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-

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-
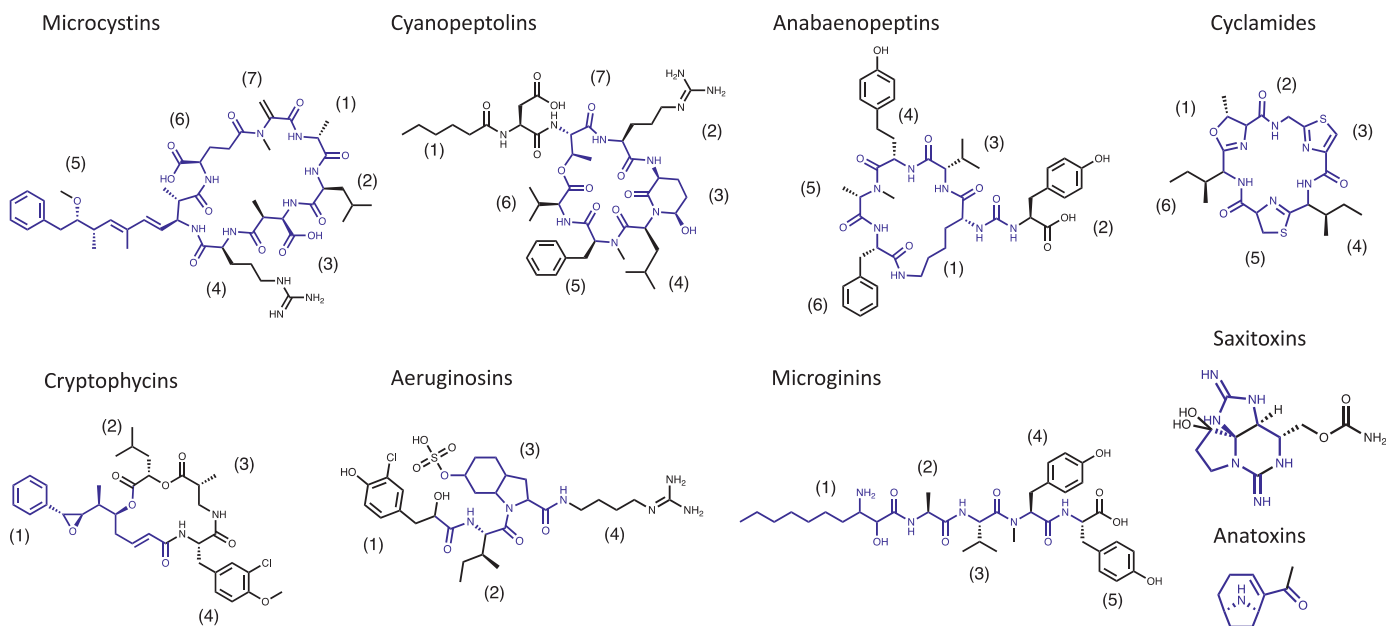
Fig. 1. Representative compounds for nine families of cyanobacterial secondary metabolites: microcystin-LR, cyanopeptolin A, anabaenopeptin A, aerucyclamide A, cryptophycin 1, aeruginosin 98A, microginin 713, saxitoxin, and anatoxin-a. Core-structures that are shared among most variants in the respective 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., (2*S*,3*S*,4*E*,6*E*,8*S*,9*S*)-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-
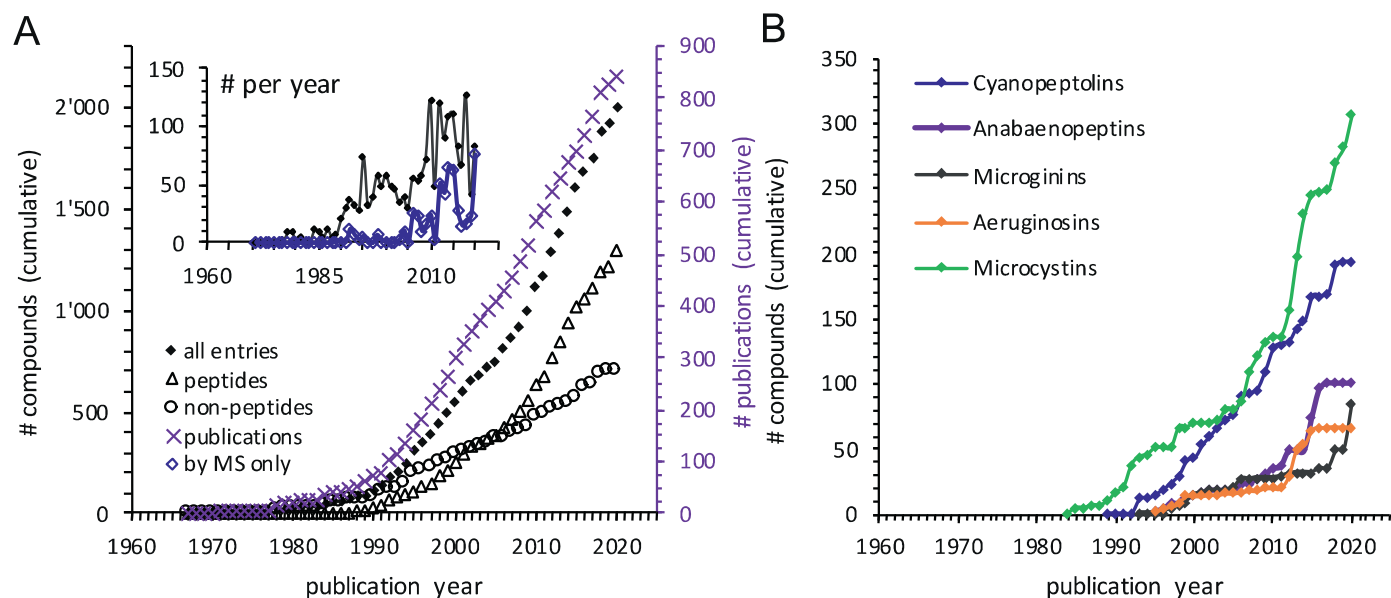
Fig. 2. (A) Cumulative number of: identified secondary metabolites from cyanobacteria (diamonds); peptide-based compounds (open triangles); non-peptide compounds (open circles), and; number of new publications (purple crosses, secondary y-axis). The inset shows the number of new compounds published each 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.
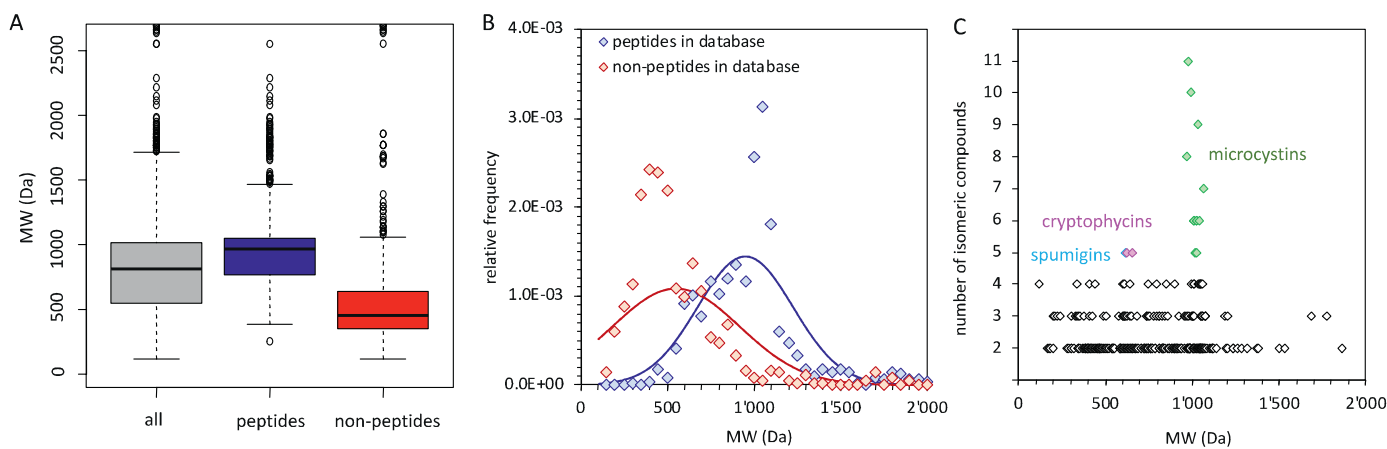
Fig. 3. Molecular weight distribution of cyanobacterial secondary metabolites: (A) Box plot of molecular weights for all metabolites (gray), metabolites containing at least one peptide bond (blue) and all other non-peptide-bond-containing metabolites (red); (B) Relative frequencies (%) of molecular weights 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.)
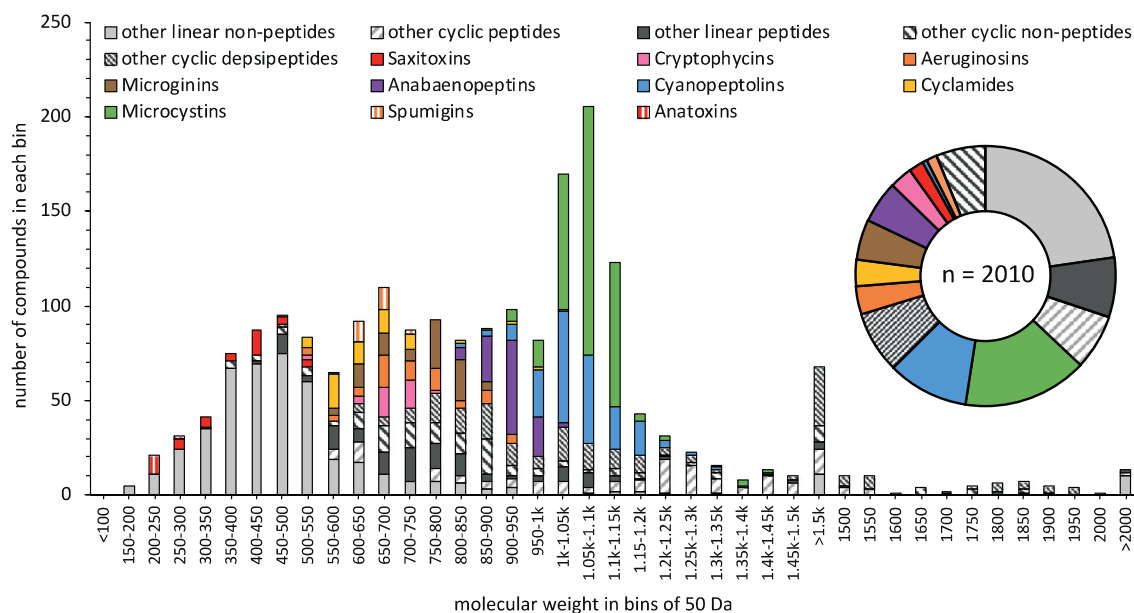
Fig. 4. Distribution of molecular weights for all cyanobacterial secondary metabolites reported in CyanoMetDB, ranging from 100 to more than 2000 Da with 50-Da bin size, showing the contribution of prominent metabolite classes and other linear and cyclic peptides and non-peptides, and the relative contribution 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 fulfil requirements 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-
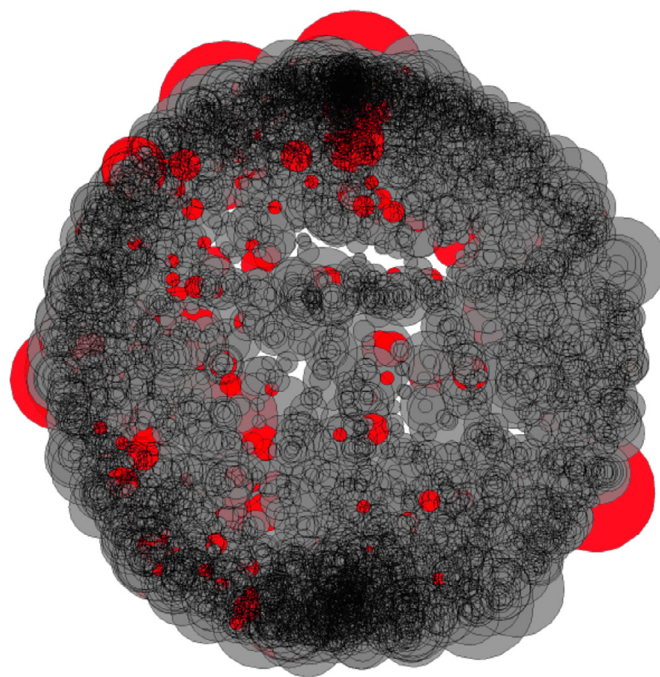
Fig. 5. Global view, highlighting positions of a preliminary set of 1640 secondary metabolites of CyanoMetDB from cyanobacteria (red) among the chemical space of more than 28,000 compounds in the Natural Products Atlas covering 239 out of 6800 nodes that indicate structural similarity (diameter scales 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 Met-Frag (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–ᴅ-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–quadrupole 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, D., 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.