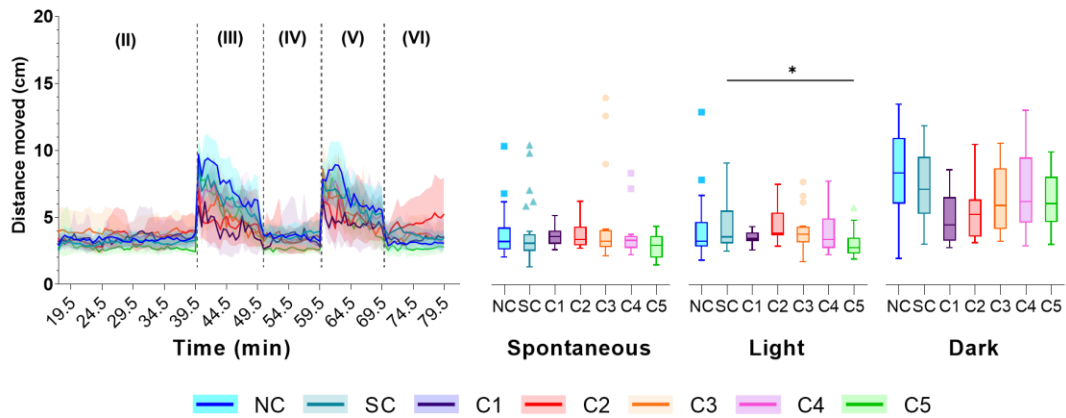
Significant alterations between groups are indicated by asterisks (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). NC: Negative control: SC: Solvent control. Data from dead and affected embryos was removed to perform the statistical analyses and plots.

Figure A24. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplots representing the average distance moved in each phase of the behaviour assay - spontaneous, light intervals, and dark intervals - of larvae exposed to different concentrations of MIRS-04 micropeptin-K139-containing fraction (#7) and the controls.

Significant alterations between groups are indicated by asterisks (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$). NC: Negative control: SC: Solvent control. Data from dead and affected embryos was removed to perform the statistical analyses and plots.

Figure A25. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplots representing the average distance moved in each phase of the behaviour assay - spontaneous, light intervals, and dark intervals - of larvae exposed to different concentrations of NPCD-01 pooled fractions and the controls.
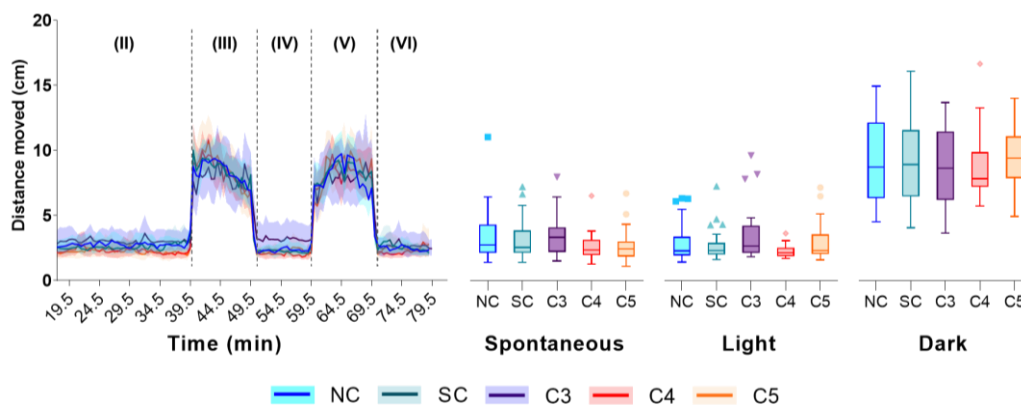
Figure A26. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplots representing the average distance moved in each phase of the behaviour assay - spontaneous, light intervals, and dark intervals - of larvae exposed to different concentrations of the cyanopeptolin-dominated fraction from the NPCD-01 strain .

Figure A27. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplots representing the average distance moved in each phase of the behaviour assay - spontaneous, light intervals, and dark intervals - of larvae exposed to different concentrations of NPCD-01 nostoginin BN741-containing fraction (#3) and the controls.

Figure A28. Time series of the accumulated distance moved (recorded every 30 seconds), and boxplots representing the average distance moved in each phase of the behaviour assay - spontaneous, light intervals, and dark intervals - of larvae exposed to different concentrations of fractions 9+10 (NPCD-01) and controls.