Application of metabolomics for the detection and evaluation of cyanopeptide mixtures in Ontario and Quebec Lakes

by

Ethan Thomas McCann

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Science

in

Chemistry with Specialization in Chemical and Environmental Toxicology

Carleton University

Ottawa, Ontario

Abstract

The magnitude and frequency of cyanobacteria harmful algae blooms (cHABs) are increasing on a global scale. Anthropogenic nutrient enrichment, and climate change are drivers of bloom formation and favour cyanobacterial dominance in most aquatic ecosystems. The release of cyanotoxin mixtures by blooms into aquatic systems pose a risk to public and ecosystem health. The majority of cyanotoxin research has focused on specific groups: microcystins (MCs), cylindrospermopsins (CYNs), BMAA, saxitoxins (STXs) (anatoxins) ATXs. Other bioactive cyanopeptides (CNPs) produced by cHABs have received much less attention. The chemistry, toxicology, and environmental concentrations of the anabaenopeptin (APs), cyanopeptolin (CPs), microginin (MGs), cyanobactin (CBs), aeruginosin (ASs) CNP groups are largely unknown despite having notable biological activities and documented co-occurrence with cyanotoxins. Untargeted, targeted, and semi-targeted mass-spectrometry based metabolomic approaches were applied to study the CNP profiles of fifty-five cyanobacteria bloom samples collected from fifteen watercourses in Eastern Ontario and Western Quebec. 117 unique cyanopeptides were identified and ninety-four quantitated. MCs and ferintoic acid A (FA A) were quantified with reference materials and other CNPs were determined semi-quantitatively. CPs and APs displayed the greatest diversity in group variants and possessed similar concentrations and occurrence to that of MCs. Based on environmental concentrations and ubiquity of the CP and AP groups, they are recommended for prioritization in future toxicological and environmental research. Additionally, the apparent rise in cHABs, has drawn attention from paleolimnological researchers. Historical records of cHABs are limited, and proxies of cyanobacteria occurrence are typically restricted to pigments and DNA – both of which are sensitive to environmental degradation. However, MCs produced by freshwater cyanobacteria are stable cyclic hexapeptides. MCs in lake sediment archives can provide context to both the occurrence and toxicity of historical cHABs. To evaluate the use of MCs as a paleolimnological proxy, an MC extraction and quantitation method was

developed and validated for four MC congeners (MR RR, MC LR, [Dha^{7}]MC LR, MC LA) in lake sediments. The method was applied to sediment cores collected from the Rideau Canal system. The MC method was combined with multiple proxies of Itrax-XRF, chlorophyll-a analyses and radioscopic dating methods to develop a chronology of historical lake conditions. The multi-proxy approach provided a clear indication of increasing biological productivity towards the surface of both cores. The apparent trend of productivity appears to be a result of climate warming and anthropogenic nutrient enrichment.

Acknowledgements

I would like to extend my upmost gratitude and thanks to my supervisor, David McMullin, for the continued guidance and support, and providing the opportunity to develop and expand my chemistry skillset with natural products and cyanobacteria metabolites. I am very grateful for the opportunity to attend and present at 12th international conference on toxic cyanobacteria (ICTC) in Toledo, OH – It was a valuable experience. Thank you to Kim McDonald (MSc.) for extracting the bloom samples analyzed here and, imparting knowledge of lab operations and all things cyanobacteria. Thank you to Dr. Justin Renaud for his analytical expertise with high-resolution mass-spectrometry and metabolomics – the knowledge of analytical chemistry I gained was monumental. Thank you to Dr. Mark Sumarah for imparting both his academic guidance and expertise with the cyanopeptide bloom project. Thank you to Dr. Jesse Vermaire for collection of Rideau sediment cores, guidance with Rideau core project, and procurement of Chl-a and gamma spectroscopy data, used here. Thank you to Dr. Frances Pick for the collection and detailed information for the cyanobacteria bloom samples collected in the National Capital Region and analyzed here. Thank you to Dr. David Miller for his guidance and providing the initial reference for this master's opportunity. Thank you to Dr. Tyler Avis for both his guidance and assistance with laboratory operations and statistics. I thank Dr. Tim Patterson for the experience and knowledge of sediment core collection and the opportunity to preform such work in New Brunswick lakes. Thank you to Dr. Nawaf Nasser for imparting his knowledge of *R* for ^{14}C age-depth modelling, the workshops were fantastic. I would like to thank Robert Banes (BSc.) for assisting with the preparation and extraction of Rideau sediment core sub samples. Thank you to Dr. John Smol for preforming the gamma spectroscopy and Chl-a analysis of the Rideau sediment cores, used here.

My cherished lab mates: Alex, Catrina, Hailey, and Keri I'm very appreciative for the timeless lab experiences, brainstorms, laughs and headshakes – you made grad school an enjoyable

experience overall and kept me sane during writing, thank you. Thank you to my better half, Rachel Clouthier, and my family (Tom, Angela and Hilary McCann) for the never-ending love and support through the ups and downs.

Table of contents

List of abbreviations:

%SSE — percent signal suppression and enhancement

(HR)-MS/MS — high-resolution tandem mass spectrometry

A/Ala — alanine

AC — aerucyclamide

ACE — angiotensin converting enzyme

ACE — angiotensin converting enzyme

AChE — acetylcholinesterase

ACN — acetonitrile

ACP — acyl carrier protein

Adda — 3-amino-9-methoxy-10-phenyl-2,6,8-trimethyldeca-4,6-dienoic acid

AEG — N-2-aminoethyglycine

Ahda — 3- amino-2-hydroxy decanoic acid

Ahoa — 3-amino-2-hydroxy octanoic acid

Ahp — 3-amino-6-hydroxy-2-piperidone

ALS/PDC — amyotrophic lateral sclerosis/Parkinsonism dementia complex

ATXa — Anatoxin-a

AOAC — Association of Official Analytical Chemists

AP — anabaenopeptin

APCI — Atmospheric-pressure chemical ionization

APM — aminopeptidase M

APM — aminopeptidase M

AS — aeruginosin

AT — acyltransferase

ATX — anatoxin

BAMA — β-amino-N-methylalanine

BMAA — β-methylamino-alanine

BW — body weight

C/Cys — cysteine

cAB — cyanobacteria algae bloom

CB — cyanobactin

cHAB — cyanobacteria harmful algae bloom

Chl-a — chlorophyll-a

Choi — 2-carboxy-6-hydroxyoctahydroindole

CIR — ratio of coherent/incoherent backscatter

CNP — cyanopeptide

CO_2 — carbon dioxide

CONISS — constrained incremental sum of squares

CP — cyanopeptolin

CPA — carboxypeptidase A

CPCC — Canadian Phycological Culture Collection

cps — counts per second

CRS — constrained incremental sum of squares

CYN — cylindrospermopsin

CYP_{450} — cytochrome P450

DAB — 2,4-diaminobutanoic acid

DART — direct analysis in real time

DCM — dichloromethane

DDA — data dependant acquisition

ddH_2O — distilled deionized water

DESI — desorption-ESI

DFF — diagnostic fragmentation filtering plot

DH — dehydratase

Dhb — dehydrobutyrine

DIA — data independent acquisition

DmA — 1,1-dimethylallyl

d.w. — dry weight

E/Glu — glutamic acid

EI — electron ionization

ELISA — enzyme-linked immunosorbent assay

ER — enoyl reductase

ESI — electrospray ionization

F/Phe — phenylalanine

FTICR — Fourier transform ion−cyclotron resonance

G/Gly — glycine

Ga — giga annum

GNPS — global natural products social

GV — guideline value

H/His — histidine

H_2-ATX— dihydroantoxin-a

hATX — homoanatoxina

HBGV — health based guideline values

HBV — health based reference values

HCO_3 — bicarbonate

Hph — homo-phenylalanine

Hpla — hydroxy-phenyllactic acid

HPLC — high preformance liquid chromatography

hTAX — homoantoxin

Hty — homo-tyrosine

HxA — hexanoic acid

I/Ile — isoleucine

ICP — inductively coupled plasma

K/Lys — lysine

KR — ketoreductase

KS — ketosynthase

KW — Kruskal Wallis test

L/Leu — leucine

LBSM — laboratory blank sample matrix

LC — liquid chromatography

LC-FLD — liquid chromatography fluorescence detector

LFB — Laboratory fortified blank

LFSM — laboratory fortified sample matrix

LGL — Laurentian Great Lakes

LOAEL — Lowest-observed-adverse-effect level

LOD — limit of detection

LOQ — limit of quantification

LWTXs — Lyngbya (Microseira) wollei toxins

M/Met — methionine

m/z — mass-to-charge ratio

MA — microcyclamide

MALDI — matrix assisted laser desorption ionization

MC — microcystin

Mdha — methyl-dehydroalanine

Mdhb — 2-(methylamino)-2-dibutyric acid

MECP — Ministry of the Environment, Conservation and Parks

MeOH — methanol

MG — microginin

MMPB — methyl-3-methoxy-4-phenylbutanoic acid

MN — Molecular network(-ing)

MP— micropeptin

MS — mass spectrometry

MS/MS — tandem mass spectrometry

MV — microviridin

N/O-Me — N-/O-methylated

NCE — normalized collision energy

NOAEL — No-observed-adverse-effect level

NOD — nodularin

NRP — non-ribosomal peptide

nRPS — non-ribosomal peptide synthase

OATPs — organic anion transporting polypeptides

OS — oscillaginin

P/Pro — proline

PCA — principal component analysis

PCP — peptidyl carrier protein

PKS — polyketide synthase

PMTs — post translational modifications

PP1/2A — protein phosphatase 1/2A

Ppant — 4’-phosphopantetheine group

PSP — paralytic shellfish poisons

PTFE — polytetrafluoroethylene

PVG — provisional guideline values

Q/Gln — glutamine

qTOF — quadrupole time-of-flight

R/Arg — arginine

RiPP — post-translationally modified peptide

RRHD — rapid resolution high definition

RSD — relative standard deviation

RT — retention time

S/Ser — serine

sp. — species

SPE — solid phase extraction

spp. — multiple species

SSR — sequential sample reservoirs

STX — saxitoxin

T/Thr — threonine

TEF — toxicity equivalence factor

Thz — thiazole

Tzc — thiazole carboxylate

Tzl — thiazoline

UPLC — ultra performance liquid chromatography

V/Val — valine

W/Trp — tryptophan

XRF — X-ray fluorescence

Y/Tyr — tyrosine

List of figures:

List of Tables:

CHAPTER 1: Literature Review

1.1. CYANOBACTERIA: BIOLOGY AND NATURAL HISTORY

Cyanobacteria, known by their common name “blue-green algae” are a phylum of Prokaryotic phototrophic microalgae which comprise a portion of natural phytoplankton communities in surface waters. They occupy illuminated environments, the most common being aquatic systems such as, freshwater, brackish and marine environments. Unlike most Eukaryotic phototrophic microalgae species cyanobacteria belong to the taxonomic domain of bacteria, distinctly unique in comparison to their Eukaryotic counterparts (i.e., diatoms, chlorophytes, dinoflagellates and chrysophytes) and therefore are not typically referred to as algae [1–4]. Rather, cyanobacteria are Gram-negative Prokaryotes having no membrane bound nucleus or organelles and reproduce asexually. However, cyanobacteria possess intercellular microcompartments such as carboxysomes for concentrating carbon dioxide (CO_2), and a thylakoid membrane system for photosynthesis and respiration [3]. Cyanobacteria must fix CO_2 for photosynthesis and dissolved CO_2 can be depleted in surface waters with significant planktic cellular growth. Carbon concentrating mechanisms enable cyanobacteria to increase the CO_2 concentrations in carboxysomes to levels at which the C-fixing enzyme Rubisco can operate effectively [1]. Five distinct inorganic carbon uptake systems have been identified for cyanobacteria to date, two for CO_2 and three for bicarbonate (HCO_3) uptake – each with independent substrate affinities and fluxes [1]. Cyanobacteria are among the very few Prokaryotic taxa that can perform oxygenic photosynthesis and respiration simultaneously in the same compartment [5]. Cyanobacteria photosynthetic electron transport occurs in the non-continuous compartmental thylakoids, whereas the respiratory electron flow takes place in both the thylakoids and cytoplasmic membrane system [5]. Cyanobacteria species produce a variety of photosynthesis pigments to harvest light, *Microcystis aeruginosa* strains exhibit the eponymous blue-green hue from the

pigments chlorophyll a and phycocyanin. Contradictory to their common name, most cyanobacteria are not blue-green in colour rather, many species produce the red pigment phycoerythrin and mixtures of yellow-orange carotenoids (i.e. echinenone, myxoxanthophyll) generating a wide range of colours, including various shades of green, red, brown, yellow and pink [1,6]. The presence of accessory pigments specifically, phycobiliproteins (i.e., phycoerythrin – blue; phycocyanin - red) enable cyanobacteria to capture light energy in a greater range of the visible spectrum (400 to 650 nm) [7]. Phycobiliproteins are components of light harvesting phycobilisomes complexes embedded in the thylakoid membranes of cyanobacteria [7,8]. Phycobilisomes act as light harvesting antennas which absorb and transfer energy to chlorophyll protein complexes and the reaction centers of Photosystems I and II, within the thylakoid membrane [7,9]. Some cyanobacteria genera (e.g., *Prochlorothrix, Prochlorococcus, Prochloron*) lack phycobilisomes and possess chlorophyll-b instead [10].

Like other phytoplanktonic species, cyanobacteria are globally ubiquitous inhabiting a multitude of ecological niches including fresh, marine and brackish-waters, soil, biological soil crusts, snow, etc. [1,2,11]. Interestingly, cyanobacteria can have symbiotic associations with different host organisms, commonly in extremely stressed conditions like volcanic ash, desert crusts, high altitudes and anthropogenically disturbed areas [11,12]. Morphologically, cyanobacteria exist as unicellular, colonial, or filamentous forms consisting of spherical, rod and spiral cellular shapes [1,13]. In the past cyanobacteria taxa were designated into five orders (i.e., Chroococcales, Pleurocapsales, Oscillatoriales, Nostocales and Stigonematales) by morphological and physiological characteristics [13]. More recent phylogenetic analyses do not reflect the previous morphology assembled taxonomy. Therefore, mislabeling of cyanobacteria into polyphyletic species and genera will require revisions in the future. Cyanobacteria maintain an extensive natural history, having persisted on Earth for ~3.5 billon years, photosynthetically contributing to the first atmospheric oxygenation event and playing a large role in shaping the composition of

Earth's crust, oceans, and atmosphere [14,15]. The Precambrian eons have been loosely termed the "the age of blue–green algae" [16]. Although, there is still minor controversy on this subject there is no doubt that cyanobacteria are ancient organisms. The long evolutionary history of cyanobacteria has provided them with many biological adaptations to persist in a variety of pelagic zones.

1.2. BIOLOGICAL ADAPTATIONS

Cyanobacteria possess broad ecological tolerances, contributing to their competitive success in a variety of environments and throughout time. Largely, this is due to their possession of highly adaptable eco-physiological traits that enable competitive exclusion of other microalgae during favorable conditions, further exaggerated by a warming climate [17]. Sunlight access, nutrient availability (carbon: nitrogen: phosphorus; C:N:P), and water temperature are the three major factors which influence cyanobacterial cellular growth. There are six advantageous eco-physiological traits, which can be specific to the genus level of cyanobacteria, (1) gas vesicle buoyancy regulation, (2) high affinity and storage of P, (3) N-fixation, (4) akinete production and, (5) light capture and, (6) increased cellular growth at warmer temperatures [18]. Most cyanobacteria species regulate their buoyancy with gas vesicles, being able to access both illuminated surface waters and nutrient rich benthic waters, most commonly in stratified lakes [1,18]. Overall, the vertical stratigraphy of cyanobacteria communities depends on the catchment hydrology, nutrient loadings and hydrodynamics of an aquatic system [18,19]. Buoyancy is dependent on light, regulated by accumulation carbohydrates during daylight photosynthesis and subsequent respiration of these products at night. In aquatic systems with heavy nutrient loading and increased vertical stratification, planktic communities shift towards buoyant fast-migrating cyanobacteria species which can efficiently access both the surface and benthic waters for resource exploitation[18,19]. The buoyancy regulation of cyanobacteria is species dependant although, expressed as common trait in bloom forming genera (i.e., *Microcystis* and *Dolichospermum*), well-adapted to systems with

stronger stratification regimes [18]. Changes in climate, such as heightened temperatures, prolonged droughts and longer water residence times are hypothesized to increase the duration of stratification and promote cyanobacteria dominance of most bloom forming genera.

In natural freshwaters P is a limiting nutrient, cyanobacteria have overcome this problem – producing phosphatases which hydrolyse phosphate to bio-available forms and sequester excess P intracellularly for storage in polyphosphate bodies [3,18,20]. For both these reasons, cyanobacteria are able to dominate during low and high P availability. In low nutrient conditions, cyanobacteria's high affinity for P allows them to outcompete other phytoplankton whereas, high P availability results in greater phytoplankton biomass dominated by cyanobacteria in lakes [18].

Nitrogen-fixation (N-fixation) is physiological adaptation specific to only some filamentous cyanobacteria genera (i.e., *Dolichospermum*, *Aphanizomenon*, *Nostoc*, *Cylindrospermopsis* and *Planktothrix)* providing a competitive advantage under low available N conditions [3,18,21]. N-fixing species produce specialized nitrogen-fixing cells called heterocysts which can metabolize N_2 gas and nitrates from the atmosphere and water [3,21]. Heterocysts are used for N_2 fixation as the nitrogenase enzyme required to catalyze the reaction is irreversibly activated by oxygen and must be spatially separated from photosynthetic processes [22]. Inside heterocysts sequestered nitrogen is reduced to simple amines and subsequently consumed for incorporation into complex biomolecules [22,23]. N-fixation is costly, requiring a significant amount of energy to both break the triple bond linking N_2 molecules and for maintenance of heterocysts [18,22]. However, energetic investments into N-fixation may be offset by the competitive advantage provided in severely N deficient environments.

To ensure survival in benthic sediments some cyanobacterial taxa can produce dormant cells, akinetes, with thick-walled, non-motile resting spores [24]. Akinetes will germinate when conditions are favorable, being known to survive temperatures up to 55°C, and remain viable up to 64 and hypothetically >100 years after deposition [18,25,26]. Akinetes production is a long-term survival and

ecological advantage processed by some cyanobacteria species (i.e., *Nodularia, Aphanizomenon, Dolichospermum*).

Cyanobacteria are regarded as strong light competitors in planktic communities due to their accessory pigmentation and phycobillosomes. Experimentally, *Synechocystis* sp. have shown to attenuate light at lower levels than other tested species and readily outcompete other phytoplankton species [27]. However, this is not consistent across all cyanobacteria genera. *Microcystis* spp. have a poor light attenuation efficiency in comparison to other cyanobacteria and phytoplankton species (e.g., *Aphanizomenon* and *Chlorella*) [28,29]. The differences in light interception among different species is a result of morphological characteristics. Filamentous cyanobacterial species appear superior to larger colonial species at harvesting light [28]. Although, dominance of filamentous cyanobacteria is not solely related to light competition, it is hypothesized that light attenuation may synergistically act with other eco-physiological adaptations to favor certain genera under specific climate scenarios [18]. In contrast, colonial genera are prone to photoinhibition with increasing sunlight iridescence [30]. The dominance of colonial cyanobacteria is more dependant on the water column stratigraphy and their buoyancy-dependent niches.

Increased temperatures mainly favor a shift in phytoplankton communities towards cyanobacteria, relative to other types of microalgae. The species of *Aphanizomenon flos-aquae, Planktothrix agardhii*, *M. aeruginosa* and *Synechococcus* spp. reach their maximum growth rates at 20, 20, 28 and 41°C, respectively [29,31]. Furthermore, *M. aeruginosa* has the fastest growth rate recorded of any cyanobacteria and Eukaryotic microalgae species, dominating over other phytoplankton when water temperatures increase above 20°C [18]. Cyanobacteria will dominate freshwater microalgae populations at higher temperatures due to their growth rate optimums and increased affinity for nutrients. The determinants of cellular growth are highly variable and dependent on water depth, solar radiation, hydrological regimes and windspeed. For example, shallow waters

paired with high surface temperatures, low precipitation and wind can result in more prolific cyanobacteria growth [32].

1.3. INCREASING OCCURRENCE AND CONCERN

Although most species of microalgae propagate and persist quite rapidly under warm eutrophic conditions, their biomass will not accumulate to the extremely high cell densities observed with cyanobacteria species [1,32]. The accumulation of cyanobacteria biomass to high cell densities creates an observable green scum at a waterbody's surface and shorelines - commonly referred to as a bloom event. Increased occurrences of cyanobacterial blooms in the past few decades are attributed to anthropogenic nutrient enrichment and climate change. The Laurentian Great Lakes (LGLs) located on the Canada-United States border are one of the largest freshwater reservoirs worldwide. Leading up to the 1960s the LGL region was a dumping ground for a variety of anthropogenic pollution, most evident in industrialized western basin of Lake Erie the shallowest of the LGLs. Throughout the mid 1900s cHABs were a common occurrence due to artificially enriched levels of N and P, from industrial and agricultural sources [33]. The first call for LGL water quality protection and environmental regulations is associated with a large surface water wildfire on a major tributary of Lake Erie. June 22nd, 1969, the Cuyahoga River caught fire, erupting in an immense blaze due to its polluted waters, saturated with sewage and industrial waste [33]. The devastating event drew galvanizing public support and national concern, setting the stage for implementation of the Clean Water Act (CWA) and Great Lakes Water Quality Agreement (GLWQA) of 1972 [33]. Both the CWA and GLWQA implemented the federal framework necessary to regulate pollutant discharge and identify restorative actions.

By the late 1960s it was understood that excess dissolved P and N from anthropogenic sources was the major driver in algal bloom growth and proliferation – and annual total phosphorus (TP) loading to the lakes was decreased. Lake Erie saw a reduction in nearly 30,000 t (33,069 tn) of TP in the late-1960s [34]. Additionally, it frequently experienced bloom events leading to the

introduction nutrient limiting regulations in 1972 as part of the GLWQA [33]. Improvements in water quality and reductions in blooms continued through the 1970s and 1980s [33,35]. However, prolific cyanobacteria blooms in the LGL region began to re-occur in the mid-1990s [36], and continued with annual prolific blooms coinciding with favorable climatic and hydrological conditions. For example, an infamous 2014 bloom caused the Toledo water crisis and a subsequent 2015 bloom in western Lake Erie reached a size of 780km^2, both are attributed to a large amount of early summer runoff from agricultural lands followed by prolonged hot and dry summer conditions [34,37].

It is currently hypothesized that the increased frequency and proliferation of blooms is linked to a warming climate. Elevated surface water temperatures and hydrologic variations from changes in climate have been shown to alter planktic community composition in favor of cyanobacteria over their eukaryotic microalgae [21,38–40]. Several studies investigating the relationship of cyanobacteria and climate have indicated compounding environmental factors that will synergistically favour cyanobacteria dominance [1,18,21,41]. Rising surface temperatures, increased atmospheric CO_2, and enhanced vertical stratification in aquatic systems all contribute to the dominance of bloom forming cyanobacteria genera[1,18,41]. Globally, the occurrence of cyanobacterial blooms is increasing both in incidence and severity [1,32]. In Canada, aquatic systems which previously did not experience blooms are now having more regular seasonal occurrence. For example, Lake Superior (the largest LGL) has experienced consistent blooms since 2010 with the most prolific events in the hot summers of 2012 and 2018 [42,43]. These bloom events were historically uncommon and surpassed previous Lake records [42,43]. Both nutrient enrichment and climate change heighten, water temperatures, thermal stratification, and nutrient availability in lacustrine systems – positively influencing cellular cyanobacteria growth.

Cyanobacteria blooms are viewed as harmful to for two reasons, (1) they produce and release mixtures of toxic/bioactive metabolites and (2) generate anoxic zones within their resident aquatic environment. Upon the senescence of a cyanobacteria bloom the decomposition of algal biomass

increases the biological oxygen demand of benthic microbes generating vertical dissolved oxygen gradients throughout the water column and ultimately creating anoxic zones. Anoxic effects can be exaggerated in stagnant waters where cyano-biomass can accumulate and persist until senescence. However, the primary concern of cyanobacteria blooms are the potent cyanotoxins and other bioactive secondary metabolites many taxa produce[44,45]. There are three general toxin classes in which cyanotoxins can be lumped into, (1) neurotoxins; (2) hepatotoxins; (3) dermatoxins. Nero- and hepatotoxins are of greatest concern due to their potency. Anoxic and toxic effects from cyanobacteria blooms are detrimental to aquatic resources – specifically, local economies, environmental industries, water quality and ecosystem health. Therefore, toxin producing cyanobacteria blooms are ascribed as cyanobacterial harmful algal blooms (cHABs). Anoxic conditions and toxin production by cHABs are linked to fish and wildlife mortalities. For example, during the early summer of 2020 330 Elephants in Botswana, SA died as a result of ingesting standing water from a natural pool containing a toxic cyanobacterial bloom[46]. Human poisoning from cHABs have been documented on all continents except Antarctica and reports of human cHAB poisonings have increased overtime, a likely result of heightened bloom occurrence and monitoring programs. However, confirmation of human deaths from cHABs are rare - limited to a single cohort exposed through renal dialysis of contaminated water at a haemodialysis center in Caruaru, Brazil, 1996.

1.4. BIOSYNTHESIS: Cyanobacteria Secondary Metabolite

Metabolism in microorganisms can be divided into two types: primary metabolism, which encompasses bio-chemical reactions and pathways necessary for survival and reproduction, and secondary metabolism, which fulfills a variety of advantageous functions, mainly acting to improve the survival and fitness of the producing organism in competition. Secondary metabolites, or commonly, natural products are often small molecules (< 2000 Da) mainly associated with bacteria, fungi and plants [47]. Biosynthesized secondary metabolites are classified by their route

of synthesis, forming several broad classifications [48]. Cyanobacteria produce a rich variety of these groups such as, (1) terpenoids and steroids, (2) alkaloids, (3) PKs (polyketides), (4) NRPs (non-ribosomal peptides) and RiPPs (ribosomally synthesized post-translationally modified peptides) [13,48]. The majority of primary metabolite functional pathways (i.e., carbohydrate, energy, amino acid and nucleotide metabolism enzymes) are highly conserved between microorganisms [49]. However, secondary metabolism does not display good evolutionary coherence between microorganisms, more commonly being genus or strain specific. Microbial natural products are mainly of interest due to their application in the pharmaceutical and chemical industries [48]. Approximately 25% of all microbial natural products are bioactive – these include, antibiotics, antitumor agents, cholesterol-lowering drugs, immunosuppressants, anthelmintic agents, antiparasitics, herbicides, ruminant growth stimulators, agricultural fungicides, bio-insecticides, and many others [48,50,51]. Moreover, roughly 75% of all antibiotics have been obtained from filamentous prokaryotes like cyanobacteria [51]. Most cyanobacterial secondary metabolites are oligopeptides or possess peptidic substructures [52–54]. The majority of these oligopeptides are synthesized by NRPS (non-ribosomal peptide synthetase) or NRPS-PKS (polyketide synthase) hybrid pathways[52,53]. A lesser number of cyanobacterial metabolites are produced by the RiPP pathway such as the chemical groups of CBs (cyanobactins) and MVs (microviridins) [52,55,56]. Overall, CNPs (cyanopeptides) derived from NRPS, NRPS-PKS and RiPP pathways account for the vast majority of reported metabolite diversity in cyanobacteria [13,53,55,57]. More than 2000 cyanobacteria natural products have been described to date [53]. Some other common biosynthetic pathways found in cyanobacteria produce alkaloids or terpenes [13,53]. The chemical diversity of secondary metabolites produced by cyanobacteria is a consequence of the numerous biosynthetic pathways and their variable molecular machinery used to produce these metabolites

1.4.1. Non-Ribosomal Peptide (NRP) biosynthesis

Non-Ribosomal Peptide synthesis (NRPS) pathways consists of a modular structure analogous to a linear assembly line, each module is responsible for selection of specific amino acid sequence and elongation of a peptide chain [57,58]. The combination of modules and their subsequent domains determines the structure of the final peptide and its biological activity [59]. There are distinct catalytic modules in NRPS responsible for each specific biosynthetic step (i.e., initiation, elongation, modification, and termination) of the peptide chain, their order matches the number and order of amino acids incorporated into the final peptide product [58,59]. Three essential domains are required within a module: the A (adenylation)-domain, responsible for selection, activation, and transfer of the selected amino acid monomer; PCP (peptidyl carrier protein)-domain, which bears the 4'-phosphopantetheine (Ppant) group responsible for transport of bound substrates and elongation intermediates; C (condensation)-domains which catalyze the transfer of PCP-bound amino acids to the nascent peptide. Further to the three domains mentioned, non-essential M (modification)-domains exist which can further modify the peptide sequence. For example, epimerization (E), formylation (F), methylation (M), heterocyclization (Cy), reduction (R), and oxidation (Ox) can occur at M domains [58,60,61]. M-domains are usually integrated at the C-terminal region of the A-domains within NRPS modules [58,61].

The first step, activation, is catalyzed by the A-domain and must be highly selective and specific toward the incoming amino acid monomer. The A-domain is a gatekeeper for incoming amino acids containing a specificity-conferring code which is responsible for substrate binding within the active site and ultimately determines the sequence of the final peptide [61]. Each A-domain contains a large N-terminal core domain and a small C-terminal subdomain which are connected by a 5-10 residue hinge [57,60]. The terminal and subdomain amino acid sequences are highly conserved throughout all adenylating modules [60]. Within each module, the A-domain first catalyzes the selection of a specific amino acid monomer with an ATP molecule and a Mg^{2+} cofactor, creating

an adenylating intermediate (commonly with aminoacyl-AMP) the activated amino acid is then transferred onto the PCP domain; however, the direct mechanism for this remains unknown [60]. Studies have observed plasticity in substrate specificity of NRPS A-domains resulting from degenerate non-ribosomal codes [61,62]. Hence A-domains possess some plasticity towards chemically similar (Phe/Tyr) or distinct substrates (Arg/Tyr). Further substrate promiscuity can arise with the C-domain affecting the substrate-specificity profile of the adjacent A-domain, by altering the stability of specific conformational states within NRPS [62].

To translocate the aminoacyl or peptidyl substrates from the A-domain to the PCP-domain a priming reaction occurs which covalently attaches the Ppant group to the peptidyl substrate via a thioester linkage, carried out by ppant-transferases. PCP-domains are also known as the “thiolation domains” due to the catalyzation of this thioester bond during elongation [61]. The now tethered amino acid is moved towards the C-domain where it undergoes a condensation reaction with the nascent peptide. While the peptide substrate is tethered to the PCP, it can also be passed to an M-domains before reaching the C-domain. The enzymes of M-domains can further modify the peptides, increasing the structural diversity of CNPs produced by NRPS [61,63]. The finished natural product is released from the distal end of the assembly line by the Te-domain (termination). The Te-domain features a highly conserved catalytic Ser residue, or sometimes a Cys. which interacts with the PCP-domain thioester, temporarily anchoring the peptide with an intermediate ester bond to the Te-domain (i.e., R-O-Te) [60,61]. By a process of hydrolysis, aminolysis, or macrocyclization the linear or cyclic product is cleaved from the Te domain and released from the NRPS framework [57,61]. Te-domains frequently act as cyclase aiding in folding the linear peptide and catalyzing an intramolecular nucleophilic attack from one end of the metabolite to another, forming a lactone or lactam linkage [61].

1.4.2. Polyketide (PK) Biosynthesis

Polyketide synthesis (PKS) pathways produce a staggering range of functional and structural diversity in cyanobacteria metabolites. There are three types of PKS: type I PKS possess a multidomain architecture; type II PKS carries each catalytic site on a separate protein, typical of aromatic polyketides; type III PKS is unique, lacking multiple catalytic domains and using an acyl carrier protein (ACP) independent mechanism to catalyze elongation [64–66]. Type I PKSs can be further subdivided into two groups, modular and iterative. In Type I PKS there are modules of repeated functional synthesizing enzyme domains. These domains catalyze all discrete reaction steps for synthesis and elongation of the PK chain. A group of domains together represents a module and controls the selection and incorporation of each PK precursor. Iterative type I PKS possesses the same core catalytic domains as the modular type however, these domains occur on a single polypeptide that generates the complete PK chain, a non-modular architecture [65]. Modular type I PKS is present in cyanobacteria, resembling NRPS in respect to its modular nature and will be discussed further. The characteristic cyanotoxin microcystin possesses a diagnostic Adda (3-amino-9-methoxy- 2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid) moiety resulting from enzymes encoded by Type I modular PKS operons mcyD–J [67]. The domains of an individual type I PKS modules generally resemble those of fatty acid synthase (FAS) in animals and fungi [64]. Type I PKS enzymes incorporate different types of carboxylic acids, frequently acetyl-CoA starter units and extending malonyl-CoA derivatives which are subsequently activated, assembled, and optionally modified [64–66]. PKS modules are formed by the enzymatic subunits of ketosynthase (KS), acyltransferase (AT), ketoreductase (KR), dehydratase, (DH), enoyl reductase (ER) and ACP [57]. For complete synthesis PKS requires a loading module, elongation modules and a terminal module. The loading module is a di-domain formed by AT and ACP, responsible for selection of the starting extending carboxylic acid unit, typically acetyl-CoA [57,65]. The AT domain

is the “gate keeper” recognizing the specific malonyl derivative incorporated into the growing polyketide chain, covalently tethering it to a sulfhydryl group of the ACP subunit [57,65]. Each elongation module contains the three catalytic domains of KS, AT and ACP, required for one extension cycle of polyketide chain. There are also a variety of auxiliary domains (similar to NRPS M-domains) which can be optionally present between the AT and AP domains, responsible for modifications to the polyketide chain [57]. The primary C-C bond forming condensation reaction, is catalyzed by a KS domain, transferring the polyketide chain to the active site by formation of a β-keto ester intermediate [57,66,68]. Here the acyl-polyketide chain can be reductively modified by any auxiliary domains positioned between the AT- and ACP-domains [65,66,68]. PKS is terminated with TE module in which the thioesterase activity is thought to catalyse the cyclization and release of the synthesized polyketide [66,68]. Modular type I PKS is predominant in cyanobacteria and can co-occur with NRPS as hybrid pathways [57,59,63]. For hybrid NRPS-PKS pathways, the product release is facilitated by the NRPS TE-domain [57,68]. Upon product release post-PKS processing by tailoring enzymes can functionalize the resultant polyketide inciting further structural diversity within these biosynthetic pathways [65,66]. A suite of tailoring enzymes provides various functionalities such as, hydroxy, carbonyl or deoxy- sugar groups, methylation and oxidation of nitrogen or carbon centres [66].

1.4.3. Ribosomally synthesized post-translationally modified peptide (RiPP) biosynthesis

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are generated by posttranslational modifications of short precursor peptides. To date over twenty different families of RiPPs have been reported [56]. RiPP biosynthesis is frequently accompanied by a suite of tailoring enzymes which provide various substitutions and functionalizations to the final peptide, conferring the exotic chemistry observed with this pathway. RiPPs are similar to NRPs due to their posttranslational modifications however they do not require the large modular enzyme

complexes to incorporate residues [56,69]. The majority RiPP pathways begin as precursor peptides with an N-terminal leader (typically ≈20 – 110 residues) which guides biosynthetic enzymes responsible for post translational modifications (PTMs) to the core peptide [56,70,71]. In special cases, a follower peptide is attached C-terminus of the core peptide in place of an N-terminal leader [71,72]. Resembling proteins, RiPPs are synthesized ribosomally, with the N-terminal leader and core peptide passed through various enzymatic domains encoding and catalyzing the stepwise elongation of the ploypeptide chain. The entire sequence of the C-terminal core peptide is mainly conserved in the final product and can harbour multiple different sites for PTMs [70,72–74]. PTMs account for the vast structural diversity of RiPP pathways, catalyzed by methyltransferases, halogenases, hydroxylases, epimerases and dehydrogenases [71,72,74]. RiPPs frequently associate Cys residues with PTMs, converting Cys thiols to, disulfides, thioethers, thiazol(in)es and sulfoxides[72,74]. Commonly, macrocyclization and modifications to the C-N terminals occur, increasing metabolic stability and susceptibility to degradation [69,72]. Upon addition of PTMs to the core peptide the N-terminal leader is removed by a proteolytic cleavage. Infrequently, C-terminal recognition sequences are required for peptide cyclization and release [72]. In large RiPP biosynthetic gene clusters commonly shared enzymes across biosynthetic pathways, hence the evolution of new structures is an apparent result of acquiring new post-translational modification enzymes [69,72,75]. RiPP post-translational enzymes show high plasticity and substrate tolerance contributing to the vast diversity of metabolites from these pathways [75]. Microviridins (MV) and cyanobactins (CB) are the two most studied groups of RiPPs produced by freshwater cyanobacteria genera of *Microcystis* and *Planktothrix* [75,76]. Cyanobactins are small cyclic CNPs produced through the posttranslational modification of short precursor peptides with a highly conserved leader sequence [69,74]. CB precursors typically possess the leader peptide, an N-terminal protease recognition sequence, a core peptide, and a C-terminal recognition sequence [56,69,72,73]. In CB biosynthesis, translation of the core peptide is followed by heterocyclization of residues, carried out by the D-protein (cyclodehydratase) [56,72]. The D-protein acts with

regioselective heterocyclization of Cys, Ser, and/or Thr residues to azol(in)es[56,69,72]. Subsequently, N-terminal proteolysis (by A protease) removes the leader peptide and recognition sequence. A separate protease (G protease) then then cleaves at the C-terminal recognition sequence and catalyzes C–N macrocyclization. Further tailoring occurs with PTMs to the mature cycles such as prenylation, oxidation, dehydrogenation [71,74,77]. The enzymes of CB biosynthesis are promiscuous and contribute to an enormous array of sequence and functional group diversity for these metabolites [72,78]. MVs are large cyclic CNPs, whose precursors typically contain a highly conserved leader peptide linked to an immature C-terminal peptide core. Post transcription, biosynthesis is initiated by two ATP-grasp ligases (dubbed MdnC and MdnB) which preform macrocyclaization. MdnC and MdnB catalyze the formation of two macrolactones and one macrolactam linkages between sidechains of C-terminal core residues [71,72]. The larger lactone linkage forms first and lactonization is typically followed by lactam formation [72]. The microviridin gene cluster does not encode for protease signatures responsible for cleavage of the leader peptide [71,72,76]. It is postulated, that the associated ABC transporter, MdnE may play a scaffolding role to verify correct processing of MV precursors [71,72]. PTMs occur with GNAT type acetyltransferase, that functionalize the N-terminal amino acid of the final MV product [71,72,76]. Interestingly the, MV modifying enzymes appear to be unique to this biosynthetic pathway, responsible for the group's unique tri-cyclic cage-like architecture [71,72].

A vast diversity of bioactive secondary metabolites are produced by cyanobacteria. The majority of these metabolites are CNPs produced via NRPS, NRPS-PKS and RiPP pathways. The recognition plasticity substrate tolerance and tailoring modifications of biosynthetic pathways contributes to a multitude of structural diversity throughout CNP groups. For example, it is possible for NRPS, PKS and RiPP enzymes to recognize more than one amino acid as its substrate [61,66,72]. If the specificity is high, consistencies between metabolites occurs and results in the formation of a chemical groupings based on shared partial amino acid sequences or non-

proteinaceous amino acids. If the amino acid specificity of the enzyme is low, this generates a highly variable position within the metabolite and thus increases structural diversity. cHABs which occur naturally can contain multiple species and/or genera each with distinct metabolomes, adding further complexity to the presence and concentrations of cyano-metabolites in the environment.

1.5. CYANOTOXINS

Toxic cyanobacteria secondary metabolites are often termed cyanotoxins. The following section is an overview of specific types of cyanotoxins produced by common bloom forming cyanobacteria. In large, cHABs attract a great deal of scientific and public attention due to the various cyanotoxins they produce. Cyanotoxins pose adverse risks to the health of humans, animals, plants and Eukaryotic microbiota, due to their toxic effects. Some cyanotoxins are exclusive products of cyanobacteria (i.e., microcystins and anatoxin), whereas others, can be naturally produced elsewhere (e.g., saxitoxins and anatoxins) [79,80]. Most cyanotoxins are low molecular weight (100-1000 Da) alkaloids and peptides. The main cyanotoxins implicated with adverse human health outcomes are microcystins, nodularin, cylindrospermopsins, β-methylamino-alanine (BMAA), anatoxins, guanitoxin (anatoxin-A(S)) and saxitoxins [81]. Human exposure to cyanotoxins mainly occurs via consumption of contaminated food or drinking water and to a lesser extent, accidental ingestion of contaminated recreational waters [81].

1.5.1. BMAA (β-methylamino-alanine)

BMAA is a non-proteinogenic amino acid thought to be neurotoxic to humans and is currently a controversial molecule (Fig. 1). To date, no toxic mode of action has been irrefutably confirmed for BMAA and its toxicology in not well understood. Although there is scepticism with BMAAs

toxicity to humans, the situation is best approached with the precautionary principle [82]. Attention was first drawn to BMAA in the late 1960s due to the occurrence of a neurological disease, amyotrophic lateral sclerosis/Parkinsonism dementia complex (ALS/PDC) appearing within a local native population of Guam [83]. ALS/PDC is similar to Parkinson's disease and dementia with varying symptoms of neurological dysfunction. ALS/PDC effected individuals exhibited a lack of mobility, memory decline, cognitive deficits and often resulted in premature death [84]. It was found that the induced toxicity was linked to the consumption of native cycad seeds. Analysis of the suspect cycad seeds resulted in the first isolation of BMAA [85]. Subsequently, Vega et al., (1968) found that BMAA induced neurotoxicity at high doses (3-14 μmoles/g b.w.) when administered by intraperitoneal injection (i.p.) into chickens and rats. A study exposed macaque monkeys *(Macaca fascicularis)* to high orally administered BMAA dosages and observed neurotoxic symptoms after a month for doses exceeding 200 mg/kg [86]. BMAA doses which induced toxicity in these model organisms were magnitudes greater than those consumed by Guam natives therefore it was suspected that other chemicals associated with cycads or mixture effects are suggested as a more possible cause of ALS/PDC [87]. For example, cycasin toxin, a known neurotoxic alkaloid was too detected in Guam cycad flour [87]. A similar case was resurrected by Spencer et al,. (1987) inciting that, the consumption of flying foxes by Guam natives would reach exposure levels similar to those of tested monkeys (>200 mg/kg orally). This assertation was subsequently discounted due to the lack of evidence [88,89]. The cause(s) of the ALS/PDC in Guam has not yet been satisfactorily explained [81]. Over the past several decades the issue has grown to be known as the BMAA-neurodegenerative disease hypothesis. Although robust exposure data is required for risk

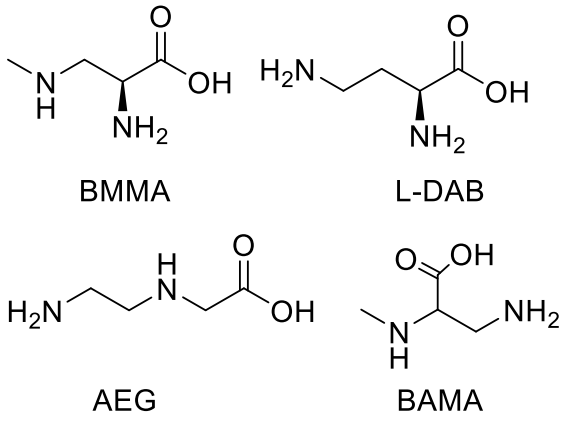

Figure 1. Chemical structures of BMAA and its isomers 2,4-diaminobutanoic acid (DAB), N-2-aminoethyglycine (AEG), β-amino-N-methylalanine (BAMA)

assessment, one key question remains unanswered – can toxic effects of BMAA be confirmed in health-relevant dose ranges?

Another major issue that has yet to be over come is the accurate detection and estimation of BMAA levels in environmental and food samples, for evaluation of human health and exposure risks. Common liquid chromatography fluorescence detector (LC-FLD) analysis methods risk misidentification of BMAA. LC-FLD analyzed samples can result in false positives and over estimated concentrations – lacking replication when analyzed with more reliable analytical methods (i.e., LC-MS/MS) [81]. Validated methods for the analysis of synthetic BMAA standards or naturally occurring BMAA in cyanobacterial samples does not yield adequate results for BMAA detection in tissue samples. A case study in this respect to this issue, is the quantification BMAA concentrations in stranded dolphins. In this study, BMAA was reported from the brains of 13 of the 14 dolphins, at concentrations of 20 to 748 μg/g, quantified by LC-FLD [90]. However, subsequent LC-MS/MS analyses we conducted with the highest concentration being 0.6 μg/g [90]. Only 4 of the 14 samples were analyzed by LC-MS/MS however these were several orders of magnitude lower than the LC-FLD results reported. Interestingly, one sample that tested negative by LC-FLD tested positive by LC-MS/MS [90]. LC-MS/MS methods albeit more reliable, are still in development for BMAA accurate detection and quantification. Underivatized BMAA cannot be adequately detected in reverse-phase liquid chromatography with a C18 column but has been studied by several investigators using HILIC columns [82]. However, HILIC columns cannot separate BMAA from its isomer BAMA (β-amino-N-methyl-alanine; Fig. 1) [82]. Therefore, it has been demonstrated that HILIC columns frequently do not accurately detect BMAA in environmental samples, contrasting with similar studies utilizing C18 columns [81,82,91]. A method adhering the guidelines of the Association of Official Analytical Chemists (AOAC) was developed by Glover et al. 2015, requiring derivatization of BMAA [91]. This method can effectively separate BMAA isomers 2,4-DAB, AEG, and BAMA [91]. Recently is has been demonstrated that BMAA

isomers have a more potent toxicity, in the following order, AEG > DAB > D-BMAA > L-BMAA (Fig. 1) [92]. Currently, few papers on BMAA report method validation, and fewer still provide sufficient details to assess validation [81,82]. Moreover, difficulties still arise without efficient separation of BMAA isomers within complex sample matrices. There is still no method available to accurately quantify underivatized BMAA and isomers.

1.5.2. Cylindrospermopsins (CYNs)

Cylindrospermopsin (CYN) is a zwitterionic tricyclic alkaloid cytotoxin first isolated and identified in 1992 from the cyanobacterium *Raphidiopsis raciborskii* (formerly *Cylindrospermopsis raciborskii*; Fig. 2*)* [93]. Attention was first drawn as a "mystery disease" occurring near Palm Island, off the coast of Queensland Australia, when 138 children and ten adults were poisoned, resulting in hepatitis-like symptoms [94]. Cultures of *R. raciborskii* produced similar effects in mice to that observed with human [81]. Later studies confirmed CYN toxin as culprit. CYN is highly water soluble, containing a, conserved polar tricyclic guanidino structural moiety (Fig. 2) that is commonly sulfated. Four additional CYN structural variants have been identified 7-epi-cylindrospermopsin (7-epi-CYN), 7-deoxy-cylindrospermopsin (7-deoxy-CYN) 7-deoxy-desulpho-cylindrospermopsin and 7-deoxy- desulpho-12-acetylcylindrospermopsin (Fig. 2). CYNs are found globally, a result of the worldwide distribution of producing cyanobacteria genera/species, these include *Oscillatoria, Aphanizomenon*, *Raphidiopsis raciborskii* and *Chrysosporum ovalisporum* [81]. The increased global presence of the toxin and analogs is likely due to improvements in water quality monitoring and analytical detection methods [95]. However, it is also likely that the organism is too, expanding into more suitable habitats made available by climate change and eutrophication [95].

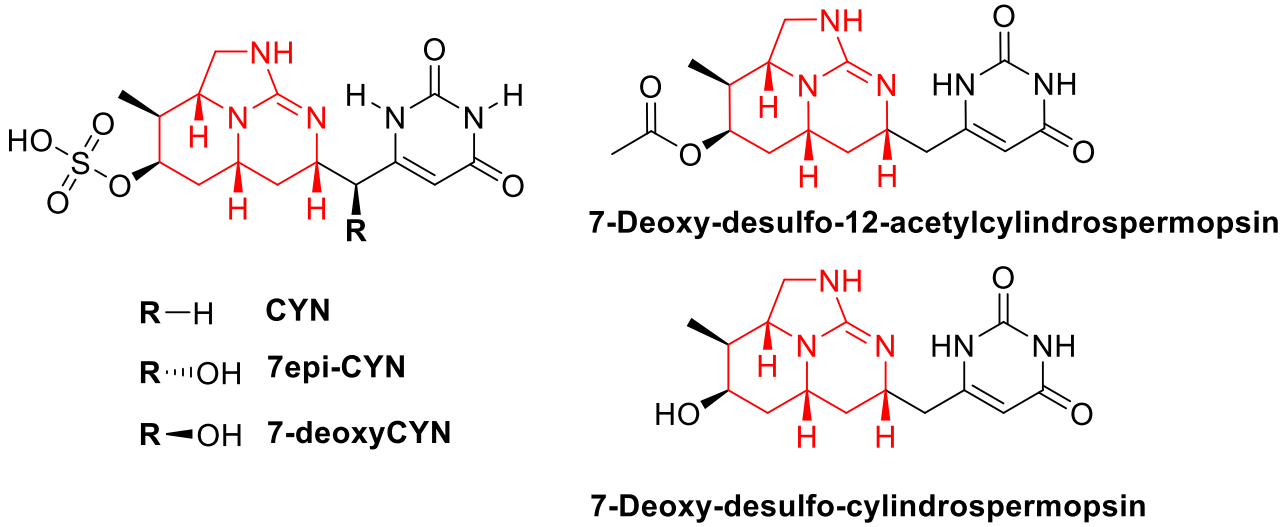

Figure 2. Structures of the 5 cylindrospermopsin congeners. The characteristic tricyclic guanidino structural moiety is indicated in red.

The toxicology of CYNs is well-understood although its several modes of action have not been completely elucidated. CYN primarily targets the liver, kidneys, and erythrocytes with multiple modes of action. Studies of radiolabelled CYN suggest that it is distributed to all major organs. Norris et al., (2001) found that ^{14}C-labeled CYN primarily concentrates in liver (20.6%) and kidney (4.3%) tissues of (Quackenbush) mice 6 h after intraperitoneal (i.p.) administration [96]. CYN mode of hepatotoxicity consists of four phases: inhibition of protein synthases, cell membrane proliferation, fat droplet accumulation, and cell death [97]. CYN symptoms of liver damage include, inflamed and swollen fatty liver, increased hepatocellular vacuolation, and liver necrosis [81]. Symptoms of kidney damage includes, abnormalities include interstitial nephritis, decreased numbers of glomerular erythrocytes, epithelial cell necrosis, and glycogen and protein deposition in renal tubules [81]. However, the direct mechanisms toxicity related to CYN modes of action still have yet to be elucidated. Cyto-, geno-, immuno-, neuro-, and endocrine and developmental toxicity modes have been reported in CYN toxicological studies [81,97]. Generally, as a result of inhibition of protein synthases, possible interactions with cytochrome P_{450} (CYP_{450}), induction of oxidative stress and DNA strand breaks, binding to estrogen receptors and affecting acetylcholinesterase (AChE) activity [97]. Overall, at low dosages, inhibition of protein synthetases

is the primary toxic effect. In contrast, at higher doses CYN toxicity appears to involve cytochrome P_{450}-dependent mechanisms, with ROS and induction of cellular stress responses possibly being involved too. Genotoxic effects have been demonstrated *in vivo* and *in vitro* studies [95,97]. There is limited viable toxicological information on CYN as many studies utilize cultured cyanobacterial cell extracts rather than the pure compound. So far, a few countries have set public guideline values for CYN. A guideline safety concentration of 1 µg/L in drinking water was originally proposed by Humpage and Falconer (2003) [98]. More recently, the WHO (World Health Organization; 2021) [81] has derived adult human provisional guideline values (PGV) for lifetime drinking-water, short-term drinking-water and recreational water, 0.7, 3 and 6 µg/L respectively. The WHO PGVs consider NOAEL (no-observed-adverse-effect level; 30 µg/kg b.w./day) based on Humpage and Falconer (2003) [98], bw (body weight; 60 kg for an adult), P (fraction of exposure allocated to drinking-water), UF (uncertainty factor; 300), C (daily drinking-water consumption; 2 L for an adult). The WHO (2021) [81] provides further recommendations for application of these guidelines. Specifically recommending, the sum of CYNs be evaluated against the PGV as there is limited evidence of the relative potency of analogs; reduction of the drinking water fraction of exposure for the lifetime and short-term drinking-water PGVs based on relative exposure data for a specific population, as food exposures will vary based on location; a standard drinking-water PGV of 0.7 µg/L value is necessary for bottle-fed infants and small children, given they can ingest a significantly larger water volumes relative to body weight [81].

1.5.3. Anatoxins (ATXs)

Anatoxins (ATXs) are secondary amine neurotoxic alkaloids with low molecular weight (~165 Da; Fig. 3) [99]. They are homotropane derivatives with an enlarged bicyclic ring (Fig. 3). Anatoxin-a (ATXa) is the most studied congener of the group with a pKa value of 9.6, existing mainly in cationic form [99]. Several congeners of anatoxins exist, the most commonly being, homoantoxin (hATX) and dihydroantoxin-a (H_2-ATX; Fig. 3) [100,101]. hATX contains an additional methyl group

on the methoxy side chain and H_2-ATX has no double-bond on the adjacent carbon to the side chain, in comparison to ATX (Fig. 3). The first report of occurrence of anatoxin-a dates back to 1951 in the USA [99]. It was then identified in Canada by P. R. Gorham in the 1960s referred to as the fast-death factor (FDF) due to its high lethality to exposed mice [102]. Subsequently, the structure of ATXa, 2-acetyl-9-azabycyclo[4.2.1.]non-2-ene was determined by C. Huber (1972) [103] using X-ray crystallographic analysis. Experimental synthesis of ATXa and hATX was demonstrated by H. Campbell (1977) [100]. An ATXa producing *Dolichoispermum* (*Anabaena) flos-aquae* strain (NRC-44) was isolated from Burton Lake, Canada and its toxicology characterized by Carmichael et al. (1975) [104].

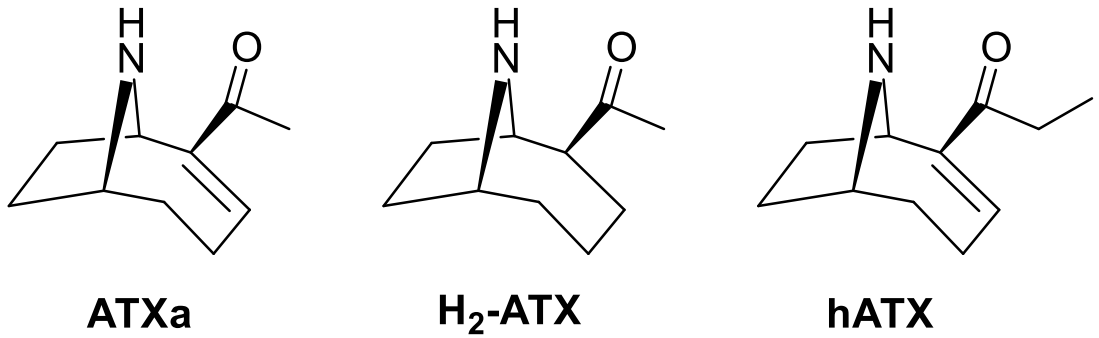

Figure 3. Chemical structures of the most common ATX analogs: ATXa (anatoxin-a), H_2-ATX (dihydroantoxin-a) and hATX (homoantoxin)-a.

Chemically, ATXs are found in the environment as mixtures of structural analogues mainly, ATX, H_2-ATX, hATX or epoxy analogues [80,81,99,100]. Each with varying degrees of toxicity based on exposure routes [99,105]. For example, hATX is as potent as ATX whereas H_2-ATX is 10-fold less toxic than ATXa, for i.p. administration [99]. A recent study by Puddick et al. (2021) indicated that H_2-ATX is approximately fourfold more toxic than ATXa with gavage administration [105].

A significant amount of toxicological information is available for ATXs although, knowledge gaps still remain for the derivation of guideline values. ATXs are a potent neurotoxin that interfere with transmission of nervous impulses in the chemical synapses within nervous systems of vertebrates. In animal cells ATXs act as an effective synaptic depolarising agent, antagonistically

competing with acetylcholine, binding to nicotinic and muscarinic receptors in both the CNS (central nervous system) and PNS (peripheral nervous system [45,81,99]. Therefore, the cardiovascular system has also been indicated as a target organ due to muscular paralysis [81]. However, ATXa does have a 100-fold selectivity for nicotinic compared to muscarinic receptors [99]. At the nicotinic receptors, ATXa binds which opens the cationic channels leading to an influx of Na^+ into the cell and an efflux of K^+ (potassium cations), resulting in overstimulation of postsynaptic receptors. The most evident and lethal effect of ATXa is death from respiratory arrest due to over-stimulation of muscles. Acute exposure studies in animals led to deaths within minutes of gavage administration [45,104]. The minimum lethal dose of ATXa being 0.25 mg/kg b.w. for i.p. administration in mice. After the administration of a sublethal single doses, mice are observed to readily recover. Additional effects attributed to ATX in cell cultures include cytotoxicity, caspase activation, apoptosis, oxidative stress induction and formation of ROS [45,99]. Many ATX producing cyanobacteria species have been reported across the globe, specifically cyanobacteria taxa belonging to the orders Nostocales and Oscillatoriaes [80]. No Canadian guideline for MAC (maximum allowable concentration) has been established for ATXs. Largely due to a lack of available oral toxicity data, specifically either a NOAEL or LOAEL [81]. Likewise, the WHO has not established PGVs for ATXs due to this issue. The WHO has derived adult human health-based reference values (HBRVs) for short-term drinking-water (≈30µg/L) and recreational water exposures (≈60µg/L) using highly conservative assumptions to define the NOAEL [81]. HBRVs are applied as a conservative approach as a result of insufficient data and, are not representative of an HBGV. However, there is insufficient data for derivation of a long-term drinking-water HBRV [81]. Analytical methods with LOQs well below 30 µg/L for ATX detection can be achieved with several established methods [45]. These include HPLC-FLD (with derivatization), LC-MS/MS and the commercially available ELISA and RBA kits [45]. Certified reference material for ATX and H_2-ATX are commercially available. More recently, Beach et al. (2021) developed a quantitative rapid screening method for ATXs in using direct analysis in real time combined with

high-resolution mass spectrometry (DART-HRMS) [106]. Although further optimization is required for application with environmental samples, the method is a promising prospect for rapid analysis.

1.5.4. Saxitoxins (STXs)

The first documented report of STX poisoning was in 1798, where crew members of an exploratory vessel near the coast of British Columbia, Canada consumed mussels resulting in unexplained illness and mortalities [107]. The toxicity of STXs was first recognized by Sommer (1932) [108], who reported lethality experiments with extracts from shellfish collected along the U.S. Pacific coast. STXs are a broad group of potent neurotoxic alkaloids with as much as 0.5 – 1 mg of pure STX mixtures constituting a lethal dose in adult humans. The STX group consists of > 50 congeners composed of a structurally characteristic 3,4-propinoperhydropurine tricyclic system containing two guanidine groups (Fig. 4) [109,110].

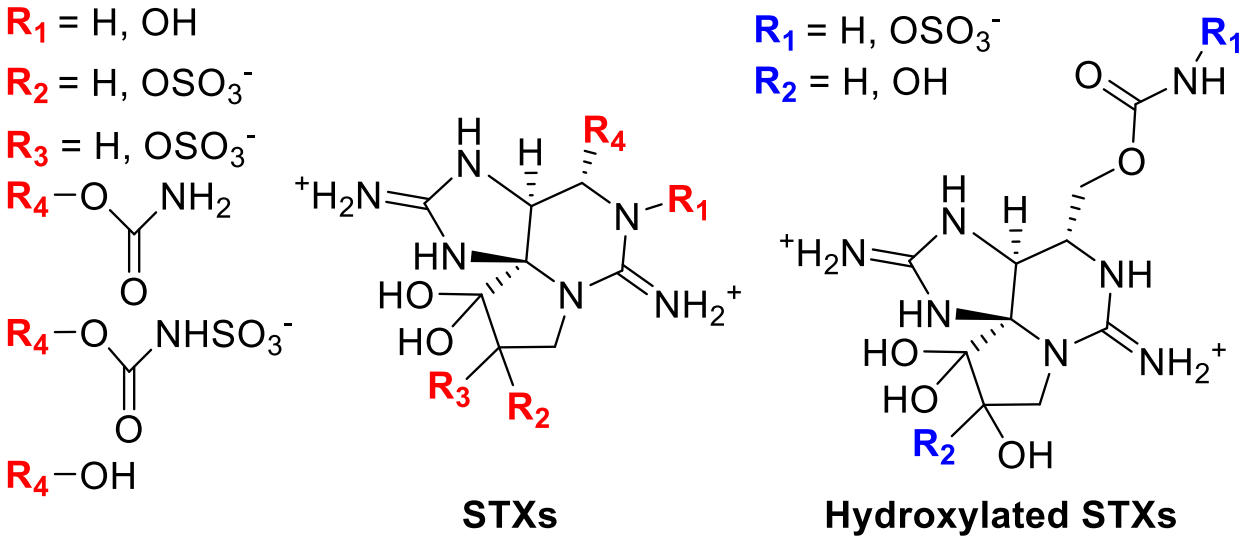

Figure 4. Representative chemical structures of STX and hydroxylated STX. STX R_1, R_2, R_3 and R_4 substitutions indicated in red. Hydroxylated STX R_1 and R_2 substitutions indicated in blue.

This characteristic moiety ties STX to a large family of guanidinium-containing marine natural products and are too responsible for the molecule's high polarity 67,110. STXs can be structurally subclassified as non-sulfated, mono-sulfated gonyautoxins, di-sulfated, decarbamoylated and some hydrophobic variants [110]. However, there are four general groups STXs are lumped into,

carbamate, N-sulfo-carbamoyl, dicarbamoyl (dc) and hydroxylated toxins – in reference to substitutions at the R_4 position (Fig. 4) [81,110]. Most known STXs are hydrophilic, with the exception of variants produced exclusively by freshwater *Microseira* (formerly *Lyngbya*) *wollei*, known as *L. wollei* toxins (LWTXs) [110]. STXs are also generally known as paralytic shellfish poisons (PSP) due to their original discovery in molluscs, with subsequent bioaccumulation, consumption, and poisonings of humans [67]. Both marine dinoflagellates and freshwater cyanobacteria are known to produce these toxins. Specifically, the cyanobacteria genera of *Dolichospermum (Anabaena), Aphanizomenon, Raphidiopsis (Cylindrospermopsis) raciborskii, Scytonema, Planktothrix and Microseira (Lyngbya) wollei* [67,81].

The toxicology and mode of action of STXs is well documented. In animal cells STX mechanism of action is based on the blocking of voltage gated ion channels (i.e., Na^+, Ca^{2+} and K^+ channels) [80,111]. STXs prevents the flow of electrical signals to peripheral nerves and skeletal/cardiac muscles. Typical neurologic symptoms include nervousness, twitching, ataxia, convulsions and muscle and respiratory paralysis. For lethal doses, animal mortalities have been observed within a few minutes of administration [81]. For humans, respiratory paralysis mortalities have been reported after 2–24 h after exposure [81]. The toxicity of STXs range significantly depending on the specific chemical congener. For example, carbamate STXs are by far the more (~60%) toxic then congeners lacking the carbamoyl side group [110]. Moreover, Di-sulfated STXs and LWTXs possess an inherently lower toxicity [110]. Toxicological studies generally report results as STX equivalents (STXeq) although, STXeq represents the total summed concentration of STX variants, thus there is a wide range in both lethal and nonlethal reported doses for STXeq. In some instances, the STXeq represents concentrations adjusted for toxicity of specific congeners. The minimal lethal dose of STXeq range from 86–788 µg/kg b.w. in adults and 25 µg/kg for children [81]. Doses in the range 140–300 µg STXeq/person are reported no or mild symptoms [81]. However, mild to moderate symptoms of paralysis and even death, followed ingestions of 460–12,400 µg

STXeq/person [81]. The EFSA (2009) identified a LOAEL for STXeq of 1.5 µg/kg b.w. assuming an adult body weight of 60 kg [112]. From the established LOAEL the WHO has designated guideline values for acute drinking-water and recreational water exposures 3 µg/L and 30 µg/L, respectively. To ensure accurate risk assessment of guideline values, the WHO (2021) recommends STXeq be calculated to indicate concentration equivalents of all STXs present, each being quantified against an analytical standard [81]. Multiplying the concentration of each STX analogue by the respective toxicity equivalence factor (TEF) before addition. In Canada federal health guideline values have not been established for STXs.

STXs have been detected in all types of freshwater worldwide generally, they are detected in only a small share of the samples at concentrations > 10 µg/L [67,80,110]. However, large concentrations of STXs (~200 µg/L) can be expected in prolific blooms and scums of producing cyanobacteria taxa [67,110,111]. As a result of their common hydrophilicity and chemical variability, STXs are one of the most challenging cyanotoxin groups to analyse. HPLC-FLD methods (requiring prior analyte oxidation) previously developed for seafood analysis of STXs in the marine environment are transferable to cyanobacterial samples [112]. Although higher specificity is achieved with reliable LC-MS/MS approaches currently used for routine analyses of STXs from marine and freshwaters. Certified reference material is commercially available for some STX congeners, but the lack of reference material for several STX congeners is still a limitation for LC-MS/MS methods.

1.5.5. Microcystins (MCs) and Nodularin (NOD)

MCs were first structurally characterised in the early 1980s and named after the cyanobacterium *M. aeruginosa* from the initial culture isolated [113]. MCs are the most extensively studied cyanotoxin group due to their global ubiquity and potent hepatotoxicity. Over 300 MC congeners have been reported to date where MC LR (Micorcystin-LR) is the most commonly studied congener, although MC LA, MC LY, MC RR, MC YR have gained more interest due to their toxicity [113–115]. Structurally, MCs are cyclic heptapeptide cyanotoxins that contain a diagnostic Adda ((2S,3S,4E,6E,8S,9S)-

3-amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoicacid) moiety at position 5 (Fig. 5). The Adda moiety is also characteristic of nodularins (NODs) which are cyclic pentapeptides produced by cyanobacteria from the genus *Nodularia* and have a similar toxicity and structure to MCs [116]. Another key difference is that NODs incorporate an Mdhb residue whereas, MCs incorporate an Mdha residue into their peptide backbone [116]. MCs mainly possess the common structure scaffold of cyclo-[D-Ala^1-X^2-D-MeAsp^3-Z^4-Adda^5-D-Glu^6-Mdha^7] [117].

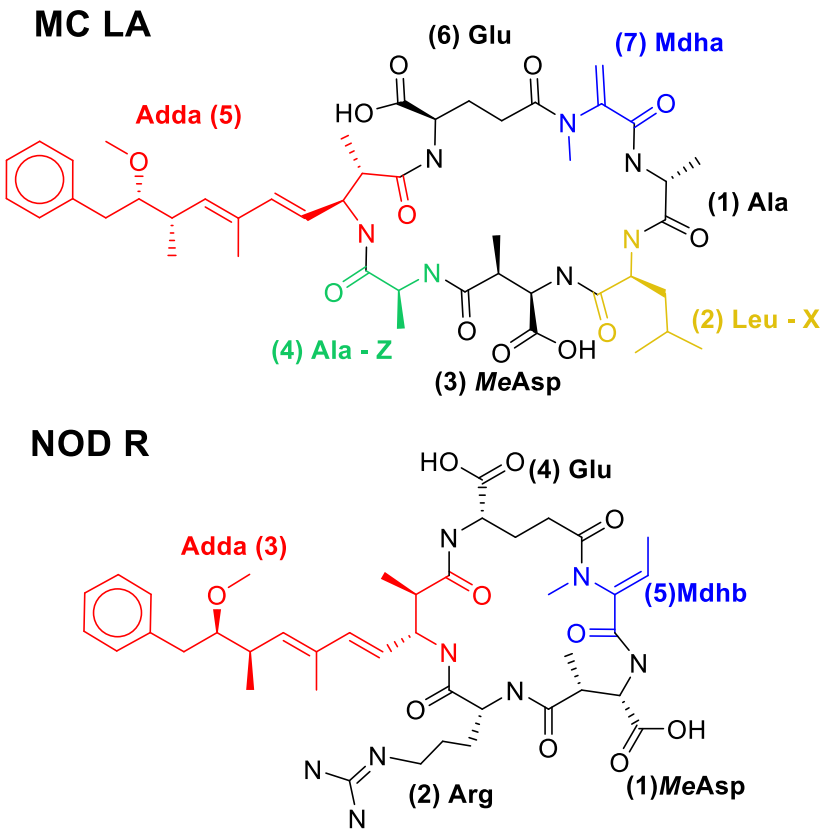

Figure 5. Chemical structure of MC LA (left) and NOD R (right) with the characteristic Adda moiety indicated in red. The Mdhb of NOD and Mdha of MCs is coloured blue. The two most variable amino acid positions in MCs, X and Z are coloured green and yellow. The two most variable amino acid positions 4(Z) and 2(X) are indicated blue and green

The X and Z residues represent positions two and four which are highly variable and determine the suffix in MC nomenclature. For example, MC LA contains Ala at position 4 (Z) and a Leu at position 2 (X; Fig. 5). Other frequent structural variations include substitution of Mdha for Dhb (dehydrobutyrine) or serine at position 7, and demethylation of amino acids at positions 3 and/or

7 [113,118]. cHABs can contain extremely high concentrations of MCs, composing up to 1% of the dry weight [119]. MC-producing cyanobacteria strains can be found in all orders of taxa such as, Chroococcales, Oscillatoriales, Nostocales, and Stigonematales [67,114]. Within orders, the distribution of MC-producers at the level of genera or species level is inconsistent. Many genera or species contain both producing (toxigenic) and nonproducing (non-toxigenic) strains [114,120]. Toxigenic strains are global distribution residing in tropical, temperate, and polar habitats [114]. Most toxigenic strains produce several MC congeners and, commonly only a few are dominant in any particular strain or cHABs [67,115]. Arguably, MCs are most common freshwater cyanotoxin on a global scale, resulting from the extreme diversity and distribution of producing cyanobacteria genre. Comprehensive toxicological information is available for this cyanotoxin group with MC LR receiving the greatest attention due to its ubiquity, abundance, and toxicity. MCs are known to be acutely and chronically toxic to animals and humans with acute LD_{50} values ranging between 50 (MC LR) and 1000 (MC RR) mg/kg b.w. for i.p. injection in mice [120]. The oral LD_{50} in mice is 5000 µg/kg b.w. is approximately 100-fold lower than the i.p. LD_{50} [120]. The toxicity by oral uptake is generally at least an order of magnitude lower than toxicity by injection as MCs cannot readily diffuse through plasma membranes, instead they require active transport into cells [78,120]. Hepatotoxic MCs are large 800-1100 Da cyanotoxins, therefore require membrane transport proteins such as organic acid transporter polypeptides (OATPs) to enter cells. OAPTs are expressed particularly in the liver and other tissues, facilitating the absorption of MCs [113,118]. Experiments have demonstrated reduced liver damage with inhibition of OATPs indicating their vital role in MC transport[119]. The site of MC metabolism is largely dictated by route of exposure, distribution, and dose[115,118]. Likewise, the toxic effects produced by different MCs congeners can vary both qualitatively and quantitatively, depending on the exposure conditions [115]. The toxicity of oral MC exposures is less than others [120].

The high expression transporters in the liver accounts for MCs selective hepatotoxicity[119]. The toxic mode of action for MC LR has been extensively studied. Protein phosphatase (PP; i.e., PP_1, PP_{2A}, PP_4 and PP_5) inhibition, is the initiating event for the MC mode of action[118,119,121]. The greatest IC_{50} of MC inhibition is expressed with PP_1 and PP2A in the 0.1 – 1nM range[121]. The non-proteinaceous Adda and Mdha structural features are key contributors to MC PP inhibition. The long hydrophobic Adda moiety facilitates reversible interactions that move the molecule into position inside the PP active site while the Mdha moiety covalently bonds within the active site pocket creating an irreversible interaction [77]. Upon PP inhibition, phosphatase dephosphorylation of cytokeratins and kinase phosphorylation occurs resulting in cytoskeletal destabilisation and subsequent apoptosis and necrosis [119]. In the liver, acute high doses lead cell-to-cell adhesion causing damage of sinusoidal capillaries and intrahepatic-haemorrhaging [78,119]. At low chronic doses phosphatase inhibition induces cellular proliferation, hepatic hypertrophy and tumour promoting activity [78,119]. The primary mode of MC detoxification is conjugation with cysteine and glutathione [118,119]. Most detoxification occurs in the liver and clearance of MCs can take a long time, up to weeks [118,120]. Clearance of MCs is congener- and conjugate-specific, having major differences in affinity and capacity between MCs congeners and their conjugates [115,119]. Therefore, the relatively similar toxicological effects of the most common MCs cannot be extended to all congeners, better expressed as toxicity equivalents.

Hepatotoxic MCs are also regarded as possible human carcinogens. MC containing cyanobacterial extracts show genotoxicity, while pure MCs yield negative results [119]. Evidence suggests MCs are not mutagens and experimental discrepancies are a likely result of the complex composition of extract mixtures and cyanobacteria cultures. Currently, cellular DNA damage from cyanobacterial extracts is thought to be an indirect effect from induction of apoptosis and cytotoxicity rather than direct mechanistic effects on the DNA [119]. MCs are classified as possibly

carcinogenic to humans (IARC Group 2B) due to their tumour promoting activity linked to PP inhibition [122].

MCs are the most regulated freshwater cyanotoxins with many countries have established guideline values for drinking and recreational waters. The WHO has designated PGVs for lifetime and acute drinking-water and recreational water exposures, ≈1 µg/L and 12 µg/L, and a recreational water PGV, 24 µg/L [81]. The PGVs are based on toxicological data for MC LR. Since MCs occur as mixtures, values should be applied with total MCs as equivalents based on the assumption that all MCs have relatively similar toxicity. In Canada the federal government has derived a tolerable daily intake (TDI) and health-based value (HBV) from toxicological evaluations of Heinze (1999) [123]. A seasonal MAC of 0.0015 mg/L (1.5 µg/L) total MCs has been established based on the Canadian HBV [124]. The current Canadian MAC and WHO PGVs are measurable by current analytical methods and achievable by current treatment technologies.

For extraction of intra- and extracellular total MCs, freeze–thaw cycling in aqueous methanol is commonly performed before analysis. Sample preparation can further concentrate MCs using evaporation or solid-phase extraction (SPE). The most reliable analytical method with the highest specificity and sensitivity for MCs is LC-MS/MS [119,125]. Accurate LC-MS/MS quantification requires reference standards for each MC congener and reference standards are commercially available for the majority of MC variants. Additionally, several non-specific total MCs analyses exist. One approach uses oxidative cleavage of the Adda moiety and subsequent analysis of the product, 2-methyl-3-methoxy-4-phenylbutanoic acid (MMPB) to quantify total MCs (and NOD) [119]. Commercially available tests include immunological (e.g., ELISA) and biochemical (e.g., PP inhibition assay) techniques [125]. These commercial tests cannot accurately identify specific MC congeners; however, they are useful for screening purposes.

1.6. CYANOPEPTIDES

Although, MCs and other cyanotoxins produce a serious health risk to humans there remains a suite of bioactive CNP metabolites with undefined toxicological implications and environmental concentrations. This section describes CNPs biosynthesized by RiPP, NRPS or combined PKS-NRPS pathways. Most CNPs are cyclic or linear non-ribosomal peptides that contain unique non-proteogenic amino acids and are grouped based on common structural features. The variable A domain sequences, selective plasticity, and modification enzymes in NRPS contribute to observed in gene clusters, especially within strains of the same species suggests metabolomic profiles are unique to the strain level [52]. The variation in biosynthetic machinery translates to a vast amount of structural diversity in CNPs. Likewise, cHABs produce complex profiles CNPs each unique, varying temporally and geographically. For these reasons, the ecotoxicology, toxicokinetics and environmental concentrations of many CNPs are not known.

1.6.1. Anabaenopeptins (AP)

Anabaenopeptins are a class of hexacyclicpeptides that share a diagnostic ureido linkage for cyclization with a D-Lys residue at position 2 (Fig. 6) [126]. The exo-amino acid residue at position 1 is joined to the cyclized ring by a carboxy group (Fig. 6). Both the exo- and remaining amino acids are highly variable and result in vast structural diversity of the congeners within this CNP group (> 100 reported; Fig. 6) [127]. AP A and B were the first isolated analogs, from the freshwater *Dolichospermum flos-aquae* (syn. *Anabaena flosaquae*) clone NRC 525-17 [128]. These metabolites are synthesized by NRPS pathways [129]. There are many trivial names for anabaenopeptins based on the producing genera. For example, oscillamides isolated from *Planktothrix* (formerly *Oscillatoria*), ferintoic acids from *Microcystis*, nodulapeptins from *Nodularia*, and lyngbyaureidamides isolated from *Lyngbya* spp.[127].

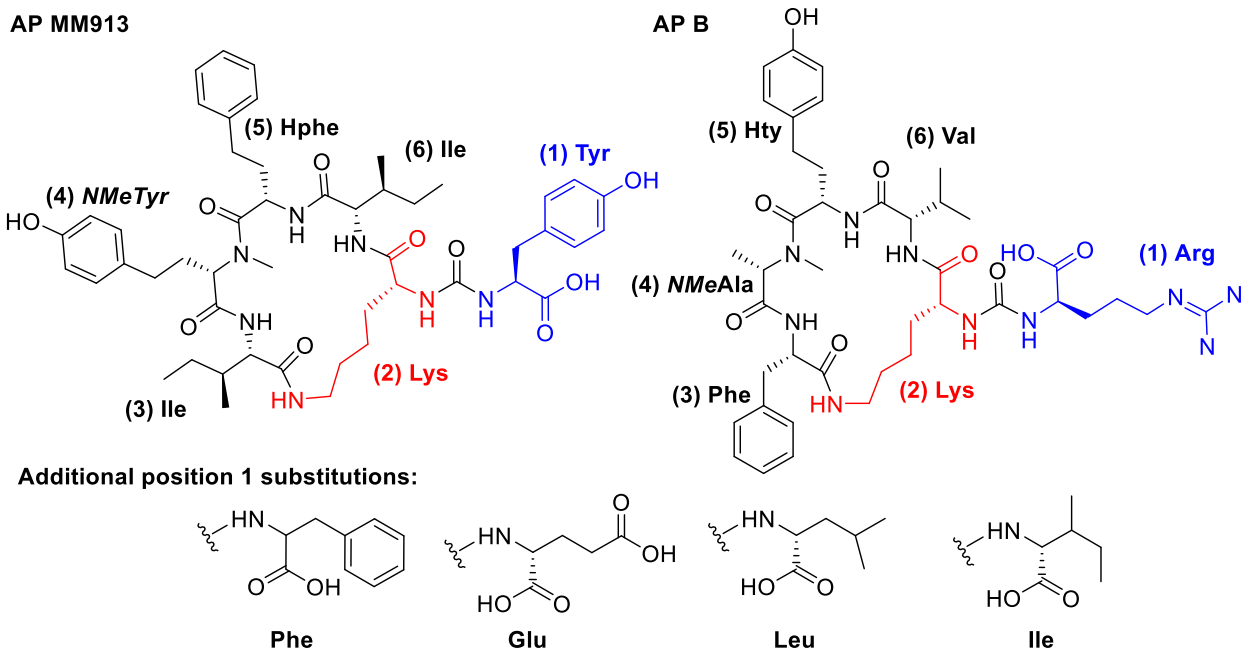

Figure 6. Chemical structures of AP MM913 (left) and AP B (right). The characteristic Lys structural moiety is indicated red and the exo residue indicated blue. Other possible substituents of the exo position (1) are shown below

These metabolites are produced by both marine and freshwater cyanobacteria, further complicating the vast chemical diversity of this class of metabolites. APs are regarded as toxic serine protease and phosphatase inhibitors with most reported to inhibition of chymotrypsin, carboxypeptidase A (CPA), elastase, trypsin, and PP_1 and PP_{2A} in the µM range [126,127,130–133]. The inhibitory activity of APs is largely controlled by the configuration of amino acids and is congener dependant. Inhibition assays have displayed APs with PP_1 IC_{50} values in the nM range. The very few studies that report environmental concentrations have detected AP in similar concentration ranges to MCs [129]. Toxic effects of APs have been demonstrated with model organisms. A 500 µg/L dose of Oscillamide Y induced a hyperactivate photomotor response to light in Zebrafish (*Danio rerio*) [134]. The crustacean, *Daphnia magna* displayed a reduction in oxygen consumption, thoracic limb activity and post-abdominal claw activity following 24 hr. exposures of AP B with an EC_{50} range of 1.7 - 27 µg/mL. Nematodes (*Caenorhabditis elegans*) displayed reduced reproduction, delayed hatching, decreased growth rate, shortened lifespan, following 10 µg/L exposures AP A, -B and -F [135]. Furthermore, it was found APs induced severe vulval integrity

defects in adult nematodes during the aging process and the three APs tested showed greater toxic effects than MC-RR, for the study [135].

Table 1. Enzyme inhibition associated with the AP group

Enzyme inhibition	IC_{50} Inhibition Range (nM)	Reference(s)
trypsin	45,000 – >45,500	[133,139]
chymotrypsin	16,000 – 50,000	[133,139]
PP_1	72 – 435,000	[127,131,133]
thrombin activatable fibrinolysis inhibitor	1.5 – 100	[132]
carboxypeptidase A	2 – 45,000	[127,130,131,133]

IC_{50} values reported in µg/mL and µM were converted to nM

The toxicity of APs is generally less than that of MCs however studies have shown that simultaneous exposers to these groups generates synergistic inhibitory effects (i.e., AP B + MC LR) greater than either individual CNP [136]. APs have been detected with MCs and other CNPs in naturally occurring blooms, with certain instances exceeding MC concentrations [137,138].

1.6.2. Cyanopeptolins (CP)

CPs were originally isolated from the freshwater *M. aeruginosa* PCC 7806 [140]. Cyanopeptolins are a group (>230 compounds) of cyclic oligopeptides containing a pentapeptide ring and a diagnostic 3-amino-6-hydroxy-2-piperidone (Ahp) moiety at position 3 (Fig. 7) [141]. Position 2 is commonly occupied by an *NMe*-aromatic residue such as Phe and Tyr, with possible halogenation of Tyr (Fig. 7) [142]. The remaining amino acid residues are highly variable, with the exo positions (S_1 and S_2) containing two or more substituents typically filled with two amino acids, or a fatty acid, such as hexanoic or octanoic acids (Fig. 7) [126]. The exo S1 position is commonly a threonine (or less common proline) linked to side chain of variable length at position S_2 (Fig. 7) [126].

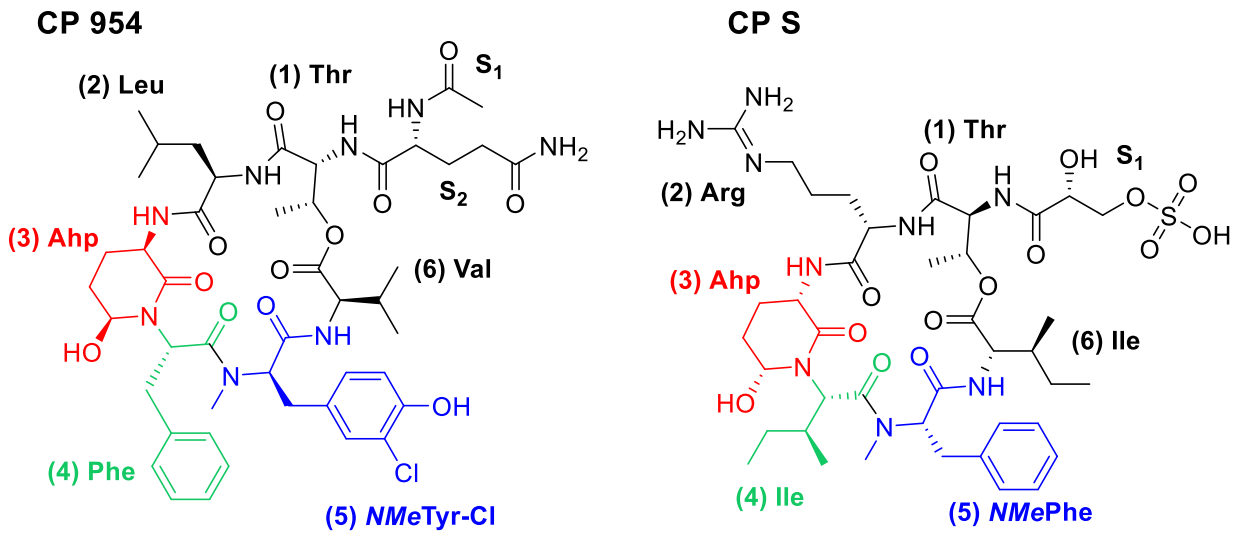

Figure 7. Chemical structures of CP 954 (left) and CP S (right). The characteristic Ahp structural moiety is indicated red. The Ahp-Phe, Ahp-Lxx, *NMe*Tyr and *NMe*Phe partial amino acids at position 4 (green) and 3 (blue) that subclassify CPs are shown.

Major variations of CPs occur at the molecules side chains, with halogenated and sulfated modifications. Although the structure of CPs is unique, compounds sharing the same Ahp residue have other trivial names indicative of the producing organism, such as micropeptins, oscillapeptins, nostopeptins, aeruginopeptins and anabaenopeptilides.

Limited studies have detected CPs coincidentally with MCs during blooms and are generally considered as non-toxic, mainly known for their inhibition of serine proteases like chymotrypsin or trypsin [135,138,143]. However, CPs are potent serine protease inhibitors including trypsin, chymotrypsin, human kallikrein, plasmin, thrombin, and elastase in the low nM range (Table 2) [126,129,144,145]. The protease inhibitory activity of CPs is structurally linked to the Ahp and position 5 residues, indicative of enzyme specificity and inhibition potency [146]. CPs can be designated into structural subclasses based on the amino acids occupying position 2 and 3, adjacent to Ahp moiety. For example, 51% of reported congeners have the partial sequence Lxx-Ahp, 32% contain Phe-Ahp, and 17% comprise other partial amino acids sequences – three broad subclasses of CPs [141]. Examples of the Lxx-Ahp and Phe-Ahp partial sequence are shown in Figure 7. Position 2 can further designate CPs with a *NMe*Phe (31% of reported congeners), and an *NMe*Tyr (36%)

[141]. The remainder of reported congeners contain, *NMe*Lxx, *NMe*Gle, *NMe*Hty, *NMe*Arg, *NMe*Lys, *NMe*Trp or methylated and/or halogenated *NMe*Tyr [126]. The toxicity of CPs has been demonstrated with few model organisms to evaluate ecotoxicological potential. For example, mortalities have been observed in freshwater crustaceans (*Daphnia* spp.; *Thamnocephalus platyurus*, LC_{50} = 8.8 µM) under experimental conditions [144,147,148]. CP 1040 is observed to induced non-specific developmental toxicity of zebrafish embryos (i.e., endpoints of pericardial and yolk sac edemas) [134]. CPs exhibited toxicity in Nematodes (*Caenorhabditis elegans*) exposed to 10 µg/L CP 1007, -1020, and -1041 showing reduced reproduction, delayed hatching, decreased growth rate, shortened lifespan [135.] For the same study, CPs had displayed a comparable toxicity to MC RR but evaluated APs were most potent [135] Resembling MCs, CPs are detected in natural blooms but commonly at lower concentrations [138,143,149]. The limited studies that reported environmental concentrations of CPs have observed similar ranges to that of APs and MCs, with production closely associated to cell abundance [129,138,149]. CPs are inhibitors of serine proteases in the nM range with specific variants displaying potent trypsin and factor Xla inhibitory effects (Table 2).

Table 2. Enzyme inhibition associated with the CP group

Enzyme inhibition	IC_{50} Inhibition Range (nM)	Reference(s)
trypsin	0.67 – > 40,000,000	133,139,144,145,150
chymotrypsin	2,500 – > 40,000,000	133,139,144,145,151
plasmin	26,000 – > 500,000	144,150
PP_1	*11,000,000 – > 50,000,000	145,150,152
thrombin	2,750,000 – > 50,000,000	132,133,140
human leukocyte elastase	85,000 – > 50,000,000	133,145
factor Xla	3.9	133

* = CP containing bloom extract

IC_{50} values reported in µg/mL and µM were converted to nM

Relative to MCs, CPs do not show the potent inhibition of PP enzymes however, their eco-toxicological potential remains a concern.

1.6.3. Microginins (MG)

Microginins are a group of linear lipopeptides formed by four to six amino acids with an N-acyl terminus. Other variants within this group have been named oscillaginins (OS) and nostoginins, from genus it was produced such as, *Oscillatoria* and *Nostoc*, respectively [153]. The MG group contains diagnostic octanoic or dodecanoic acid derivatives at their N- terminus, representative of the PKS synthesis pathway [126,154]. The common N-terminal diagnostic residues are Ahoa (3-amino-2-hydroxy octanoic acid) and Ahda (3-amino-2-hydroxy decanoic acid) which can be methylated and/or halogenated (typically with chlorine) [154]. For example, chlorination of the terminal Ahda carbon occurs in the reported congener oscillaginin A (OS A). The Ahda moiety is most prevalent, found in over 90 reported microginin congeners, making up 69%. Whereas the Ahoa moiety is found in only 6% [142].

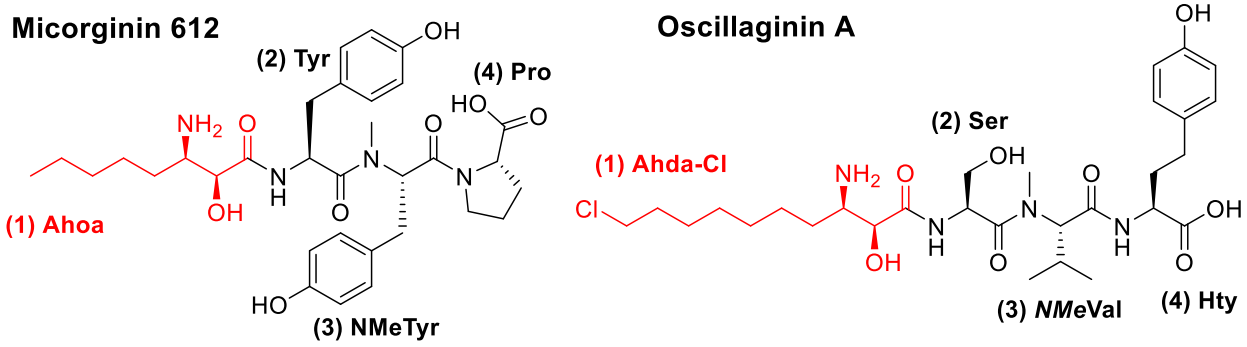

Figure 8. Chemical structures of MG 612 (left) and OS A (right). The characteristic methylated Ahda and Ahoa structural moiety is indicated red. In the case of OS A, the N-terminal Ahda is chlorinated

Methylation of the amino functionality occurs in the reported congener MG 612 (Fig. 8). Position 3 contains an *NMe*-amino acid, with no specificity towards the chemistry of the substituted amino acid seen with the substitution from *NMe*Tyr to *NMe*Val (Fig. 8). MGs have been shown to inhibit exopeptidases such as, leucine aminopeptidase, aminopeptidase M (APM) and angiotensin converting enzyme (ACE) at low µM concentrations (Table 3) [126,154,155]. Co-occurrence of MGs

with MCs has been reported during cyanobacteria bloom events. Exposures of cyanobacterial extracts primarily containing MG FR3 exhibited mortality to the crustacean, *Thamnocephalus platyurus* (LC_{50} = 0.99 mg/mL) [156]. Several studies report toxicity of MGs using cyanobacteria extracts exposers to model organisms however, assays using the purified compound are required to further evaluate MG toxicology [156,157]. Little is known about the structural features contribution to MG inhibition and the groups relative natural concentrations.

Table 3. Enzyme inhibition associated with the MG group

Enzyme inhibition	IC_{50} Inhibition Range (μM)	Reference(s)
leucine aminopeptidase	0.98 – > 5000	158–160
aminopeptidase M	2.7 - > 100	155,160
angiotensin-converting enzyme	6.8 – > 1000	155,158,159,161,162

IC_{50} values reported in μg/mL and μM were converted to nM

1.6.4. Cyanobactins (CB)

Cyanobactins are a large group of ribosomal synthesized natural products that undergo a variety of post-translational modifications described below [70]. Cyanobactin biosynthetic gene clusters are common within cyanobacterial species (10–30 % of species) [73]. Cyanobactin gene clusters are most frequently found within the cyanobacteria genera *Scillatoria, Arthrospira,* and *Microcystis.* The term cyanobactin encompasses cyclic peptides containing heterocyclized amino acid residues, commonly azol(in)es) and prenyl groups (Fig. 9). The term cyanobactin encompasses cyclic peptides within heterocyclized amino acids and further grouped into tenuecyclamides, trichamides, lyngbyactins, microcyclamides, aerucyclamide, aeruginosamides, and anacyclamides. The heterocyclized residues of oxazoline and thiazolines can be further oxidized to oxazoles and thiazoles, increasing the structural diversity of the group [70]. The diversity of cyanobactins originates from the biosynthetic gene clusters of the group that posses, large variations of core peptide sequences, unique combinations of post-translational modifications,

and differing lengths of the mature peptides [70,73]. Currently, not all novel products emerging from the cyanobactin family adhere to the original definition of cyanobactins. For example, the reported cyanobactins, anacyclamides, kawaguchipeptins, prenylagaramides, piricyclamides, scytodecamide and trikoramide do not possess heterocyclized amino acid residues. In addition, cyanobactin synthesis pathways include highly modified linear peptides with rare N-prenylated and O-methylated terminals [70,73]. Many cyanobacteria genera produce cyanobactins with over a hundred variants identified [163]. The quantity and type of heterocycles and, number of proteogenic amino acids generally dictate the subclassification of cyanobactins.

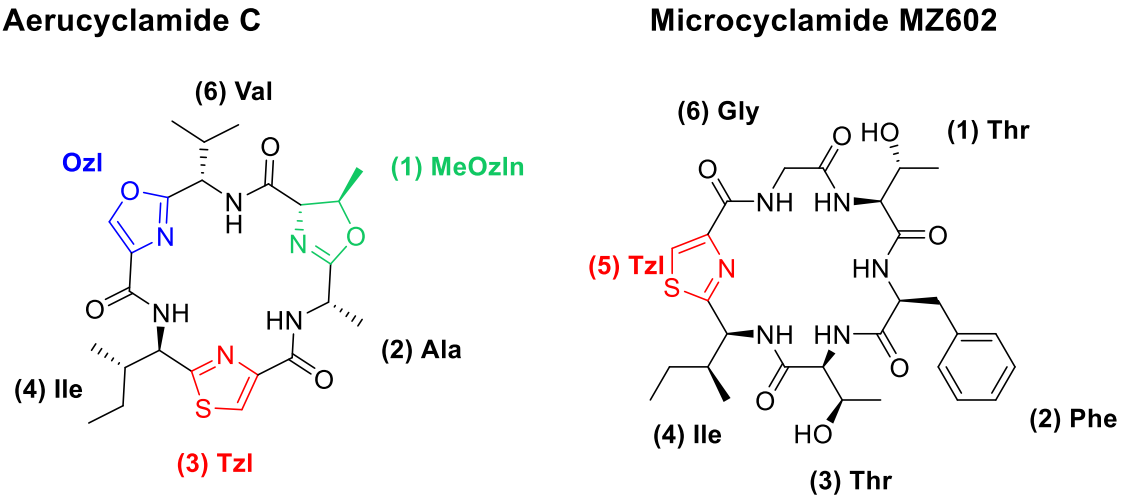

Figure 9. Chemical structures of AC C (left) and MA MZ602 (right). The characteristic heterocylized thiozol (red) and oxazol (blue) and methoxazoline (green) structural moieties are indicated.

Cyclic cyanobactins typically possess six residues (Fig. 9). In the case of aerucyclamides, they possess three proteinaceous amino acids and three heterocycles in alternating order (Fig. 9). A vast amount of structural variation arises with substitution of the proteinaceous amino acids, the heteroatoms within the heterocycles and the oxidation state of such heterocycles. For example, both oxazolines and thiozolnes and their oxidized residues are found within variants of microcyclamides and aerucyclamides. CBs often show cytotoxicity against different cancer cell lines but have also shown antiviral, antimalarial, and allelopathic activities [70]. Additionally, CBs pose eco-toxicological effects to model organisms. For example, Aerucyclamides (AC) A and B

have shown toxicity to fairy shrimp (*Thamnocephalus platyurus*) with LC_{50} values of 30.5 and 33.8 µM [163]. Currently, it is unclear if CB toxicity is ecologically relevant as limited data is available on their bioactive properties and few studies have produced data comparable to other CNPs groups.

1.6.5. Aeruginosins (AS)

Aeruginosins are a group of small linear tetrapeptides that contain the diagnostic non-proteinaceous amino acid 2-carboxy-6-hydroxyoctahydroindole (Choi; Fig. 10). The AS group consist of > 40 known structural variants [126,164]. The C-terminus of ASs contain either an, Arg, agmatine (Agm), argal, argininol or argininal. The N-terminus of ASs contain a p-hydroxyphenyl lactic acid (Hpla) derivative that can be halogenated or sulfated (Fig. 10) [164]. These metabolites typically appear in structurally related mixtures of compounds with substitutions of the proteinaceous amino acids in position 2 (Fig. 10). The hydroxyl side group of the Choi and the three R positions of the Hpla residues can be functionalized with halogens, sulfates, phosphates, and other moieties (Fig. 10). ASs are known serine protease inhibitors like variants of the CP and AP group [139]. Several marine and freshwater cyanobacteria genera produce AS variants such as *Microcystis, Oscillatoria, Nodularia* and *Planktothrix* [165]. ASs have been found to inhibit, chymotrypsin, elastase, plasmin, factor XIa and thrombin in the µM range (Table 4) [126,129,166–170]. The group is associated with potent trypsin inhibition, possessing IC_{50} values in the nM range (Table 6) [171].

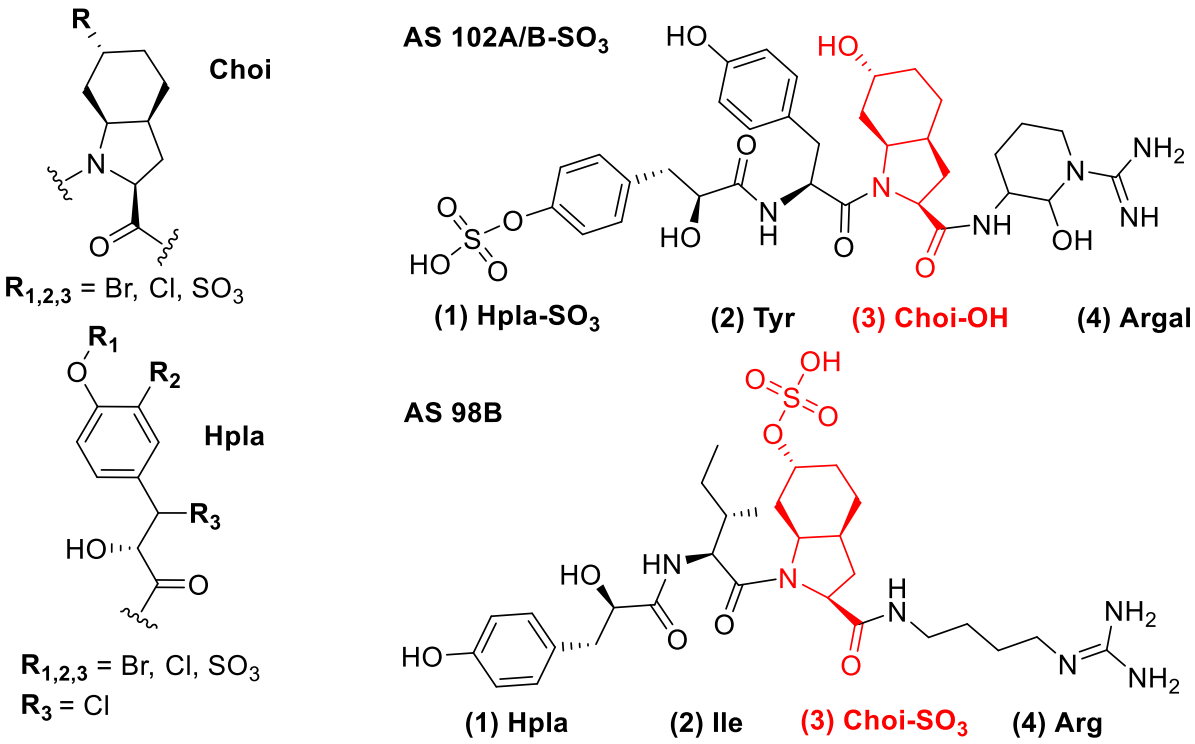

Figure 10. (left) The Choi (2-carboxy-6-hydroxyoctahydroindole) and Hpla (a p-hydroxyphenyl lactic acid) structural moieties and their substitution positions. (right) Chemical structures of AS 102A/B-SO_3 (top) and AS 98B (bottom) with choi indicated red.

ASs possibly retain the greatest trypsin inhibitory effects comparative to other CNPs groups and second to that of CPs. Major parts of AS biosynthetic gene clusters are highly similar between cyanobacterial genera; however, the genes coding tailoring enzymes largely differ, resulting in the substitutional variation within the group [126,165].

Table 4. Enzyme inhibition associated with the AS group

Enzyme inhibition	IC_{50} Inhibition Range (μM)	Reference(s)
trypsin	0.0047 – >100	139,166,168,169,171
chymotrypsin	13.9 – >100	167,169
thrombin	0.04 – >100	166–170,172
elastase	28 – >100	167,169
factor Xa	26* – 140	167,170

IC_{50} values reported in μg/mL and nM were converted to μM

1.6.6. Microviridins (MV)

MVs are one of largest CNP produced by cyanobacteria, ranging from 12 – 20 amino acids [173]. Most MV possess a mass between 1400-2000 Da, with 14 amino acids. Several freshwater genera of cyanobacteria are known to produce this group of compounds via their RiPP biosynthetic pathway (i.e., *Microcystis, Planktothrix, Anabaena, Nostoc* and *Nodularia*) [154,173]. MVs possess a characteristic tricyclic structure architecture inherent of the lactone and lactam linkages. Typically, the distinct rings of MVs are formed by two ω-ester and an ω-amide bond, cyclized by post-translational ATP-ligase enzymes. Mono- and bicyclical structures may also be formed, potentially resulting from absence of PTM enzyme(s) (i.e., MV SD1652) [72,173,174]. Without characteristic non-proteinaceous amino acids, classification this chemical class in respect to other CNPs relies on the presence of uncommon proteogenic amino acids (i.e., Trp). However, analytical parameters such as, LC-MS retention time ranges and doubly and triply charged precursor ions, aid in designating the group. The original, MV A was first isolated in 1990 from toxic *Microcystis viridis* (NIES-102) in Kasumigaura Lake (Japan) with the sequence Ac-Tyr(I)$_1$-Gly(I)$_2$-Gly(I)$_3$-Thr$_4$-Phe$_5$-Lys$_6$-Tyr(II)$_7$-Pro$_8$-Ser$_9$-Asp$_{10}$-Trp$_{11}$-Glu(I)$_{12}$-Glu(II)$_{13}$-Tyr-OH$_{14}$ [175]. Other MV congeners generally substitute various amino acids and ester linkages at different positions, with positions 1, 2, and 5 appearing most variable [173,176]. MVs have shown enzyme inhibition activity against serine-type proteases with potency linked to residues position [173,176,177]. For example, MV J, possess a unique substitution of Arg at position 5, suggested to be the key binding site for potent trypsin inhibition with this congener [177]. Furthermore, MVs containing Leu at position 5 (i.e., MV B, C, G, H,) are potent elastase inhibitors with IC_{50} values ranging from 0.01 – 58 μM [173]. In contrast, MV congeners with Met, Tyr, or Phe (microviridin D, E, F) display little or no activity [173,177]. Typically, most MV congeners possess the lowest IC_{50} values for Chymotrypsin ranging from 0.1 – 60 μM [173]. The large number of residues and PTM enzymes involved in microviridin

biosynthesis, offers various chemical modifications and amino acid substitutions resulting in the high chemical variation within this CNP group.

1.7. LC-(HR)MS/MS

(Liquid chromatography and high-resolution mass spectrometry)

The detection of single analytes from environmental samples, which frequently contain complex mixtures of chemicals is a challenge. Integration of multiple analytical techniques can overcome this issue. Liquid chromatography (LC) provides effective separation of complex chemical mixtures and mass spectrometry (MS) can give chemical data for all ionizable compounds, providing retention times (RT) and precursor *m/z* of each chemical. In MS mass detection is performed by a mass analyzer which manipulates ionized compounds, assigning specific mass to charge (*m/z*) ratios then subsequently, outputting them to the detector where they are detected and converted to a digital signal [178]. Recent advances in LC-MS instrumentation have demonstrated utility in non-targeted data collection and the analysis of complex sample mixtures. Non-targeted LC-MS workflows generally consists of three distinct steps: 1. separation, 2. ionization and 3. MS method and instrument. The first requires chromatographic separation of the sample, either gas or liquid. Typically, high-performance liquid chromatography (HPLC) or ultra-performance LC (UHPLC) are applied in high-resolution metabolomic studies with liquid samples. UHPLC offers greater sample throughput although, both chromatographic methods are reliable for effective separation of chemical constituents. Second to separation by chromatography, ionization of the metabolites entering the mass spectrometer is an important step of MS experimental workflows. A variety of ionization methods exist such as, electron ionization (EI), direct analysis in real time (DART), inductively coupled plasma (ICP), matrix assisted laser desorption ionization (MALDI), atmospheric-pressure chemical ionization (APCI), electrospray ionization (ESI) and desorption-ESI (DESI) [178,179]. For most (HR)MS studies electrospray

ionization (ESI) is most often used, dependant on a samples physical state [180]. In LC-MS workflows, ESI is the optimal ionization method and used for multiple reasons. The main advantage to ESI is its ability to effectively ionize and handle larger thermally unstable biomolecules such as, proteins, peptides, oligopeptides, and some inorganic molecules [178,181]. Furthermore, ESI is a soft ionization method – meaning, it does not readily induce in-source fragments of the precursor molecule. Therefore, the resulting MS spectrum contains the dominant molecular ions and adequate ionization of both large and small biomolecules within the liquid phase. A common feature of ESI is the generation of salt adducts such as sodiated and ammoniated ions with the use of buffered solvent systems. This is advantageous as the combination of adducts with other observed ions can be confirmatory. Lastly, the MS method applied is directly related to the type of MS chemical data acquired dependant on the instrument, detector, and the experimental goal. For example, not all MS instrumentation possess a collision cell (i.e., MALDI) for acquisition of tandem MS (MS/MS) data. Within a collision cell analyte molecules collide with inert gas, commonly Argon, and the resulting fragments are detected and displayed as a product ion spectrum. Hence, the *m/z* of fragment ions in MS/MS spectra are representative of a chemicals structural features. MS/MS instrumentation with high resolution (HR) mass analyzers such as orbitraps or quadrupole time-of-flight (qTOF), can provide highly sensitive structural mass data with excellent resolution over wide detection dynamic range for most ionizable chemicals and product ions within a sample [182,183]. Historically, (HR)MS analyses were limited to double-focusing magnetic sector or Fourier transform ion−cyclotron resonance (FTICR) instrumentation [184]. Advances in mass analyzing technology has yielded new-age mass analyzers such as time-of-flight (TOF) and Orbitrap capable of providing (HR)MS/MS data with accessible bench-top instrumentation. Additionally, classical low resolution MS methods such as (i.e., Triple Quadrupole and MALDI) are enhanced by the incorporation HR mass analyzers.

For the current study, an Orbitrap single-configuration mass analyzer is used which provides high-resolution tandem mass spectral (HR)MS/MS data, ideal for sample-based profiling and metabolite identification [183]. Orbitraps are capable of performing data dependent (DDA) and data independent acquisition experiments. DDA methods collect both full scan MS and MS/MS spectra for ionized compounds that meet pre-defined criteria (i.e., precursor ion intensity, retention time, diagnostic product ions, etc.). Conversely, data independent acquisition (DIA) methods collect MS and MS/MS data for every ion within a defined *m/z* range. DDA modes are frequently applied to experiments regarding multiple analytes within highly complex samples such as, metabolomics, proteomics and lipidomics in clinical research, the food and beverage industry, drug discovery and regulatory toxin/contaminant screening [180,181]. DIA experiments are applied for similar instances as DDA and DIA is generally considered it non-biased successor, analyzing all compounds rather than a pre-defined selection. DIA methods generate vast amounts of data in comparison to DDA, hence the latter is still experimentally favored to optimize runtime duration and post acquisitional processing.

(HR)MS/MS chemical data provides structural information as product ions in MS/MS spectra, directly related to fragments of an analytes chemical structure. Additionally, (HR)MS/MS allows the signal-to-noise ratio to be substantially improved for most analytes. (HR)MS/MS analyses are well suited for applications that elucidate low-abundance, high complexity, or otherwise difficult samples. The collection of (HR)MS/MS data is vital for identification of reported compounds and classification of structurally related compounds since, compounds with shared structural features will generate common product ions upon fragmentation. Non-targeted DDAs provide high-resolution mass data for all abundant and ionizable chemicals within a sample [180]. Coupling (HR)MS/MS with chromatographic techniques streamlines analyses generating continuous chemical data of individual chromatographic peaks and associated structural information.

LC with (HR)MS/MS is capable of providing vast amounts chemical information for each ionizable metabolite within a sample with good reproducibility. The simultaneous collection of characteristic chemical data such as, retention time (relates to metabolite polarity), precursor ion *m/z*, and multiple product ions generated by MS/MS; for thousands of compounds in a given sample is optimal for the screening of natural product extracts. Full scan (HR)MS and (HR)MS/MS data provides isotopic analysis and a molecular formula based on the mass defect principle. Every isotope of an element has a unique mass defect, meaning their mass is slightly lower than the sum of the mass of their protons and neutrons [185]. Therefore, a metabolite will have a characteristic mass defect corresponding to the number and type of atoms it contains[185,186]. Similarly, isotopic peak distribution in full MS scans can indicate the presence of elements with characteristic isotopic rations. With determination of molecular formula for a chemical, identification is simplified by comparison to literature. A limitation to (HR)MS analyses is the inability to differentiate isobaric compounds (i.e., compounds with the same molecular formula). For example, the immonium ions generated for the amino acids leucine (*m/z* 86.0969) and isoleucine (*m/z* 86.0969), are isobaric, therefore cannot be differentiated with MS [178,186]. Overall, structural isomers and stereochemistry cannot be elucidated with (HR)MS data alone. The challenge of discriminating compounds based on a single HR molecular formula alone is optimized with comprehensive databases and libraries of previously of reported natural products. Cyanobacteria natural product spectral databases are essential for rapid dereplication of metabolites and distinguishing novel metabolites. Current research has focused on the creation of publicly available natural product databases, established by highly curated protocols. For example, the collaborative creation of CyanoMetDB in 2021 – a comprehensive public database of cyanobacterial secondary metabolites, opened the door for effective designation of cyanobacteria natural products and standard protocols for database curation of novel metabolites [53]. The chemical information provided by the acquisition of large (HR)MS and (HR)MS/MS

datasets is a gold standard for expansion of chemical libraries and rapid dereplication of metabolites in metabolomic studies.

1.8. METABOLOMIC ANLAYSES

Metabolomics is the comprehensive analysis of all metabolites within a biological specimen or samples at a given time, known as the metabolome [181]. Multiple -omics technologies exist such as genomics (DNA), transcriptomics (RNA), and proteomics (proteins), which measure the metabolic potential of organisms, whereas metabolomics provides detailed information on the organic compounds that are the endpoints of biological processes [181,187]. Metabolomics is not defined by any one particular experiment but encompasses the study of metabolomes in a comprehensive way. Most metabolomic workflows consist of extraction from a biological sample followed by chromatographic separation and mass spectrometry for simultaneous measurement of various metabolites [187,188]. Metabolomics studies can involve non-targeted screening methods where thousands of unknown metabolite features are profiled and compared against measured qualitative conditions [187]. Screening experiments prove useful in identifying novel metabolite natural products associated with a change in biological or physiological conditions. Semi-targeted metabolomics experiments are more frequently employed and often are more useful. Semi-targeted metabolomics can quantitate known metabolites and simultaneously detect thousands of unknown features that can potentially be identified by subsequent data processing methods [186,187]. Targeted experiments provide the greatest understanding with absolute concentrations of biomolecules but require analytical standards[187]. Metabolomic studies require complex analytical instrumentation coupled with post-data acquisition techniques to comprehend discover metabolic profiles and variations in biological systems [188,189]. There is no limitation on the sample types that are suitable for metabolomics; however, sample preparation must be in accordance with sample type and the specific metabolites of interest. Metabolomic data interpretation differs markedly depending on the biological system from which the metabolites originated [186,187]. Metabolomic

approaches are advantageous multi-tired analyses to comprehend the diversity of cyanobacteria secondary metabolites in naturally occurring blooms.

1.8.1. Non-targeted Metabolomics of Cyanobacteria Metabolites

Large (HR)MS and (HR)MS/MS datasets must be interpreted with sophisticated post-data acquisition strategies to comprehend the variation in metabolomic profiles [186,190]. Software-based multivariate analyses provide the means to visualize metabolomic variations. *R* language and statistical software and packages of *xcms*, *FactoMineR*, and *MetabolAnalyze* allow for comprehension of variables (metabolites) contributing to metabolome diversity. Further, multivariate statistical analyses can be used to assess and visualize the variation CNP metabolite profiles of cHABs. Specifically, principal component analysis (PCA), factor loading plot analysis are used to interpret peptide profiles. Subsequent statistical tests can indicate individual CNP contributions to overall profile variation and identify trends. PCA is a means of describing the greatest variability of a multivariate dataset in a 2-dimensional space [191]. Reducing the dimensionality of large datasets condenses variation within a dataset into fewer dimensions – most effective when working with large multivariate datasets [191]. This test is non-specific and converts variable covariances from a covariance matrix into sets of new uncorrelated variables deemed principal components. In essence principal components represent the dimensions which explain the variance, consist of eigenvectors and eigenvalues. Eigenvectors of the covariances represent the directions of the axes that explain variance. Eigenvalues are unit-scaled coefficients of eigenvectors and indicate the proportion of variance explained each for each principal component dimension. Therefore, the proximity samples plotted in principal component dimensions directly corresponds to similarities in the variables that define them are more similar than samples plotted further away along the principal components. Factor loading plots depict the individual covariances (eigenvectors) by scale magnitude (eigenvalues) for the selected principal components. These plots visualize the contribution of each variable to the distribution of the

samples within the PCA. Within LC-(H)MS based metabolomics, variables are identified compound peaks from LC, detected by (HR)MS/MS. Statistical tests for non-parametric data are well suited for chemical data generated from LC-MS. Non-parametric tests can indicate statistical significance of variables that define variance within the PCA. The current study applied, the non-parametric Kruskal-Wallis (KW) test to all features considered for PCA. A Benjamini-Hochberg correction was applied to reduce false positives.

1.8.2. Semi-targeted Metabolomics of Cyanobacteria Metabolites

The chemical information provided by the acquisition of large (HR)MS and (HR)MS/MS datasets is optimal for semi-targeted metabolomic approaches. Screening and analysis of LC-(HR)MS/MS raw data for characteristic features for chemical groups such as retention time ranges, precursor ion *m/z* ranges, and diagnostic product ions can effectively filter massive datasets for all metabolites with a particular chemical feature relating to a chemical group or compound. Diagnostic fragmentation filtering (DFF) is one of such techniques, filtering large chemical data sets for identification of multiple diagnostic product ions. DFF is a feature of the open-source mass spectrometry data analysis software, mzMine2 and allows for the identification of diagnostic product ion pairs or combinations which may be tailored to a specific chemical class [141]. DFF advantageously uses structural similarities in product ion spectra to indicate structurally similar chemicals [141]. DFF searches output all metabolites within the large data set that contain specified product ions within their MS/MS.

1.9. PROJECT AIM

1.9.1. Chapter 2

Cyanobacteria harmful algae blooms (cHABs) release mixtures of biologically active CNPs into freshwater systems. Many of these poorly studied compounds pose undermined risks to both human and ecosystem health – having the potential to negatively impact local economies and

industries that depend on freshwater resources. The biological factors that promote cyanobacteria growth are broadly understood; however, the chemistry, biological activity, and environmental concentrations for the majority of CNPs are not. MCs, the most extensively studied CNPs group, however, much less is known about the chemistry, structural diversity, toxicology, and environmental concentrations of lesser studied CNPs groups of APs, CBs, CPs, MGs and ASs. The application of mass spectrometry and metabolomic data processing techniques are powerful tools used to decipher mixtures of compounds in environmental samples. The current thesis work applies LC-(HR)MS/MS and metabolomic analyses to decipher CNPs profiles produced by cHABs. The primary objective of this study was to determine the CNP profiles and their respective concentrations for fifty-five bloom samples from fifteen watercourses collected from eastern Ontario and western Quebec. The study will further our understanding of the toxigenic potential of cyanobacteria populations from different freshwater systems in the region and help prioritize common and abundant cyanopeptides for further study (hazard characterization and exposure assessment). Here, a DDA LC-(HR)MS/MS approach was used to detect 117 unique CNPs from intracellular extracts of cyanobacteria bloom samples, focused on the groups of MC, AP, CP, MG, CB, and AS. MCs and select other CNPs were quantified with reference materials. Using a LC-(HR)MS/MS metabolomic approaches of Global Natural Products Social (GNPS) Molecular networking (MN) and diagnostic fragmentation filtering (DFF), the discrete CNPs profiles of the fifty-five cyanobacteria algal blooms were elucidated and bloom concentrations of APs, CPs, and MGs determined semi-quantitatively. The goal of this chapter is to gauge the toxigenic potential of some less studied CNPs groups (i.e., AP, CP, MG) with comparison of their relative concentrations and ubiquity to that of MCs and others.

1.9.2. Chapter 3

Few long-term records of cyanobacteria occurrence exist from both monitoring programs and sediment archives. This is a challenge when assessing whether the frequency and magnitude of

cHABs has increased in recent years or whether this trend is related to improved monitoring efforts. MCs are regarded as the most prevalent cyanotoxins in freshwater systems. Upon senescence of cHABs, MCs can subsequently deposit into surface sediments and adsorb to particulates. The detection of MCs in lake sediment has been explored by few previous studies and existing analytical methods require laborious extraction procedures. The overall goal of this section is to develop, validate and apply an efficient and effective analytical method for MC detection in lake sediment. Here, an (HR)LC-MS, (HR)MS/MS analytical method was developed and validated for the extraction and quantification of MC LR, [Dha^{7}]MC LR, MC RR and MC LA in lake sediments. MCs were sequentially extracted by sonication, vortex and orbital shaker and recovered by SPE. MC separation and quantitation is achieved by HPLC coupled to a Quadrupole Orbitrap mass spectrometer. To evaluate the MC method's use, it was applied with two lake sediment cores collected from the Rideau Canal system along with radioisotope dating (i.e., ^{14}C and ^{210}Pb), geochemical (i.e., Itrax-XRF) and Chlorophyll-a analyses. Historical Rideau Canal conditions were inferred using a ^{210}Pb chronology and a multiproxy approach that combined geochemical, Chl-a and MC data. This study provides an effective analytical method for quantitation of MCs in Lake sediment to further the understanding of cHAB occurrence and toxicity with respect to a historical context.

CHAPTER 2: Deciphering the Cyanopeptide Diversity of Cyanobacteria Blooms in Ontario and Quebec

2.1. INTRODUCTION

Canada is richly endowed with freshwater resources possessing 7% of the Earth's renewable water supply, the third largest worldwide [192,193]. The Laurentian Great Lakes are a prime example, straddling the Canada-U.S. boundary, posses 18% of the world's freshwater [192]. With 563 lakes

larger than 1003 km^2 and an estimated total of 31,752 freshwater lakes, Canadian freshwaters are both a highly valuable resource and commodity [193,194]. Throughout the 20th century anthropogenic driven changes in land use, invasive species and climate have caused pervasive change in freshwater lacustrine and fluvial systems [195,196]. In many of the settled areas of Canada, freshwater is polluted and is unsuitable for human consumption, recreational and industrial use - usable only at the high cost of treatment [194]. The provision of safe drinking and recreational waters is an ongoing challenge that requires a multi-barrier approach including, water quality management and monitoring, government legislation, public involvement, regulatory guidelines, technological solutions, and continued research [193,194]. In drinking water management, monitoring source waters is a crucial step to protecting public and environmental health. Source water protection is a vital component of water resource protection in Canadian policy and frequently involves a multi-tiered approach among stakeholders enhance water quality or control pollution. Currently, Canada has no enforceable standards for safe drinking water and no extensive federal water-monitoring program to provide detailed, timely reporting on the state of water resources [197,198]. At the community level, Ontario Conservation Authorities (CAs) provide regular water quality monitoring. However, this is not consistent across provincial boundaries, and the size of monitoring programs is limited by funding. In 2016, the Canadian Lake Pulse Network was launched, an academic-government research partnership monitoring of 680 lake watersheds [197]. This is the most comprehensive water-quality monitoring network to date in Canada. As monitoring efforts increase so must the screening capability of analytical methods to encompass potential new contaminants and toxins. Recently, cHABs have frequently posed an adverse challenge for freshwater monitoring and management, mainly as a result of the harmful metabolites they produce. For example, cyanotoxins (toxic cyanobacteria secondary metabolites) are the greatest concern due to the risks they pose to human health. Positive correlations between toxic cyanobacteria and adverse health effects have been well demonstrated in several countries including Canada [32,124]. Many freshwater cHAB studies report the increasing frequency and

magnitude of blooms since 1990s as a consequence of anthropogenic driven environmental change [38–40]. The common bloom-forming cyanobacterial genera of *Microcystis, Aphanizomenon, Cylindrospermopsis, Dolichospermum, Nodularia, Planktothrix*, *Oscillatoria*, and *Trichodesmium* have all been shown to produce cyanotoxins [199]. The variability and diversity of toxin production by cHABs is a specifically challenging subject for water quality management. Toxic cyanobacteria species are globally distributed and can produce complex mixtures of cyanotoxins specific to the stain level [54,154]. Several groups of cyanotoxins exist, with potential hepatotoxic, cytotoxic, and neurotoxic effects. The complexity and variability of cyanotoxins creates a surmounting challenge for water quality monitoring and treatment. Microcystins (MCs), cylindrospermopsin, anatoxins and saxitoxins are all common classes of cyanotoxins found globally [67]. Arguably, MCs are the most studied cyanotoxin group being, peptidic hepatotoxins and possible human carcinogens (IARC group 2B) [119]. Beyond MCs there exists a suite of cyanobacterial peptides or peptidic metabolites (cyanopeptides; CNPs) whose bioactivities and environmental concentrations are generally unknown [129,142]. To date, over 2000 CNPs (including cyanotoxins) have been isolated from cyanobacteria [53]. The CNPs groups of Aeruginosin (AS), Anabaenopeptin (AP), Cyanobactin (CB), Cyanopeptolin (CP) and Microginin (MG) have displayed bioactivity and their co-occurrence with cyanotoxins are seldomly reported during bloom events [129,143]. These less studied cyanopeptides environmental concentrations and eco-toxicological effects in Canadian freshwaters remain unknown. Representative structures of these CNPs groups are shown in Figure 11. Recent studies suggest that bioactive CNPs mixtures can affect the cohabiting aquatic organisms differently – outlining the need to monitor and potentially regulate CNPs for the protection of ecosystem health [143].

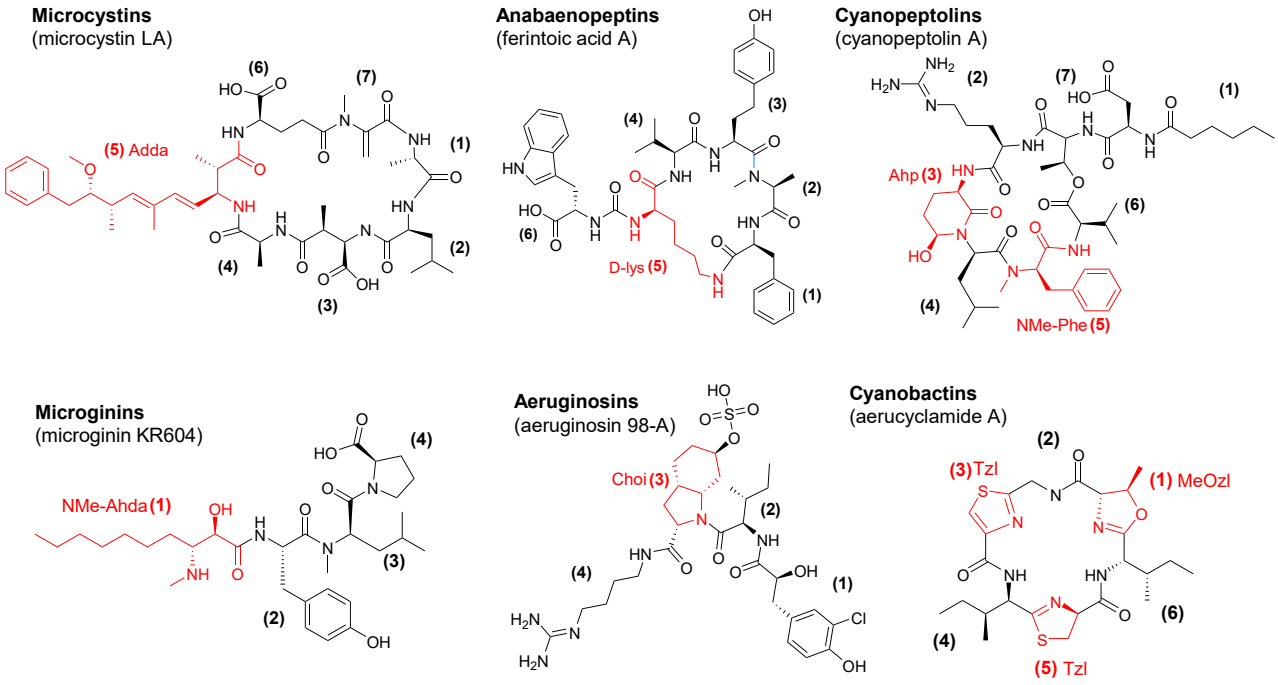

Figure 11. Representative structures of CNP groups detected from blooms in this study. Structural features characteristic to each CNP group are annotated in red.

The chemical complexity and diversity associated with CNPs is a difficult challenge for risk assessment activities. To overcome the diversity and mixture complexities of CNPs, they require sensitive and precise analytical methods for detection and differentiation of metabolite profiles. Non-targeted liquid chromatography with high resolution tandem mass spectrometry (LC-(HR)MS/MS) provides the ability to rapidly profile complex metabolite mixtures from crude natural product extracts [142,186]. Combining, LC-(HR)MS/MS analyses with post-acquisition metabolomic strategies such as, Global Natural Product Social (GNPS) molecular networking (MN), diagnostic fragment filtering (DFF) and multivariate tests can enhance the comprehension and interpretation of metabolite profiles [142,143,188].

Application of highly sensitive and accurate mass analyzers (i.e., orbitraps and qTOFs) using data-dependent acquisition (DDA) experiments can generate reliable (HR)MS/MS datasets for metabolomic studies – providing mass data for the majority of ionizable chemicals within a sample. DDA acquisition methods are a requirement for (HR)LC-MS metabolomic analyses as

they retain both (HR)MS and (HR)MS/MS spectra for ionized compounds that meet pre-defined analysis criteria. This enables the collection of vast amounts of (HR)MS/MS data for most ionizable chemicals in a sample such as, retention time (RT), precursor ions and product ions. Large LC-(HR)MS/MS datasets provide shared structural features as common product ions upon fragmentation, vital to dereplicate previously reported compounds and classify structurally related compounds. CNPs groups have characteristic structural moieties which define them (Fig. 11). Hence, structural moieties of a respective cyanopeptide group will generate shared product ions in MS/MS. Furthermore, variation within the peptide sequence of CNPs can be elucidated with the use of shared immonium ions of proteinaceous and non-proteinaceous amino acids. The use of positive heated electrospray ionization ($HESI^+$) in MS can cause the formation of mass adducts during ionization, these are charged molecules that can be manipulated by the mass spectrometer and generate multiple precursor ions for a single compound. CNPs are similar to other polypeptide natural products and will form common mass adducts such as sodiated $[M+Na]^+$, dehydrated $[M+H-H_2O]^+$, ammoniated $[M+NH_3]^+$ and doubly charged $[M+2H]^{2+}$ ions apparent in mass spectra the same charged precursor state. Although, multiple precursor adducts can complicate metabolomic analysis, they can also be confirmatory when present. Spectral complexity is detrimental to CNP specificity in LC-MS approaches, particularly when multiple adducts corresponding to the same CNP can occur, therefore MS/MS scans are more accurate for quantitation. Moreover, MS/MS fragments associated with a neural loss of the precursor ion (i.e., H_2O, CO, CO_2 and SO_3) are useful in confirming compound identity. Putative structural information is generated for each compound by comparison of MS and MS/MS data and proposed fragmentation patterns of the reported structure. A limitation to MS approaches is the inability to differentiate isobaric compounds. For example, Lxx is used to annotate either leucine (Leu) or isoleucine (Ile) except when reporting a known compound. Combining, LC-(HR) MS/MS analysis with metabolomic tools of GNPS MN algorithm and DFF creates a synergistic approach towards the dereplication of natural products and discovery of novel compounds. Further introduction of

multivariate analysis can elucidate the variability between metabolite profiles and identify commonly detected metabolites. These metabolomic tools enable the curation, visualization, and interpretation of thousands of obtained MS/MS data from analytes within an extracted environmental sample, an otherwise surmounting task to perform manually. The utility of metabolomic profiling approaches has been demonstrated in cultured cyanobacteria and can be extended to naturally occurring cyanobacteria samples [142,143]. It has been demonstrated that LC-(HR)MS/MS metabolomic analyses of complex cHAB CNPs mixtures can provide a comprehensive understanding of metabolite profiles, and aid in both dereplication, and novel discovery of metabolites produced by cyanobacteria [142,143,200]. For the current study, a LC-(HR)MS/MS metabolomic approached was used to identify more than 100 CPNs within intracellular extracts of cyanobacteria bloom samples from fourteen Canadian lakes and one river. Nine MC congeners and the reference materials of CP 982, MG FR1 and FA A were quantified with targeted LC-(HR)MS/MS methods. A semi-quantitative LC-(HR)MS/MS method was used to quantify the CP, AP and MG CNP groups using a single reference material for each.

2.2. METHODS

1.1 Sample Collection:

Fifty-five cyanobacteria bloom (cAB) samples were collected from fifteen separate freshwater watercourses throughout the years of 2009 to 2019. The sampling region and each watercourse location is depicted in Figure 12. The sampled watercourses were located in the Canadian capitol region of Eastern Ontario (ON) and Western Quebec (QC; Fig.12).

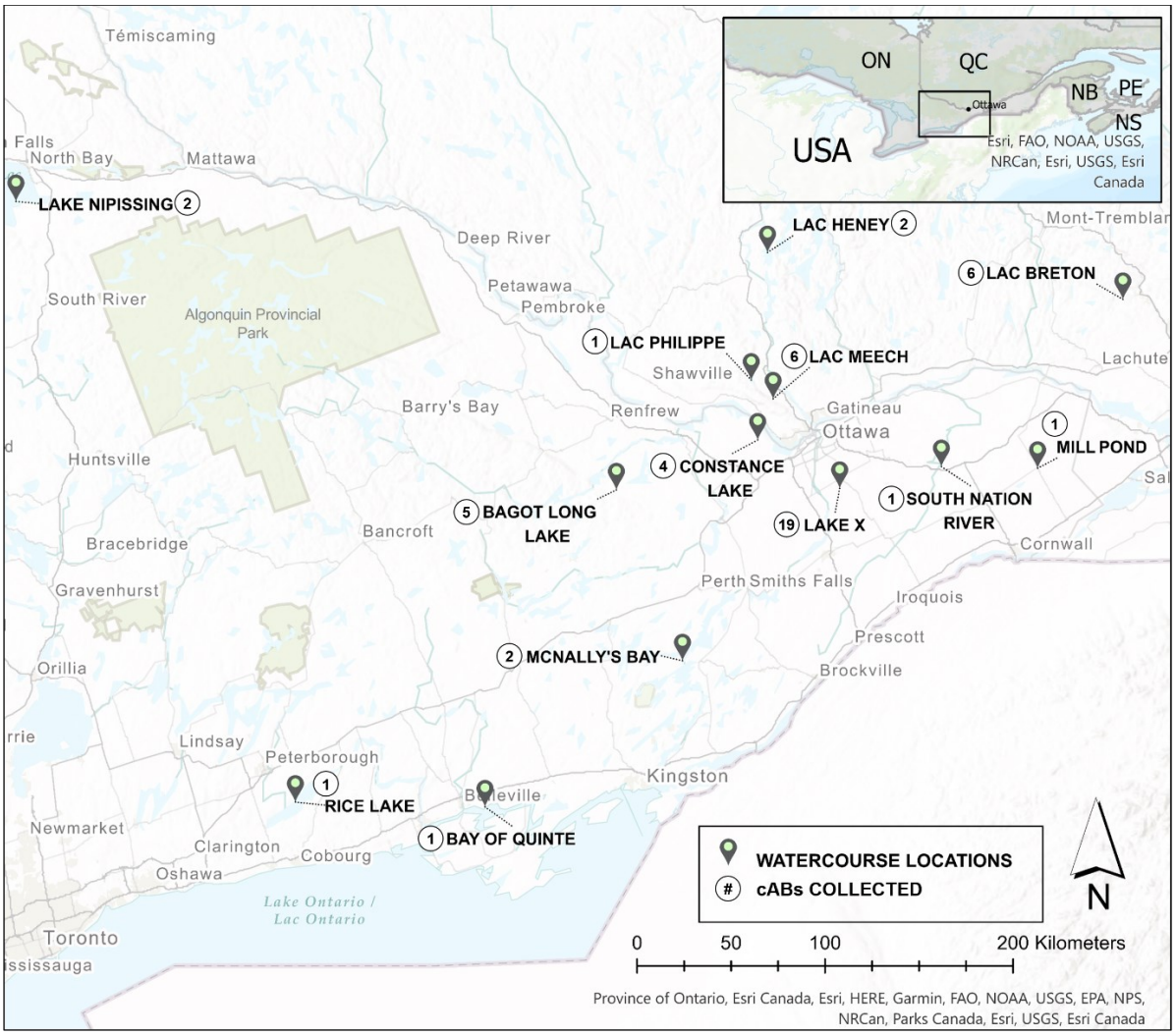

Figure 12. Arc®GIS Pro generated map, indicating watercourse locations and number of collected cyanobacteria algae bloom (cAB) samples. Watercourse names are annotated with text callouts at each pinned location.

Each cyanobacteria bloom sample was collected by surface water grab sampling into a 1 L clear plastic wide-mouth sample jar. The samples were transported in a cooler with ice packs and placed in refrigerator storage at 4°C, prior to filtration and extraction. For the purpose of this study samples will be referred to as cyanobacteria algae blooms (cABs) rather than cHABs as not all blooms meet the predefined criteria of a harmful algal bloom. The majority of watercourses sampled are small freshwater inland lakes. With exception of the Bay of Quinte cAB from Lake Ontario and the South Nation River cAB, near Castleman ON, the only fluvial system sampled.

Two of the provided cAB samples (i.e., C-Shore and Standout) did not have available sampling data and were not included in Figure 12.

2.3. Experimental

2.3.1.1. Reagents and Standards

MC certified reference materials of MC LR (CAS: 101043-37-2); [Dha7]MC LR (CAS: 120011-66-7); MC LA (CAS: 96180-79-9); MC RR (CAS: 111755-37-4), procured from the National Research Council of Canada's (NRC) Metrology Research Centre (Halifax NS, Canada). FA A (CAS: 176327-91-6) and MG FR1 (CAS: 11135249-50-1) refence materials were procured from Enzo Life Sciences, Inc. The individual material of CP (cyanopeptolin) 982 was purified in-house from a cultured M. aeruginosa CPCC 464. MC mixtures at 1 µg/mL wee prepared in 50% aqueous methanol (ddH2O and HPLC grade methanol, Sigma-Aldrich) containing standards of MC RR, MC LR, MC LA, [Dha^{7}]MC LR and NOD R.

2.3.1.2. Intracellular CNP Extraction

A 35 mL volume of each cAB sample was collected to a dry glass microfibre filter paper (Whatman, GF/C, diameter 47 mm, ~1.6 µm) by vacuum filtration (Millipore, stainless steel). The filter and harvested biomass were dried with an oven overnight (24 hrs) at 40 °C. Dry filters with biomass were transferred to test tubes (16 x 100mm, VWR) and 14 mL of 80% aqueous methanol was added. The tubes were vortexed and sonicated for 30 second intervals, then placed in a freezer at -20 °C for 1 hr. Samples were then thawed at ambient room temperature (≈24 °C). The vortex-sonication-freeze-thaw extraction process repeated twice more. The resultant methanolic extracts comprised of intracellular CNPs were passed through 0.22 µm PTFE syringe filters (Choice, ChromSpec, Inc.) into 7 mL amber glass vials (Type I, Class B amber borosilicate glass, Fisher Scientific; 15-425 threat cap with PTFE liner, Sigma-Aldrich) and dried under a gentle stream of nitrogen gas at 35 °C. Dried extracts were reconstituted with 1.5 mL methanol and

transferred to 2 mL amber HPLC vials (Agilent). HPLC vials were dried under a gentle stream of nitrogen gas and stored at -20 °C prior to LC-(HR)MS/MS analysis.

2.3.1.3. LC-(HR)MS/MS Conditions

All LC-(HR)MS/MS data was collected using an Agilent 1290 HPLC coupled to a Q-Exactive Quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific). All extracts were reconstituted in 1 mL of 90% aqueous methanol and 5.0 µL volumes were injected onto a C18 Eclipse Plus RRHD column (2.1 × 50 mm, 1.8 µm; Agilent Technologies) maintained at 35 °C and flow rate of 300 µL/min. The solvent program used HPLC-MS grade Water with 0.1% formic acid (mobile phase A) and acetonitrile with 0.1% formic acid (mobile phase B; Fig. 13).

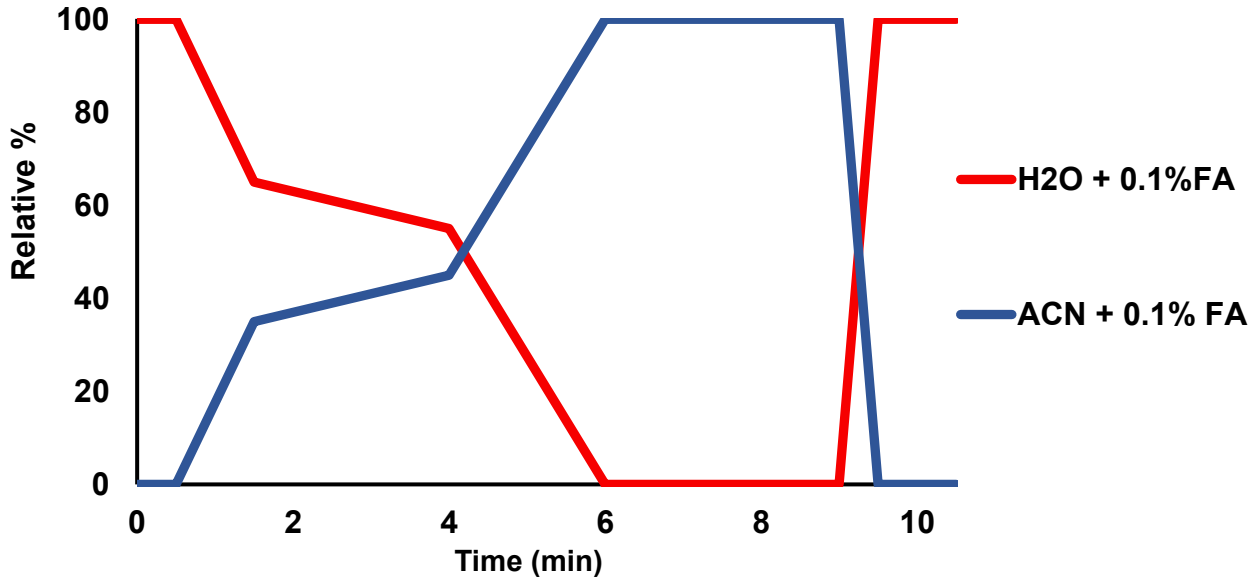

Figure 13. HPLC solvent program for LC-(HR)MS/MS analysis. The mobile phase consisted of acetonitrile (ACN) and water, each containing 0.1% formic acid (FA)

The 10.5 min method started with 100% mobile phase A. Mobile phase B was increased from 0% to 35% over 1.5 min and again to 45% over 2.5 min. Mobile phase B was increased to 100% B over 2 min, and held at 100% B for 3 min prior to returning to 0% B over 0.5 min and maintained for 1 min.

2.3.1.4. High resolution tandem mass spectrometry, (HR)MS/MS

All (HR)MS/MS data was obtained using a Quadrupole Orbitrap mass spectrometer equipped with a HESI-II probe, heated electrospray ionization source (HESI). The analytes were ionized with $HESI^+$ under the following parameters: capillary voltage, 3.9 kV; capillary temperature, 400 °C; sheath gas, 17 units; auxiliary gas, 8 units; probe heater temperature, 450 °C; S-Lens radio frequency level, 45.00. Targeted (HR)MS/MS data was acquired for MC LR, [Dha^7] -LR, -LA, -RR, -WR, -LY, -LF, -YR, FA A, CP 982 and MG FR1. A data-dependent acquisition (DDA) mode was used. This included a full MS scan at 35,000 resolution, with a scan range of 150.0000 – 2000.0000*m/z*; automatic gain control target, 1×10^5; and a maximum injection time of 128 ms. Analytes were targeted from each full scan for MS/MS analysis using 4.0-Da isolation and a 0.8s retention time window. MS/MS scans used the following conditions: resolution, 17,500; automatic gain control target, 2×10^5 ; max intensity threshold, 100ms. The specific LC-(HR)MS/MS parameters were used for each targeted analyte are listed in Table 5. Semi-targeted (HR)MS data was acquired using a DDA mode that included a full MS scan at 35,000 resolution, with a scan range of 100.000 – 1450.0000 *m/z*; automatic gain control target, 1×10^6; and a maximum injection time of 128 ms. The five highest intensity ions were selected for MS/MS using a 1.0-Da isolation window, and analyzed with the following conditions: resolution, 17,500; automatic gain control target, 1×10^6 ; max IT, 64ms; stepped normalized collision energy (NCE), 30-45; intensity threshold, 1.6×10^5 ; dynamic exclusion, 7 s. Precursor ion *m/z* and RTs for semi-targeted analytes are listed in Table S2.

Table 5. LC-(HR)MS/MS parameters for targeted quantification and LC-(HR)MS parameter for semi-quantification.

CNP	*m/z*	Formula	RT (min)	Quantifier, qualifier ion *m/z*	NCE	Method equation	R^2	LOQ (µg/mL)	LOQ RSD%
Targeted quantification									
MC RR	519.7915 $[M+2H]^{2+}$	$C_{49}H_{75}N_{13}O_{12}$	2.69	135.0804, 213.0858	38	$y = 1E^{+08}x + 771376$	0.999	0.001	3%
MC LR	995.5556 $[M+H]^+$	$C_{49}H_{74}N_{10}O_{12}$	3.30	135.0804, 213.0858	45	$y = 2E^{+07}x - 7637$	0.998	0.001	14%
[Dha^7]MC LR	981.5404 $[M+H]^+$	$C_{48}H_{72}N_{10}O_{12}$	3.34	135.0804, 107.0855	45	$y = 2E^{+07}x - 11645$	0.999	0.001	12%
MC WR	1068.5515 $[M+H]^+$	$C_{54}H_{73}N_{11}O_{12}$	3.58	135.0804, 159.0908	38	$y = 2E^{+05}x - 7496$	0.998	0.05	11%
MC LA	910.4921 $[M+H]^+$	$C_{46}H_{67}N_7O_{12}$	5.28	135.0804, 213.0858	41	$y = 9E^{+05}x - 1919$	0.999	0.05	4%
MC LY	1002.5183 $[M+H]^+$	$C_{52}H_{71}N_7O_{13}$	5.42	135.0804, 213.0858	41	$y = 6E^{+05}x - 2909$	0.999	0.01	10%
FA A	867.4400 $[M+H]^+$	$C_{46}H_{58}N_8O_9$	3.87	84.0810, 114.0547	38	$y = 2E^{+07}x + 87504$	0.999	0.001	9%
MG FR1	728.4230 $[M+H]^+$	$C_{38}H_{57}N_5O_9$	3.06	100.1120, 146.1167	33	$y = 2E^{+07}x + 12471$	0.998	0.001	13%
Semi-targeted quantification									
CP 982	965.4770, $[M+H-H_2O]^+$	$C_{38}H_{57}N_5O_9$	5.24			$y = 2E^{+09}x - 213394$	0.999	0.005	N/A
FA A	867.4400, $[M+H]^+$	$C_{46}H_{58}N_8O_9$	3.87			$y = 1E^{+09}x - 133071$	0.999	0.001	6%
MG FR1	728.4229, $[M+H]^+$	$C_{38}H_{57}N_5O_{10}$	3.06			$y = 1E^{+10}x + 235520$	0.999	0.005	10%

2.3.1.5. CNP (Cyanopeptide) IDENTIFICATION

Characteristic LC-(HR)MS/MS data such as, precursor ion *m/z*, MS/MS spectra, retention time and, diagnostic product ions and patterns were used to identify individual CNPs from the raw data files using *Xcalibur™ (version 3.0.63, Thermo Fisher Scientific)* software [201]. Previously reported CNP metabolites were identified using their characteristic LC-(HR)MS/MS data and novel CNPs were identified using LC-(HR)MS/MS data in addition to neutral losses, mass adducts and two metabolomic tools – GNPS MN and DFF (*MZmine2*). Chemical formulas, elemental compositions and ring-and-double-bond equivalents were determined from the (HR)MS data using *Xcalibur™ (v3.0.63)*. Potentially novel CNPs were designated into a group based on retention time range, precursor ion *m/z* range, presence of adducts, neutral loss fragments, diagnostic product ions and molecular network analysis. Putative structural information was generated for each compound by comparison of MS and MS/MS data and proposed fragmentation on patterns of the reported structures. Furthermore, structural variants of a CNP group were described by comparing the MS/MS spectra of reported CNPs and evaluating the associated mass error (ppm)

of the precursor ion *m/z* to a molecular formula. Additionally, small chemical modifications and amino acid substitutions were characterized by this method. Peak information of all individual CNPs was gathered into a comprehensive inclusion list for all fifty-five bloom samples. This included retention time, precursor ion *m/z*, adduct *m/z*, *m/z* of product ions, and which sample it was detected. The in-house inclusion list identified previously characterized CNPs by comparison of high-resolution precursor and product ion *m/z* and, acknowledged any CNPs not previously reported. A key limitation of this (HR)MS/MS based approach is the inability to discriminate residue configurations, nor isobaric residues. For example, Lxx is used to annotate either leucine (Leu) or isoleucine (Ile) except when reporting a known compound. Multiple adducts for the same metabolite can occur and were removed from the peak list by identification of the characteristic mass differences and conformations of MS/MS patterns.

2.3.1.6. Quantitation

Peak area integration and quantification was performed using *Xcalibur™ (v3.0.63)*. Analytes were quantified using a 1/x weighted linear calibration curve generated from the associated chemical standard, under identical analytical conditions. Calibration curves for all nine analytes were prepared at concentrations of 0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1 and 5 µg/mL. Automated peak integration was conducted using the Genesis peak detection algorithm, 7-point smoothing, enabled valley detection and 0.5 S/N threshold. Targeted quantitation used the (HR)MS/MS quantifier ion of each analyte (Table 5). Targeted analytes were designated as detected by the presence of both a qualifier and quantifier *m/z* in (HR)MS/MS. (HR)MS calibration curves of CP 982, MG FR1 and FA A were used semi-quantitate all identified CNPs of their respective group. A linear 1/x weighted calibration curve was used for semi-quantitation. The most intense precursor ion from (HR)MS scans of each non-targeted analyte was used for semi-quantitation. The method parameters for each CNP groups' semi-quantitated reference standard are listed in Table 5. The Semi-targeted CPs, APs and MGs were designated as detected by presence of the precursor ion

at a specific RT (Table S2). Semi-quantitated values ≥ 100 µg/g were rounded to the nearest tenth and those ≥ 1000 µg/g rounded to the nearest hundredth.

2.3.1.7. Limits of Detection and Quantitation

The limit of detection (LOD) was defined as the lowest concentration level of the calibration curve wherein five consecutive injections produced a detectable signal. The limit of quantitation (LOQ) was defined as the lowest concentration at which the peak areas' relative standard deviation (RSD) of five consecutive injections is below 20% [202]. All limits of the analytical methods are listed in Table 5. As there was no commercially available standard for all semi-quantitated CNPs, the concentrations were estimated using the retro analysis method, wherein the concentrations of a single available standard for the CNP group was used to calculate concentrations for all compounds of that respective CNP group using the calibration curve of the standard, assuming that both compounds ionize in a similar manner [202].

2.3.2. Metabolomic Processing

LC-(HR)MS/MS data files require multi-step metabolomic processing wherein .RAW files are converted to usable formats; pre-processing filters remove the majority of unwanted peaks and, manual processing further verifies and validates the peak dataset. Manual deconvolution is a necessary final step as many pre-processing filters cannot differentiate between a precursor ion and multiple mass adducts or neutral losses. Although this can complicate metabolomic analysis, adducts and neural losses can also be confirmatory when present in the same (HR)MS spectrum. Likewise, MS/MS fragments associated with a neural loss of the precursor ion (i.e., $-H_2O$, -CO, $-CO_2$ and $-SO_3$) are useful in confirming compound identity. Using the application *msConvert(v3.1.19)* and the peak picking filter of MS levels 1-2, all RAW LC-(HR)MS/MS data files were converted to a centroid format (.mzML). *msConvert* was procured from the open-source cross-platform software library *Proteowizard*. Further metabolomic processing of the .mzML files

used (1) *R(version 4.1.3)* language and packages (*xcms* [203], *MetabolAnalyze* [204], *FactoMineR* [205]), (2) DFF module from the open-source mass spectrometry platform *MZmine2* [206] and, (3) GNPS MN algorithm [207].

The metabolomic processing required for multivariate analysis of the large .mzml dataset utilized *R(v4.1.3)* language. Peak areas that met the parameters displayed in Table 6. were extracted from all .mzML files using the *R(v4.1.3)* xcms package. Subsequent manual deconvolution of the extracted peaks was done to verify designated CNPs.

Table 6. Peak picking parameter values applied with the bioinformatics R package, xcms.

Parameter	Polarity	Scan pre-filter*	*m/z* deviation (ppm)	s/n ratio cutoff	Peak width range (secs)	Noise level
Set point	+	5	1	5	12-15	$1E^{+6}$

*Number of scans detected with a minimum intensity of 5000.

This removed all peaks outside retention time window of 2.0 – 6.5 minutes and a *m/z* range of 450 – 1450 Da for most cyanopeptides. Any peak areas less than the same peak area of the blank multiplied by fifty, were removed. All extracted peaks were further assessed manually in *Xcalibur™(v3.0.63)* to confirm identity. Multiple adducts corresponding to the same CNP compound occurred and were removed by manual identification of mass shifts and MS/MS patterns.

METABOLOMIC ANALYSES

R Statistical Analyses: LC-(HR)MS/MS data

All processed peak areas were $\log_{10}$ transformed and pareto scaled using the *R(v4.1.3)* package metaboanalyze. The extracted log scaled peak areas were used to for subsequent multivariate analysis. A principal component analysis (PCA) was performed on 95 of the 117 CNP metabolites detected. Twenty-two peaks not conforming to the set peak picking parameters (Table 2) were not considered in the PCA. *R(v4.1.3)* packages of *factoextra* and *FactoMineR* were used to

generate PCA and loading plots. All dimensions of the PCA were regarded, with dimensions one and two describing the greatest variance for the dataset. A non-parametric KW test with was performed on the ninety-five CNPs considered in the PCA to investigate the statistical significance of each features' contribution to the overall variation between watercourse CNP profiles. Resulting p-values were adjusted using a Benjamini–Hochberg correction.

GNPS Molecular networking (MN) analysis

The GNPS MN algorithm was developed by researchers from the Scripps Institute and UC San Diego. MNs are a computational method that allows visualization and interpretation of the vast amount of chemical data acquired from MS/MS analyses. Therefore, GNPS MNs are visual displays of the chemical profile for samples analyzed by MS/MS [207]. Commonly, environmental and natural product extracts used to study metabolites are diverse mixtures containing both known and unknown metabolites. MNs operate on the established concept that that structurally related compounds will generate similar product ions and product ion patterns in MS/MS. This allows for identification of structural similarities throughout an (HR)MS/MS spectral dataset and propagation to potentially related unknown compounds. The GNPS MN algorithm uses a modified cosine scoring scheme that determines the spectral alignment (similarity) of two MS/MS spectra with COS scores ranging from zero (dissimilar) to one (identical) [207]. In a MN each chemical spectrum is represented as a node (circle), and spectrum-to-spectrum alignments as connecting edges (lines) between nodes. Thus, structurally similar nodes are connected by edges due to shared fragment patterns reflected in their MS/MS spectra.

The seeding GNPS MNs with known MS/MS spectral data provides the ability to anchor known metabolites to structurally similar unknown compounds – allowing quick dereplication of reported metabolites. This saves time and resources, allowing for greater focus on novel compound discovery. The open-source format of GNPS MN makes it an efficient and effective choice for research projects. To generate the MN, .mzML files were uploaded to GNPS network using

FileZilla (free FTP client; https://filezilla-project.org/). The.mzML data files considered in the MN analysis included, the fifty-five bloom samples, and the seeded spectra of all reference standards used (e.g., MC LR, [Dha^{7}] -LR, -LA, -RR, -WR, -LY, -LF, -YR, FA A, CP 982, MG FR1) and two extracts of previously studied cultured strains of *M. aeruginosa* (CPCC 300 and 632). The MN was generated using: a precursor ion *m/z* tolerance of 0.02 and a product ion *m/z* tolerance of 0.03, an MS/MS cosine similarity score of 0.6, a minimum of six matching fragment ions and cluster component size of two. All clusters with single or double nodes not corresponding to a CNP were removed manually. The network was filtered with laboratory blank MS/MS data, removing any peaks present in both the samples and the blanks. The produced .graphML file was extracted from the GNPS website and visualized using *Cytoscape(v3.9.0)* (https://cytoscape.org). Further manual deconvolution within the Cytoscape MN included removal of metabolites (1) outside the retention time range of 2.0-6.5 min, (2) outside a mass range of 450-2000 *m/z*, (3) unwanted mass adducts generated in HESI^{+}, and (4) any remaining single nodes not identified as a CNP. A GNPS spectral library search was preformed using all spectral libraries available (forty-four libraries) within the GNPS Spectrum Library Knowledgebase. All fifty-five samples were searched against the libraries to partially identify additional compounds, separate to CNPs.

DFF (diagnostic fragment filtering, MZmine2)

DFF is a semi-targeted metabolomic approach which targets a molecule's specific structural features as diagnostic product ions, searching an untargeted LC-(HR)MS/MS dataset for chemicals containing the specific molecular feature of interest. MS/MS spectra contain characteristic product ions that are unique to the structural features for an analyzed metabolite. Therefore, structurally related compounds will generate similar product ions within their MS/MS spectra. CNPs are grouped based on the presence of unique non-proteinaceous amino acids or partial amino acid sequences which generate diagnostic product ions that can be used for CNP group identification. Some identified molecular features of CNP groups are identified and their

product ions summarized in Table 7. These structural features were targeted using DFF in this study of CNP diversity. DFF is a complementary to other metabolomic methods, allowing for quick and efficient searching of diagnostic product ions without the limitation of additional parameters (s/n ratio, noise level, peak width). Most useful when searching for product ions at low noise levels (below 50,000) – not always feasible for parametrically constrained metabolomic approaches.

Table 7. Unique structural features and their diagnostic product ions characteristic of different CNP groups.

Group	Structural features	Product ions *m/z*
MC	Adda	135.0803, 163.1113
	Mdha	213.0870
CP	Lxx-Ahp	181.1331
	Phe-Ahp	215.1167
	Val-Ahp	167.1178
AS	Choi	140.1099, 122.0966
	Agma	281.1914
MG	Di, mono, non-Cl-Ahoa	168.0338, 134.0727, 100.1122
	Di, mono, non-Cl-*NMe*Ahoa	182.0494, 148.0883, 114.1278
	Di, mono, non-Cl-Ahda	196.0639, 162.1039, 128.1423
	Di, mono, non-Cl-*NMe*Ahda	210.0795, 176.1195, 142.1590
AP	D-Lys	84.0816
	*NMe*Ala	58.0658
	Arg-CO	201.0985
	Hty	150.09134
	*NMe*Hty	164.1070
	Hph	134.0962
	Trp	159.0914
CB	Tzl	86.0059
	Tzln	88.0226
	MeOx	84.0448
	Tzln-Lxx	199.0904
	Tzl-Lxx	198.0904
	Tzl-Ala	155.0270, 198.0579, 223.0898

2.4. RESULTS

117 individual CNPs were detected from the fifty-five bloom samples studied. Sixty-three previously reported CNPs were identified and fifty-four were designated as potentially new CNPs. APs and CPs were the most diverse groups detected with thirty-three and thirty-two variants,

respectively, followed by MGs with twenty-one, CBs with fifteen and ASs with nine (Table 8). MCs were the least diverse class with only seven congeners detected (Table 8). The specific number of group variants detected are summarized by watercourse in Table 8. Lake X is a privately-owned watercourse located south of Ottawa, the landowners wish to remain anonymous and is designated as such. Lake X contained 91% of CNPs variants detected (106) and the most MC congeners (6). Lac Heney contained the second highest diversity with 61% CNP variants (71). Bay of Quinte and Constance Lake had the second (5) and third (4) greatest number of MC congeners. Interestingly, the Bay of Quinte shared an equal amount of five MC, CB and CP variants – the only watercourse wherein MCs shared the greatest CNP group diversity.

Table 8. Number of unique CNPs detected for each group displayed by watercourse. CNP groups abbreviated as follows: AP; anabaenopeptin, AS; aeruginosin CB; cyanobactin, CP; cyanopeptolin, MC; microcystins, MG; microginin.

		CNP group						
Watercourse	**Samples**	**MC**	**AS**	**AP**	**CB**	**CP**	**MG**	**Total**
Lake X	19	6	8	32	13	28	19	**106**
Lac Heney	4	3	5	25	13	15	11	**71**
McNally's Bay	2	2	2	14	7	4	3	**32**
Lac Meech	6	2	1	9	2	9	4	**27**
Lac Breton	6	2	7	3	1	12	1	**26**
Bagot Long Lake	5	2	–	6	2	10	4	**24**
Bay of Quinte	1	5	3	3	5	5	–	**21**
Constance Lake	4	4	3	–	2	8	2	**19**
Rice Lake	1	3	2	5	4	1	2	**17**
C Shore Out	1	2	1	4	3	–	–	**10**
Stand Out	1	–	1	1	5	–	3	**10**
Lac Philippe	1	1	1	5	–	1	1	**9**
Mill Pond	1	2	2	1	–	–	–	**5**
Lake Nipissing	2	2	1	1	–	–	–	**4**
South Nation R.	1	–	–	–	–	3	–	**3**
Total	**55**	**7**	**9**	**33**	**15**	**32**	**21**	**117**

A semi- targeted method was used to quantify all detected variants for the CNP groups of AP, CP and MG. The optimized LC method used provided effective separation of CNPs and decent peak

resolution (Fig. 14). Select CPs and APs had the lowest LC resolution within a 3.0 to 3.5 min RT window (Fig. 14). Linear CNPs eluted in the 2.2 – 3.5 min RT window and the majority of cyclic CNPs eluted within the 2.5 – 4.5 min RT window. Cyclic CBs and more hydrophilic MCs had the greatest RTs times ≥ 5 min.

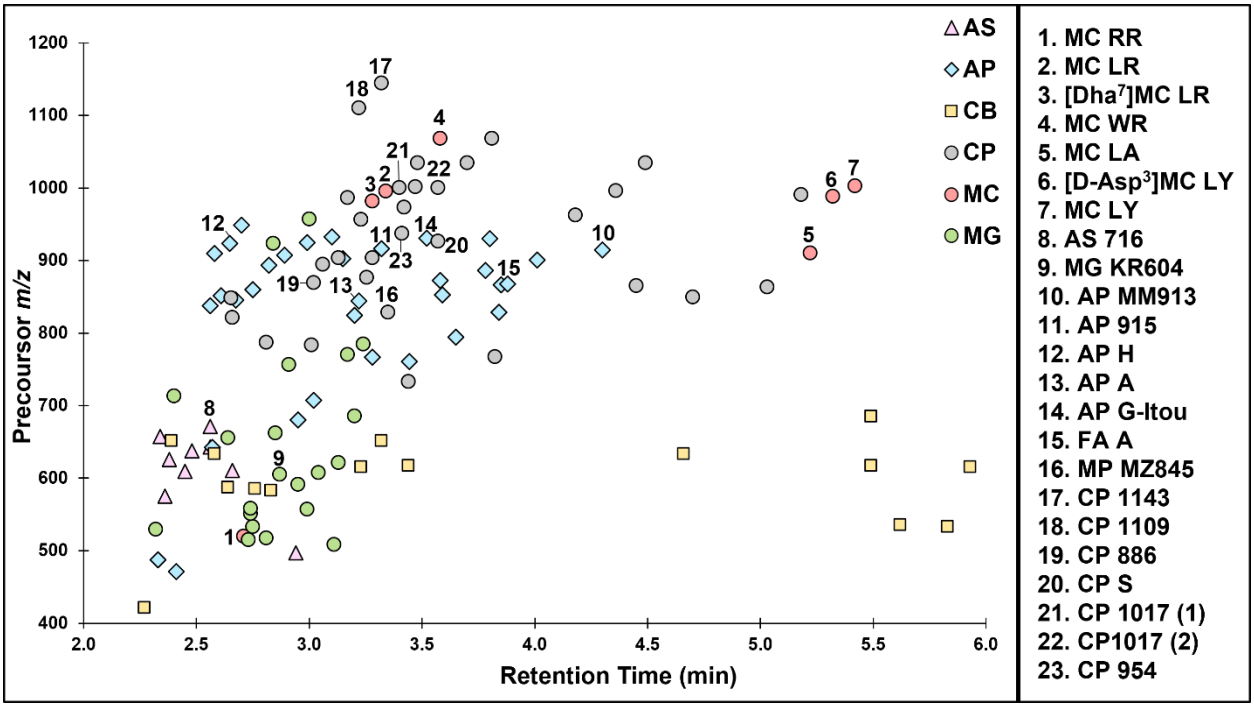

Figure 14. Retention times of the 117 CNPs identified plotted with the precursor ion *m/z*

1.2 GNPS Molecular Network

Seeded MNs are useful tools that supplement multiple metabolomic approaches, aiding in quick and effective identification of metabolites. The generated MN is shown in Figure 15. The resulting network for the fifty-five cAB samples contains 984 nodes and 2103 edges indicating structurally similar metabolites. The network displays a total of 103 clusters with more than two nodes (Fig. 14). Fourteen clusters and 117 nodes represent the detected CNP classes of MC – cluster #10; AS – clusters #1 and #2; AP – clusters #3 and #4; CP – cluster #8 and #9; MG – clusters #11, #12, #13 and #14; CB – clusters #5 ,#6 and #7. Each CNP cluster is annotated and outlined with black border in the network (Fig. 15). Individual CNPs were identified by (HR)MS/MS, retention

time, diagnostic product ions, and matches with library or database searches. Eighteen clusters were further identified as non-CNP compounds by the GNPS spectral library search, each is annotated with a dashed blue border. Seventy-one clusters and 442 nodes remain unidentified in the MN. The network was seeded with known MS/MS data for two previously studies *M. aeruginosa* strains [142]. The nodes corresponding to seeded spectra from *M. aeruginosa* CPCC 300 and CPCC 632 are displayed as “dark blue” and “light blue” nodes, respectively (Fig. 15).

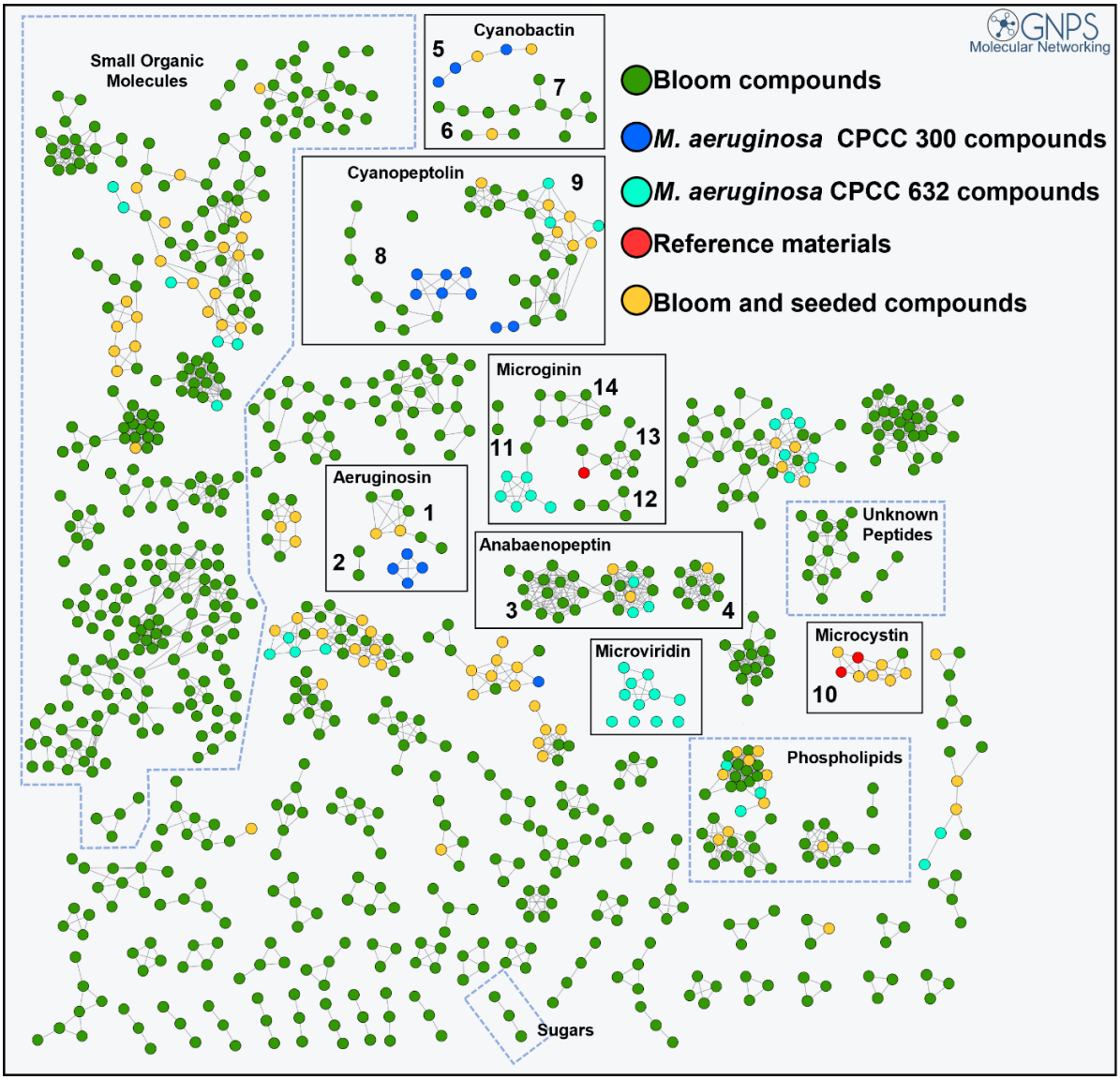

Figure 15. GNPS molecular network generated from metabolite (HR)MS/MS spectra from fifty-five cAB samples. Parameters used for network generation: precursor ion *m/z* tolerance; 0.03 Da, product ion *m/z* tolerance; 0.02 Da, minimum cosine similarity score; 0.6, minimum matching product ions; 6, minimum cluster size; 2. CNPs were targeted with the following LC-(HR)MS parameters: 450-1450 *m/z* range,

retention time window of 2.0-6.5 min. Deconvolution removed all metabolites present within the blank and redundant CNP feature adducts. The network was seeded with (HR)MS/MS spectra from *M. aeruginosa* CPCC 300 and CPCC 632 intracellular extracts and select reference standards. Nodes represent individual compounds, edges (gray lines) connect compounds which share similar product ions in their fragment spectra. CNP chemical groups are annotated with black boxes and labelled, groups specific clusters are numbered. Nodes shared by seeded spectra and the cABs are coloured gold. Clusters annotated with dashed blue lines indicate compounds identified by spectral library searches.

Nodes representing the seeded MS/MS spectra for the reference standards are coloured "red" (Fig. 15). Nodes corresponding to spectra from analyzed cABs are coloured "green". Any nodes overlapping with both the seeded and cAB spectra are coloured "gold" (Fig. 15). *M. aeruginosa* CPCC 300 is considered "toxic" as it produces MCs whereas, CPCC 632 is considered non-toxic as it does not produce MCs. *M. aeruginosa* CPCC 632 produces microviridins, chosen to seed the MN for effective identification of this metabolites of this group. MVs are large oligopeptides containing 12 to 20 amino acids, being difficult to identify using LC-(HR)MS/MS without available standards. The MN contains one single microviridin cluster from seeded MS/MS data of CPCC 632. No microviridins were detected in the fifty-five cAB samples.

2.4.1.1. Microcystins (MCs)

MCs are the most studied CNP group the date, being potent hepatotoxins and human carcinogens with well-documented toxicological information. The vast majority of MCs contain the non-proteinaceous beta amino acid Adda (3-Amino-9-methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid), and to a lesser extent the Mdha (methyl-dehydroalanine) [113]. The MS/MS spectra of MCs, contains diagnostic product ions which correspond to these diagnostic structural features. For example, *m/z* 135.08044 and *m/z* 163.11185 of the Adda moiety and *m/z* 213.08575 for Glu-Mdha backbone. Using metabolomic approaches such as MN and DFF these characteristic product ions are targeted to rapidly identify all MCs from samples [141,142]. The generated MN contained one MC cluster (#10). The MN was seeded with reference standards of eight MC congeners, MC LR, [Dha^7] -LR, -LA, -RR, -WR, -LY, -LF and -YR. Two MC congers from the seeded reference materials were not detected in the cAB samples, MC LF and -YR. Overall, nine

previously reported MC congeners were detected. [D-Asp^3]MC-LY was the only MC detected without an available reference standard. All seven of the detected MCs within the MN have been previously reported, MC LR, [Dha^7]MC LR, MC LA, MC RR, MC WR, MC LY and [D-Asp^3]MC LY. The identity of all MCs, excluding [D-Asp^3]MC LY were confirmed by LC-(HR)MS/MS of certified reference materials (NRC, Halifax). For determining the identity of the [D-Asp^3]MC LY (*m/z* 988.5023 $[M+H]^+$), it was 14.0160 mass units less than MC-LY, suggesting a demethylation of Asp at position 3. When comparing the MS/MS of MC LY and [D-Asp^3]MC LY, presence of the product ion *m/z* 392.2186 $[Tyr\text{-}Asp\text{-}Leu+H]^+$ suggests Asp in position 3 for this congener. Two seeded MCs form certified reference materials remain in the MN and were not detected in the samples, MC LF and MC YR.

2.4.1.2. Cyanopeptolins (CPs)

CPs are a CNP group of cyclic oligopeptides containing hexacyclic ring and an exo side chain of variable length at positions S_1 and S_2 (Fig. 16). CPs are structurally classified by the presence of a 3-amino-6-hydroxy-2-piperidone (Ahp) residue at position 3 of their hexacyclic ring (Fig. 16) [54]. Position 1 linking the exo side chain is commonly threonine residue [126]. Overall, CPs were the second most diverse class of CNPs with thirty-two congeners detected. The cABs containing the greatest diversity of CPs expressed by the watercourses of, Lake X, Lac Heney, Lac Breton and Bagot Long Lake. The shared Ahp moiety is ubiquitous for the class resulting in shared common product ions between congeners. The amino acid adjacent to the diagnostic Ahp group, position 4 (Fig. 16) subclassifies cyanopeptolin congeners [142]. This substitution remains fairly consistent within the group, making subclassification advantageous in characterizing the vast structural diversity for CPs [52,54,141]. Determination of shared partial amino acid sequences with diagnostic product ions provides further structural subclassification. For example, CP 880 is a Lxx-Ahp containing cyanopeptolin because it contains a Ile (position 4) adjacent to the diagnostic Ahp moiety (position 3) and results in the following diagnostic product ions; *m/z* 209.1283 [Ile-Ahp–

$H_2O+H]^+$ and *m/z* 181.1331 [Ile-Ahp–H_2O–CO+H]$^+$ (Fig. 16) [142]. Moreover, NP BN920 is a Phe-Ahp containing CP

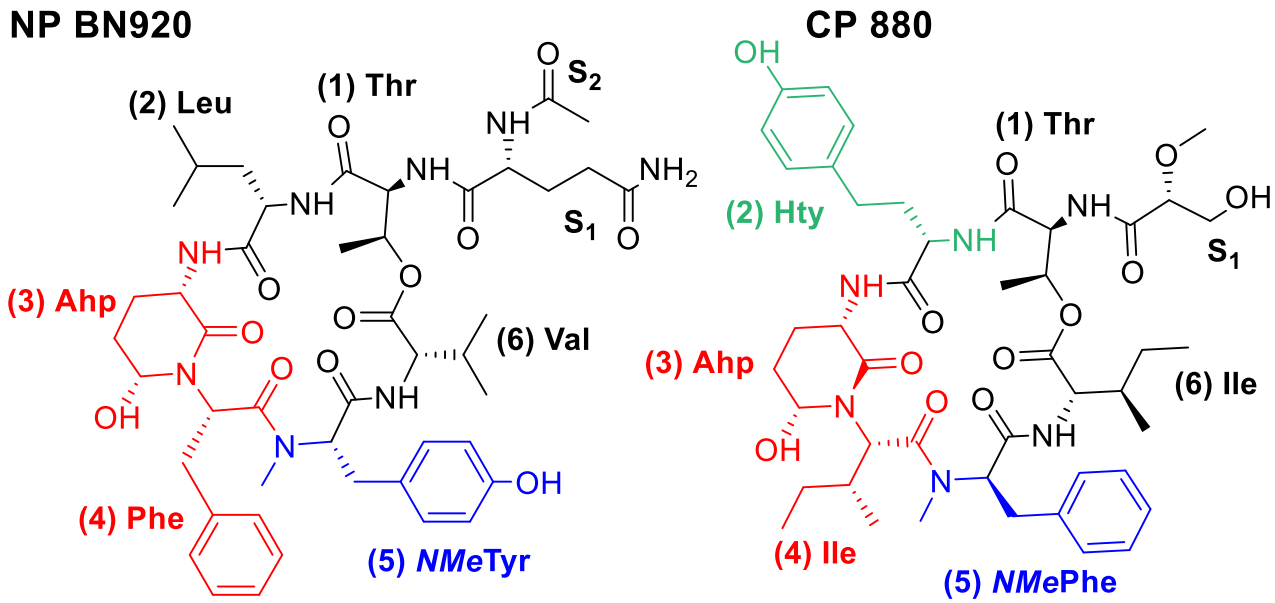

Figure 16. The chemical structures of nostopeptin BN920 (left) and cyanopeptolin 880 (right) representative of each subclass. Ahp-Phe and Ahp-Lxx and the *NMe*-aromatic position 3 substitution.

having a Phe residue at position 4 adjacent to the Ahp moiety and results in diagnostic product ions of *m/z* 243.1121 [Phe-Ahp – H_2O+H]$^+$ and *m/z* 215.1174 [Phe-Ahp–H_2O–CO+H]$^+$ (Fig. 16). Greater structural and sequence information commonly arises with the presence of the *NMe*-aromatic amino acid immonium ion at position 5 such as, *NMe*Phe (*m/z* 134.0962), *NMe*Tyr (*m/z* 150.0910), and *NMe*Tyr-Cl (*m/z* 184.0518; Fig. 16). The generated MN contained two CP clusters # 8 and # 9, with eight and twenty-two nodes, respectively (Fig. 17). The MN was seeded with seventeen CPs produced by *M. aeruginosa* CPCC 300 and 632. Six of the seeded CPs were detected in the bloom samples. Overall, thirty-two CP congeners were detected, fourteen previously reported and eighteen new CPs were detected (e.g., CP 866, CP 882, CP 875, CP 955, CP 1000, CP 1051 (C), CP 1051 (B), CP 1085, CP 1109, CP 784, CP 805, CP 838, CP 886, CP 981, CP 990, CP 1003, CP 1007, CP 847). The subclassification for congeners of this group is visualized as the two CP clusters of the MN. CP 847 was the only cyanopeptolin that did not

cluster and remained as a single node (Fig. 17). CP 847 contained the partial sequence NMeTyr-Val-Ahp a combination not shared by CPs of either cluster. Cluster # 8 contains Lxx-Ahp and Val-Ahp CPs associated with an *NMe*Phe or *NMe*Hty within their amino acid sequence.

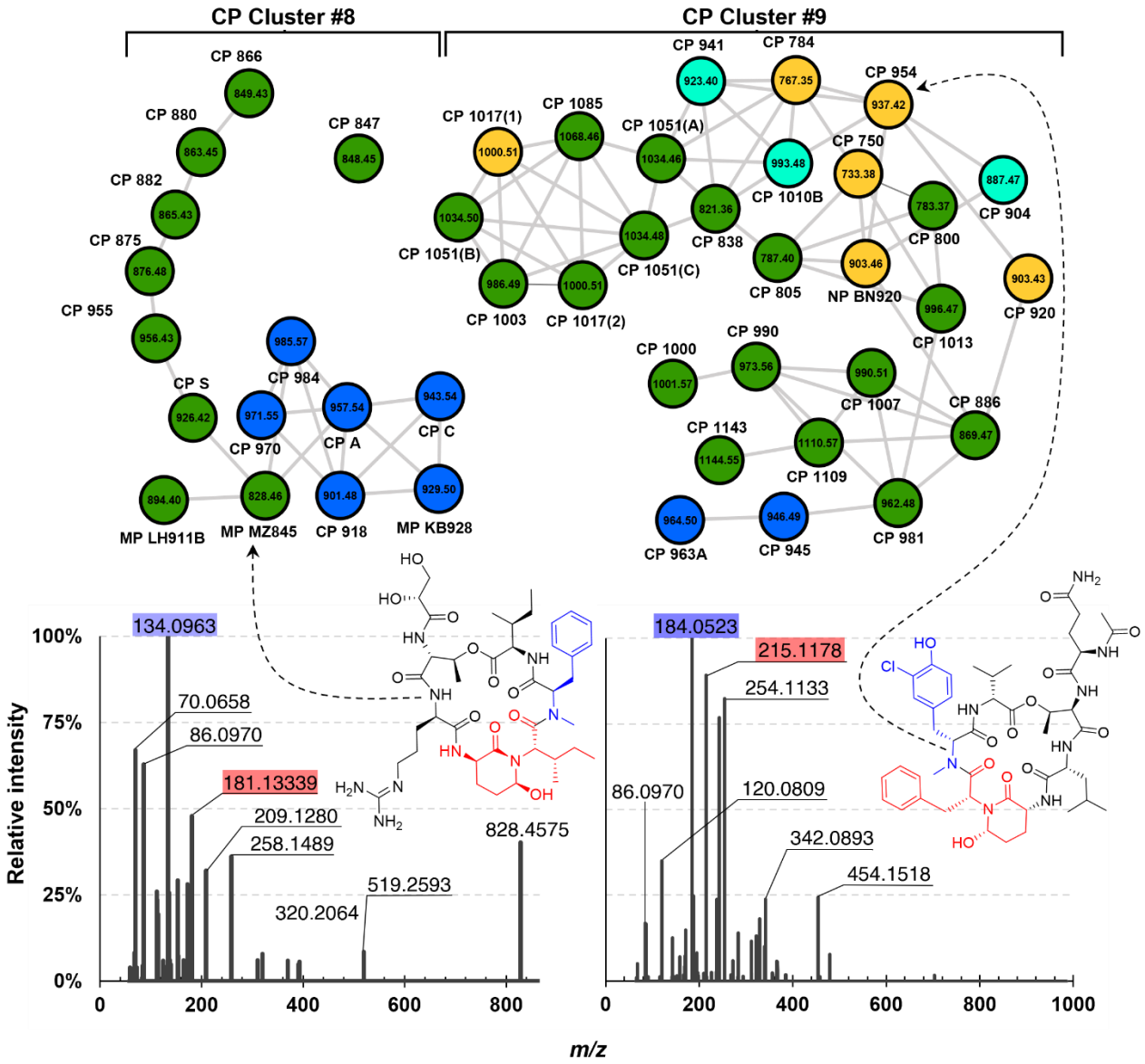

Figure 17. (Top) GNPS molecular network clusters of the cyanopeptolin CNP group. Cluster nodes are annotated with each congeners m/z (inside) and chemical name (outside). (Bottom) The chemical structures and MS/MS spectra of micropeptin MZ845 (MP MZ845; left) and cyanopeptolin 954 (CP 954; right), examples of each subclass and separate GNPS cluster. Diagnostic product ions are indicated MS/MS spectra. For Lxx-Ahp-*NMe*Phe or -Hty cyanopeptolins *m/z* 134.0961 (*NMe*Phe; blue), *m/z*

181.1331 ([Ahp-Lxx–CO–H2O]$^+$; red) and for Lxx- or Phe-Ahp- cyanopeptolins with either a *NMe*Tyr or *NMe*Tyr-Cl, the *m/z* 184.0523 (*NMe*Tyr-Cl; blue), *m/z* 215.1178 ([Ahp-Phe–CO–H2O]$^+$; red) are present.

Cluster # 8 CPs were confirmed by the diagnostic product ions of *m/z* 181.1331 [Ahp-Lxx–CO–H2O]$^+$, *m/z* 167.1178 [Ahp-Val–CO–H2O]$^+$, *m/z* 134.0961 (*NMe*Phe), and *m/z* 164.1069 (Hty). Cluster #9 contains Phe-Ahp and Lxx-Ahp containing CPs associated with an *NMe*Tyr or *NMe*Tyr-Cl within their amino acid sequence. In addition, CPs containing Hty are only found in cluster #8. Cluster #9 CPs were confirmed by the diagnostic product ions of *m/z* 181.1331 [Ahp-Lxx–CO–H2O]$^+$, *m/z* 215.1167 [Ahp-Phe–CO–H2O]$^+$, *m/z* 184.0517 (*NMe*Tyr-Cl) and *m/z* 150.0912 (*NMe*Tyr). For example, MP MZ845 contains the partial sequence Lxx-Ahp-*NMe*Phe and CP 954 contains Phe-Ahp-*NMe*Tyr-Cl, representative of their separate cluster groupings within the MN (Fig. 17). The substitution of the amino acids at CP positions 2, 3, 4 and 5 and the exo-S1 and -S2 side chain results in sufficient product ion diversity to generate separate network clusters for subclassification of cyanopeptolins. Furthermore, structural subclassification of CPs resulted in three distinct groups of CPs within the MN.

2.4.1.3. Anabaenopeptins (AP)

Anabaenopeptins (APs) are a group of hexapeptide CNPs, with a heptacyclic core comprised by five amino acid residues and one in an exo position (position 1) [126]. This group of CNPs is of interest as its congeners demonstrate a similar mechanism of toxicity to that of MCs, inhibition of protein phosphatases [126,129]. APs share a ureido linkage for cyclization by a D-lysine residue at position 2. This linkage is shared between all congeners within the group, connecting the exo residue at position 1 to the residue at position 3 by a carboxy group [127]. Other residue positions, apart from the lysine, are variable contributing to the groups' large structurally diversity [127]. Overall, APs were the most diverse group of CNPs with thirty-three detected variants. Furthermore, APs were detected in all watercourses studied. The greatest number of variants was observed in Lake X, Lac Heney, McNally's Bay, Lac Meech, Lac Breton, and Bagot Long Lake. The dominant common product ions observed in MS/MS spectra of APs are *m/z* 84.0812

and *m/z* 129.1018, characteristic of the lysine residue. The generated MN contained two AP clusters #3 and #4, with thirty-three and ten nodes (Fig. 18). The MN was seeded with five APs produced by *M. aeruginosa* CPCC 632 and the certified reference material FA A. The three of the seeded APs

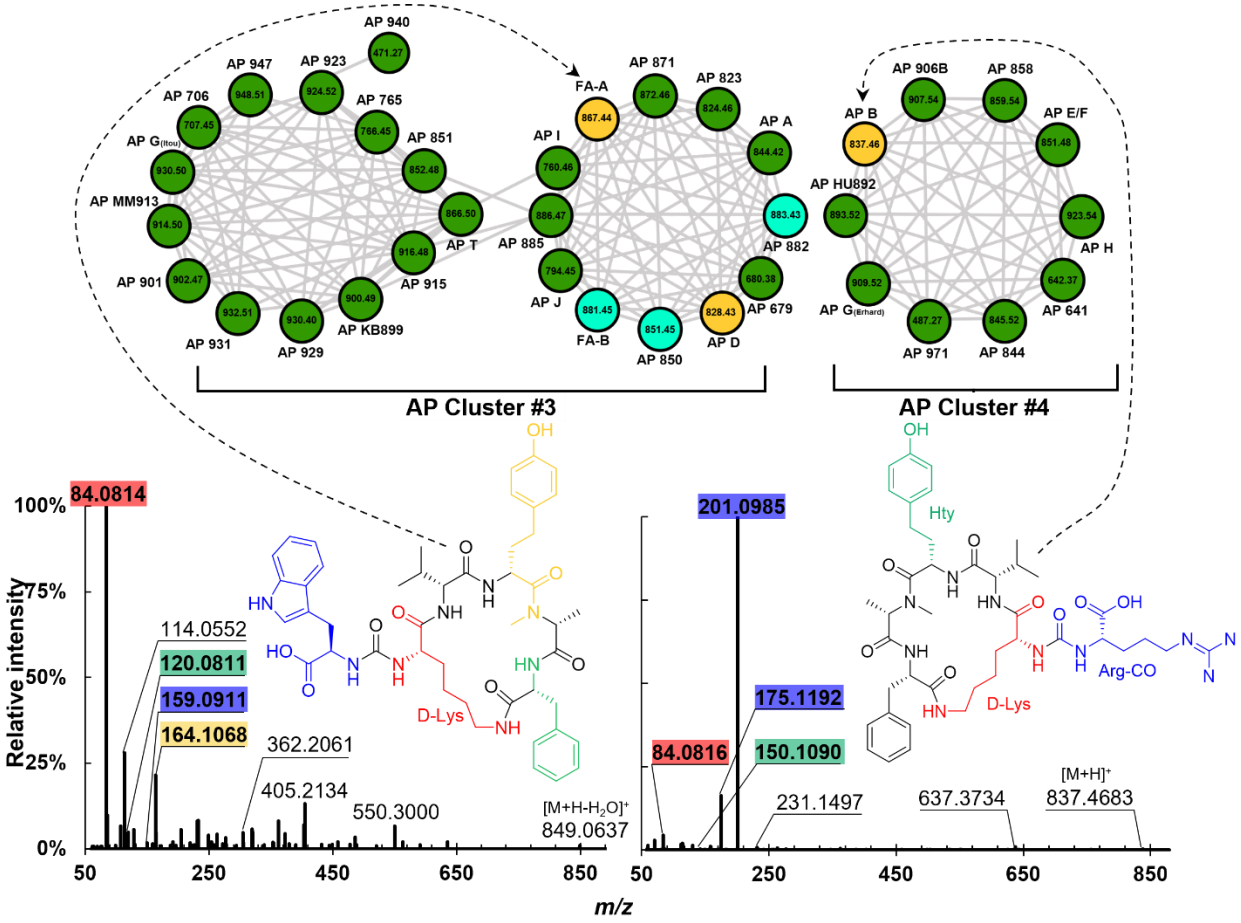

Figure 18. (Top) GNPS molecular network clusters of the anabaenopeptin CNP group. Cluster nodes are annotated with each congeners *m/z* (inside) and chemical name (outside). (Bottom) The MS/MS spectra and chemical structures of ferintoic acid A (FA A; left) and anabaenopeptin B (AP B; right). Diagnostic product ions and their putative structures are annotated by colour: *m/z* 84.0816 (Lys; red), *m/z* 120.08078 (Phe; green), *m/z* 150.0913 (Hty; green), *m/z* 164.1070 (*NMe*Hty; yellow), *m/z* 201.0985 (Arg-CO; blue), *m/z* 175.1192 (Arg + 2H; blue) and *m/z* 159.0914 (Trp; blue).

(i.e., FA A, AP B and AP D) were detected in cABs. Three additional seeded APs, FA B, AP 850 and AP 882 from CPCC 632 were not detected in cABs. A total of thirty-three AP congeners were identified by the MN, eighteen congeners were previously reported. The identity of reported AP congeners was verified by diagnostic product ions within their MS/MS spectrum and mass errors of the precursor ion. For example, immonium ions confirming the identity of AP D are *m/z* 58.0658

(*NMe*Ala), *m/z* 72.0813 (Val), *m/z* 84.0812 (Lys), *m/z* 120.0808 (Phe), and *m/z* 150.0909 (Hty); representative of the [Arg_1-D-Lys_2-Val_3-Hty_4-*NMe*Ala_5-Phe_6]. Fifteen AP congeners within the network clusters could not be matched to a previously identified metabolites through library or database searches and are considered potentially new APs. Eleven potentially new APs are within cluster #3 and, four in cluster #4. The two separate AP clusters in the MN are representative to the structural variation at the exo position, specifically the presence or absence of the Arg-CO at this position (Fig. 18). APs within cluster #3 contained either a Lxx (*m/z* 86.0969), Phe, Hty or Trp at the exo position, confirmed by the corresponding diagnostic product ions, *m/z* 159.0914 (Trp) and *m/z* 164.1069 (HTyr; Fig. 18). APs within cluster #3 possess an Arg-CO at the exo position, confirmed by the product ions of *m/z* 201.0985 $[Arg\text{-}CO+H]^+$ and *m/z* 175.1192 $[Arg+2H]^+$ (Fig. 18).

2.4.1.4. Microginins(MG)

MGs are linear lipopeptides comprised of four to six amino acids with an N-acyl terminus containing a diagnostic octanoic or dodecanoic acid moiety [126]. The groups characteristic N-terminal 3-amino-2-hydroxydecanoic acid (Ahda) and 3-amino-2-hydroxyoctanoic acid (Ahoa) residues result from a PKS pathway [126]. The C-terminal end of the Ahda and Ahoa polyketide moieties can be tri-, di, or mon-chlorinated and the N-terminal can be methylated adding greater structural complexity to the group [142]. Characteristic product ions are generated from the fragmentation of the MG polyketide moieties. Changes in chlorination states of the diagnostic fragment ions are observed as a mass jump of 33.9606 units, associated with the exchange of a chlorine for a hydrogen atom. The Ahoa product ions are, *m/z* 100.1122, *m/z* 134.0727, *m/z* 168.0338, indicative of the non-, mono-, di-, and tri-chlorinated states, respectively [142,208]. Product ions of Ahda residue include *m/z* 128.1433, *m/z* 162.1039 and *m/z* 196.0645 [142,208]. N-methylation of the Ahoa and Ahda restudies is observed as a mass increase of 14.0151 units. Many MGs contain the *NMe*Lxx residue, the product ion of which is isobaric to that of the Ahoa residue (*m/z*

100.1122). This indistinguishable issue is a limitation to structural elucidation using (HR)MS/MS for the group. The generated MN contained five MG clusters #11, #12, #13 and #14, with two, four, seven, and eight nodes, respectively. The MN was seeded with five MGs produced by *M. aeruginosa* CPCC 632 and the certified reference material of MG FR1. None of the seeded MGs overlapped with cAB nodes. In total, twenty-one MG congeners were detected and fourteen were previously reported such as, MG 507, MG 557, MG 514, MG 516, MG 528, MG 532, MG 661, MG 755, MG 551, MG KR604, MG 606A, MG620A/B and MG 684. The Ahoa (3-amino-2-hydroxy octanoic acid) and Ahda (3-amino-2-hydroxy decanoic acid) residues characteristic of MGs produced similar diagnostic product ions within their MS/MS spectra. However, the structural substitutions of the polyketide moiety were substantial to generate four sperate clusters. Cluster #11 contained MG 507 and MG 557 which possessed the *NMe*Ahda moiety confirmed by the corresponding diagnostic product ion of *m/z* 142.1589 (*NMe*Ahda). Cluster #12 contains four MGs (e.g., MG 712, MG 755, MG 922, MG 956) associated with an *NMe*Tyr (150. 0912) or an *NMe*Tyr-Cl (*m/z* 184.0517) residue in their MS/MS. MG 755 contained the Ahoa structural moiety confirmed by *m/z* 100.1122. However, the MS/MS spectra for MG 712, MG 922 cand MG 956 lacked diagnostic product ions corresponding to the terminal polyketide moiety. Three MGs in cluster #12 lacked the Ahoa or Ahda structural moieties of the group possibly suggesting an alternate N-terminal residue for these MGs. It's postulated that MG 922 and 956 possessed the Ahda moiety confirmed by *m/z* 168.1383 ($[Ahda+COC+H]^+$), although this could not be validated with additional product ions. Cluster #13 contains seven MGs with the Ahoa and Ahda-Cl moiety associated with *NMe*Lxx and Lxx residues (e.g., MG 514, MG 516, MG 528, MG 528, MG 532, MG 556, MG 590 and MG 661; Fig. 19) The seeded node of MG FR1 was contained in this cluster which possessed the amino acid sequence [Ahda-Ala-N-Me-Leu-Tyr-Tyr]. The cluster #13 groupings is indicative to the isobaric nature of the Ahoa and *NMe*Lxx product ions. MGs of this cluster were confirmed by the product ions of *m/z* 100.1122 (Ahoa), *m/z* 100.1122 (*NMe*Lxx), *m/z* 162.1039 (Ahda-Cl) and *m/z* 86.0964 (Lxx). Cluster #14 contained eight MGs which contained

the *NMe*Ahda moiety and MG 551 which contained the Ahda moiety confirmed by the diagnostic product ion of *m/z* 128.1423 (Ahda) and *m/z* 142.1590 (*NMe*Ahda; Fig 19). The remaining peptide sequence was variable for Cluster #14 MGs containing product ions for Val (*m/z* 72.0813), *NMe*Ala (*m/z* 58.0658) Pro (*m/z* 70.0651), *NMe*Lxx and *NMe*Tyr.

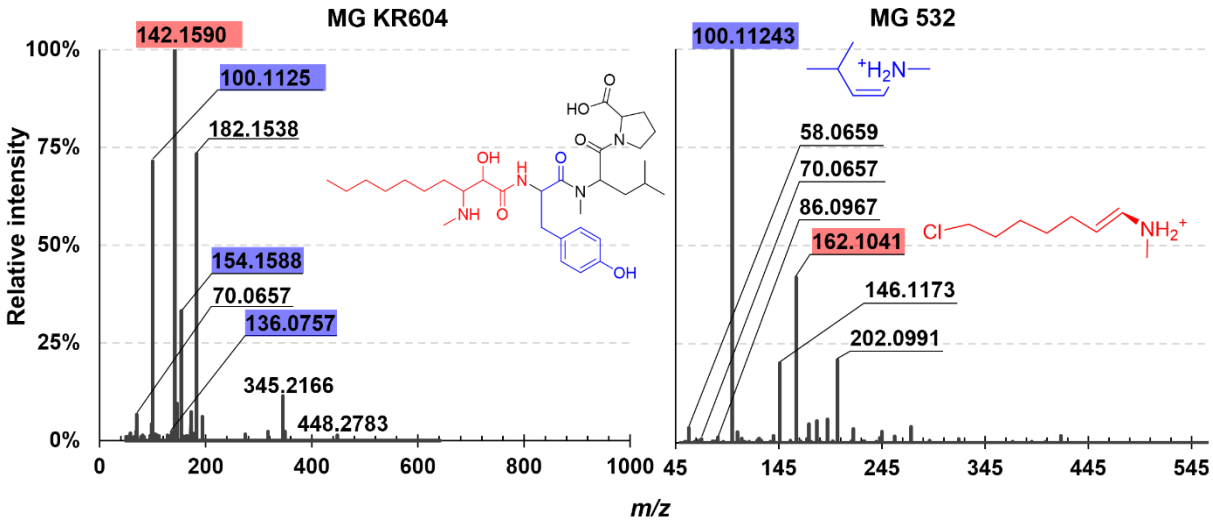

Figure 19. Chemical structure of MG KR604 and MS/MS spectra of microginin KR604 (left) and MS/MS spectra of MG 532 (right). Diagnostic product ion structures are shown for, *m/z* 100.1122 (*NMe*Lxx) and *m/z*, 162.1039 (Ahda-Cl). Product ions are displayed for *m/z* 142.1590 (*NMe*Ahda), *m/z* 70.0658 (Pro), *m/z* 86.0969 (Lxx), *m/z* 58.0659 (*NMe*Ala), *m/z* 136.0752 (Tyr), *m/z* 154.1588 ([Tyr-H2O+H]+) confirmatory of MG structures.

Previously reported MGs were identified by spectral library databases searches (GNPS and CyanoMetDB spectral libraries) and confirmed with diagnostic product ions. For identification of potentially new MGs DFF was used to compliment spectral library searches and diagnostic product ions. Two Microginins, MG FR12 and MG FR13 in cluster #14 were identified from unpublished work using the CyanoMetDB library [53]. Five potentially new Microginins were identified based on diagnostic product ions, MG 590, MG 712, MG 922, MG 956 and MG 654.

2.4.1.5. Cyanobactins (CB)

Cyanobactins are a group of highly modified CNPs that contain both proteinaceous amino acids and heterocyclic residues [70]. In CB RiPP synthesis, precursor peptides are recognized by a single heterocyclase enzyme responsible for addition of azolines (commonly Tzln (thiozoline) and

MeOxln (methoxyzoline) in the core peptide sequence [74]. Azolines residues form a bond between the amino acid side chain heteroatom (oxygen or sulfur atom) from Thr, Ser, and Cys and the preceding amide bond. Azoline residues can be further oxidized to Azoles (i.e., Tzln (thiozole) and MeOxln (methoxazole) which occur in majority of the reported cyanobactins, increasing structural diversity of this chemical group [70]. Further diversity occurs with substitution of proteinaceous amino acids surrounding the Azol(in)e residues. These metabolites may be cyclized or linear peptides cyanobactins (cyclamides) and linear cyanobactins [70,74]. The potential for oxidation/reduction of the heterocyclized resudues, amino acid substitutions, and linear/cyclic cyanobactins heightens structural diversity and translates to dissimilarities in diagnostic product ion patterns. The generated MN contained three CB clusters #5, #6 and #7 with two, ten and three nodes, respectively. The MN was seeded with five CBs produced by *M. aeruginosa* CPCC 300. Two of the seeded MGs, aerucyclamides (AC) A and B were detected in cABs and represented by nodes in cluster # 5. In total, fifteen CB variants were detected, ten previously reported. Cluster #5 contains AC A and AC B, the former has the partial sequence Tzln-Lxx which is confirmed with the product ions *m/z* 86.0064 (Tzln), *m/z* 154.0681 ([Tzln-Lxx–$CONH_2$+H]$^+$) and *m/z* 197.0741 (Lxx-Tzln; Fig. 20). AC B and has the partial sequence Tzl-Ile which is confirmed with the product ion *m/z* 199.0741 related to the oxidized residue of the Tzln product ion seen in the spectra for AC A (Fig. 20). The oxidized congener, AC B contains the Tzl (*m/z* 88.005). The seeded CB nodes of AC C, CB 518 and CB 550 were found in cluster # 5. Cluster # 7 contains one previously reported CB, microcyclamide (Md) GL616 identified by the presence of [Tzl-Ala]$^+$ (*m/z* 198.0580), [Tzl-Ala-CO+H]$^+$ (*m/z* 155.0272), [Lxx-Tzl-Ala]$^+$ (*m/z* 223.0898), Phe (*m/z* 120.0808), Thr (*m/z* 74.0600) and [M+H]$^+$ (*m/z* 617.2747) product ions in MS/MS spectra (Fig. 20). The remaining nine CBs in cluster # 7 could not be matched to previously identified metabolites by library or database searches and are considered potentially new variants. Partial amino acid sequences could not be elucidated for these compounds.

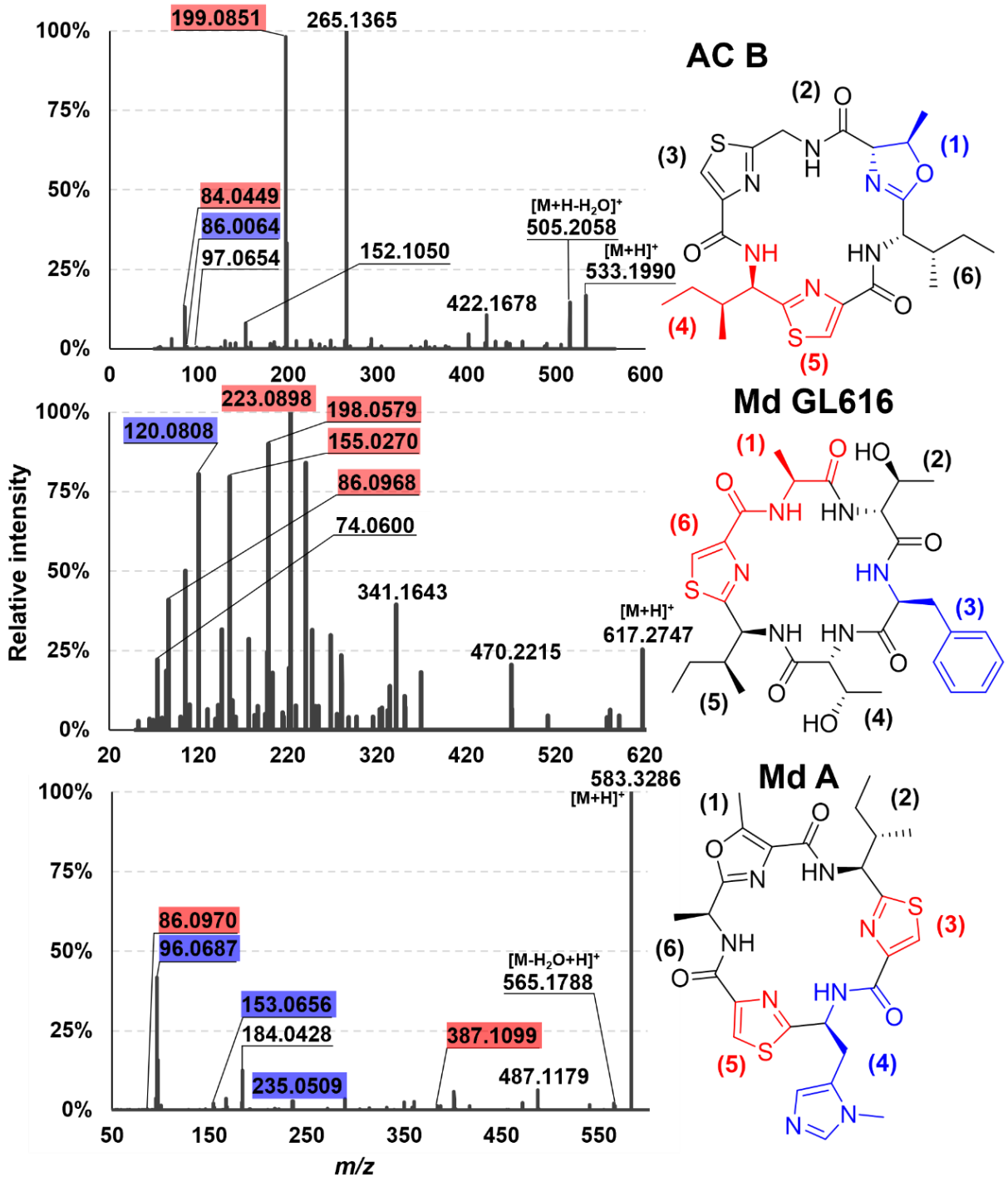

Figure 20. Chemical structures and MS/MS spectra of aerucyclamide (AC) B (cluster #5), Microcyclamide (Md) GL616 (cluster #7) and Md A (cluster #6). Each cyanobactin was confirmed by precursor ion *m/z* and diagnostic product ions from the MS/MS: AC B, *m/z* 84.0449 (MeOxl) *m/z* 86.0969 (Lxx), and *m/z* 199.0904 [Tzln-Lxx]$^+$; Md GL 616, *m/z* 120.0808 (Phe), 74.0600 (Thr), 155.0270 ([Tzl-Ala–CO+H]$^+$,

198.0579 [Tzl-Ala]$^+$, 223.0898 [Lxx-Tzl-Ala]$^+$; Md A, (MeOxl), *m/z* 86.0064 (Tzln), *m/z* 96.0687 (MeHis), *m/z* 153.0656 (MeOxl-Ala), 387.1099 [Ile-Tzl-MeHis-Tzl]$^+$

Cluster #6 contains three previously identified CBs, Md A, and its reduced congeners Md 584 and Md 586 [142,163]. Md A has the following hexapeptide sequence [MeOxl-Ile-Tzl-MeHis-Tzl-Ala] (Fig. 20). The molecule is highly oxidized containing multiple reduction sites, which differentiates the three congeners within the cluster. Cluster #6 nodes share the product ion for MeHis at *m/z* 96.0687 and MeOxl-Ala at *m/z* 153.0656 confirming the common core structure between these molecules. Md 584 and Md A contain the product ion for Ile-Tzl-MeHis-Tzl (*m/z* 387.1099). The position of the oxidized Tzl group in Md 584 is indicated by the product ion sequence MeHis-Tzl (*m/z* 235.0509). MeHis-Tzl is apparent in both MS/MS spectra of Md A and Md 584 whereas, Md 586 contains MeHis-Tzln (*m/z* 237.0660) implying the Tzl residue at position 3 is oxidized to Tzln (Fig. 20). Md 586 contains MeHis-Tzln and MeHis-Tzl (*m/z* 389.1231) implicating both Tzl residues are oxidized to Tzln. The cyclic and linear nature of these peptides associated heterocycles, and multiple oxidation/reduction sites is a surmounting challenging when identifying CB variants in natural product extracts.

2.4.1.6. Aeruginosins(AS)

Aeruginosins (ASs) are small linear tetrapeptides that contain a characteristic 2-carboxy-6-hydroxyoctahydroindole (Choi) and a C-terminal Arg or Arg-derivative [126]. The N-terminus of ASs contains a hydroxyphenyl lactic acid (Hpla) group that can vary in substitutions along the aromatic ring, commonly halogenation in the meta positions of the ring [52]. Further, variation arises with, O-R modifications of the Choi moiety such (i.e., sulfate or hydroxy groups), substitutions of the C-terminal Arg-derivative and amino acid substitutions within position 2 [52,126]. The MS/MS spectra of AS metabolites contain diagnostic product ions generated from the characteristic Choi moiety. Two product ions related to the Choi moiety are observed at *m/z* 140.1064 and *m/z* 122.0962, the latter is 18.0102 mass units less indicating a dehydrated product ion [52].The generated MN contained two AS clusters #1 and #2 with seven and two nodes, respectively. The MN was seeded

with nine ASs variants produced by *M. aeruginosa* CPCC 300. A total of nine ASs were detected from bloom samples, two were previously reported (e.g., AS 688 and AS 722) the remaining seven were identified as potentially new ASs. All congeners were confirmed for the group by the presence of product ions characteristic of the Choi moiety, *m/z* 140.1066 (Choi), *m/z* 122.0964 $[Choi\text{-}H_2O]^+$. Cluster #1 contained ASs which processed a N-terminal Hpla-Phe group confirmed by the presence of a Phe immonum ion and the diagnostic product ion *m/z* 284.1268 $[Hpla\text{-}Phe\text{-}CO+H]^+$. At the C-terminal, cluster #1 ASs contained an agmatine (Agma) residue confirmed by diagnostic product ions, *m/z* 281.1914 (Agma), *m/z* 314.2199 $[OH\text{-}Choi\text{-}Agma]^+$. New AS 576 was an exception as no product ions for a C-terminus residue (i.e., Agm, argininol, argininal) were observed, suggesting it does not possess a C-terminal residue. AS 610 was the only non-sulfated AS detected. Sulfated are confirmed by a neutral loss of 79.9571 *m/z* from the precursor ion corresponding to the loss of an SO_4 group. Cluster #2 contained sulfated AS 716 and AS 750 both possessing an Argal residue at their C-terminus confirmed by the diagnostic product ions, *m/z* 266.1863 $[Choi\text{-}Argininal\text{-}H_2O+H]^+$, *m/z* 291.1817 $[Choi\text{-}Argininal\text{-}NH_2\text{-}H_2O+H]^+$ and *m/z* 100.0760 (Argal fragment). AS 716 contained an Hpla-Phe at the N-terminus whereas, AS 750 contained a chlorinated Hpla-Phe confirmed by $[Cl\text{-}Hpla\text{-}Phe\text{-}CO+H]^+$ ion.

2.4.2. Multivariate Analysis

Untargeted metabolomics approaches used in this study involved the application of non-parametric statistics, such as principal component analysis (PCA) and factor loadings plot. These statistical analyses can interpret and visually portray the metabolome variations of cAB samples. In combination with GNPS molecular networking, non-parametric analysis are synergistic tools for describing the CNP profiles of cABs. The objective of the metabolomic analyses was to identify the most common cyanopeptides from bloom samples which can be prioritized for subsequent studies: hazard characterizations and exposure assessments. The acquisition of large high-resolution DDA (HR)MS/MS data sets, combined with metabolomics analyses are highly useful

for visualization and interpretation of CNP profiles [142,143]. PCA aims to describe the variation between complex samples using a collection of multiple variables that define each sample. To understand the CNP metabolome variations of the fifty-five cAB samples, ninety-five of the 117 CNP peaks detected with the *R(v4.1.3) xcms* package were considered for PCA. The pareto scaled $\log_{10}$ peak areas of the selected nighty-five CNP features were analyzed with the packages *MetabolAnalyze and FactoMineR* in *R(v4.1.3)* to generate a PCA and a factor loading plot (Fig. 21-1). The factor loading plot is complementary to the PCA, providing visual insight into which CNP variables contribute to the variance and resemblance of the metabolomes visualized in the PCA. The CNP groups and individual compounds that differentiate each bloom and watercourse were elucidated through investigation of the factor loadings plot. The PCA plot displayed as dimensions one (60.2%) and two (7.6%) which describe the majority of variation comparatively, for all dimensions (summed total variance = 70.1%). The use of alternate PCA dimensions was explored but none provided a greater explanation of the overall variance. The majority of blooms sampled share similar CNP profiles, visualized in the PCA as a large group of blooms overlapping along the x-axis, left of the y-axis. The CNP profiles of cABs within this grouping possess MCs and variable CNPs from the groups of CB, CP, AP and MG. However, many of these profiles contain a low diversity in CB, CP and AP variants. Lake X blooms (blue) occupy 3 of 4 quadrants of the PCA, possessing the majority of CNP variation in both diversity and abundance. Two CNP profiles from McNally's Bay (Upper Rideau Lake, ON; violet) and Lac Heney (QC; orange) are adjacent to several Lake X samples, left and right of the y-axis, respectively. These two samples share similar CNP diversity and concentrations to that of Lake X. Interestingly, three Lac Breton blooms (green) appear to be rather unique in their CNP profiles grouping separately from the remainder of samples, towards the top-left of the PCA. The corresponding loadings plot (Fig. 21-2) visualizes each of the ninety-five CNP features considered by the PCA. The coordinate locations of individual CNP vectors within the loadings plot is determined by both its incidence and abundance of a CNP relative to all profiles. CNPs are annotated within the feature loadings

plot to highlight the most abundant CNPs detected. The vast majority of CNPs contributing to the variation within Lake X profiles are associated with the groups of APs, MGs and CBs. Lake X cAB profiles situated below the y-axis and further right comprise the greatest diversity of APs, CBs and MGs. The seven Lake X samples are situated above the y-axis differ minimally to the other Lake X CNP profiles but do contain a greater abundance of MCs, MGs and CPs (Fig. 21-1). The single McNally's Bay and Lac Heney samples situated near Lake X in the PCA contain similar profiles of APs and MCs; however, they lack MG and CP diversity (Fig. 21-2). The three Lac Breton blooms grouping to the upper left of the PCA are unique, resulting from of a lack of APs and the presence of CP 1143 and CP 1109 within their profiles (Fig. 21-1, -2). Similar to the more diverse cABs, Lac Breton contained MCs and CP 887. Overall, the Lac Breton booms had minimal diversity in MG and AP variants conversely to, Lake X, McNally's Bay and Lac Heney profiles. Further investigation of specific CNP features contributing to the profile variance between watercourses was conducted using the *dplyr* package in *R*. A KW test (non-parametric) with Benjamini-Hochberg correction was performed for each CNP feature considered for PCA (nighty-five). The test determined which metabolites were statistically significant in contributing to the cAB profile variation between the fifteen watercourses ($p < 0.05$). The loadings plot contains the CNP features indicated by significance (significant = red; non-significant = black; Fig. 21-3). Forty-four CNPs were deemed statistically significant and fifty-three non-significant in contribution to the overall watercourse profile variance. A goal of this study is to determine which CNPs are most ubiquitous throughout the fifty-five bloom samples. For this reason, non-significant CNPs in contribution to the watercourse-to-watercourse variation, were of further interest. Similar to most studies, MCs were the most common group across watercourse profiles with 6 (87%) non-significant congeners. MC RR was the only significant MC and predominantly found in Lake X. MC WR was not considered in the PCA as it did not meet the pre-set peak pick parameters. The CP group consisted of the second most non-significant variants, 66% (21/32). MGs and ASs had 39% (8/21) and 44% (4/9), respectively.

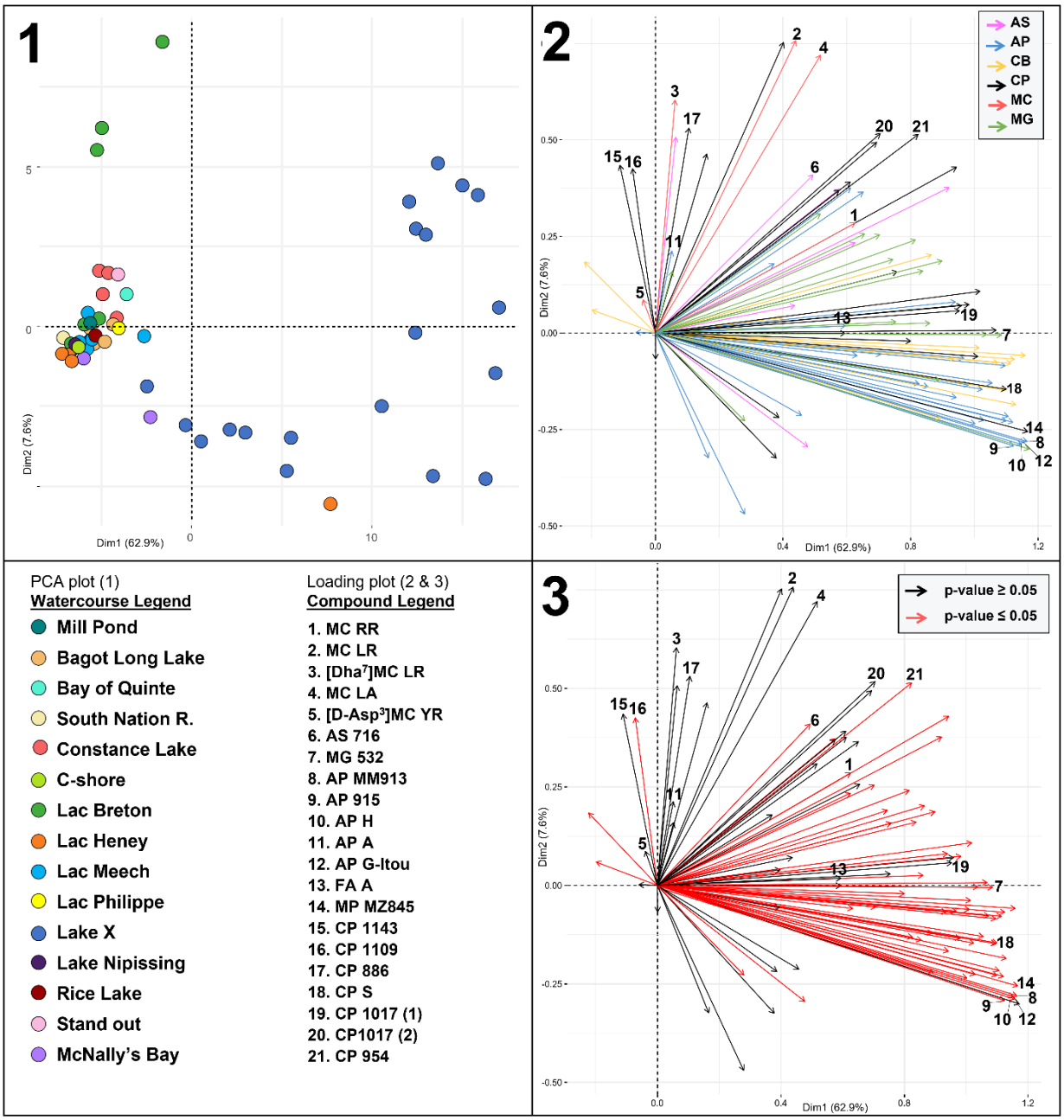

Figure 21. (1) Principal component analysis (PCA) plot of CNP features from 55 blooms samples collected from 15 watercourses. Of the 117 CNPs detected, 95 are considered in the PCA as they had peak noise threshold above $1E^{+06}$. PCA plotted watercourse features are identified by the corresponding legend, bottom left. **(2, 3)** Loading plots contain 95 of the 117 CNPs detected. Loadings plot CNP features are annotated by the corresponding compound legend, bottom left. **(3)** Red vectors indicate statistical significance in contribution to variance of individual watercourses considered for the PCA ($p < 0.05$, KW test); black vectors have no significant variation ($p > 0.05$). **(2)** Vectors are coloured based on CNP chemical group legend at the top-left and abbreviated as follows: MC - microcystins, AP - anabaenopeptin, AS - aeruginosin CB - cyanobactin, CP - cyanopeptolin, MG – microginin. Annotated

cyanopeptolins and anabaenopeptin features have cumulative semi-quantitative concentrations >100 µg/g.

variants being non-significant. APs and CBs produced the majority of watercourse profile variance with only 33% (11/33) and 20% (3/15) non-significant congeners, respectively. The most ubiquitous CNP groups found within the fifty-five blooms samples follow this order: MC > CP > AS > MG > AP > CB. The ten most non-significant CNPs are listed as follows: AP 706 (pval=0.91), AP 923 (pval=0.88), CP 805 (pval=0.69), CP1051 C (pval=0.61), CP 886 (pval=0.58), AS 736 (pval=0.58), AS 609, (pval=0.58), CP 1051 B (pval=0.58), AP A (pval=0.46), CP 920 (pval=0.46).

2.4.3. LC-(HR)MS/MS Quantitation of Select Cyanopeptides

A targeted LC-(HR)MS/MS method was applied to quantify MC-LR, [Dha^7] -LR, -RR, -LA, -LY, -LF, -YR, WR and FA A, MG FR1, CP 983. These are the cyanopeptides where commercial, or in-house, standards were available. MC LA and -WR had the highest detection limits with LOQs of 0.05 µg/mL, followed by MC LY with 0.01 µg/mL (Table 5). The LOQ for CP 982 and MG FR1 was 0.005 µg/mL whereas, the remaining targeted compounds had LOQs of 0.001 µg/mL (Table 5). All targeted analytes had LOQ values equivalent to LOD, with exception of MC LY (LOD = 0.01 µg/mL; Table 5). To optimize targeted quantification of MCs, product ions of the at *m/z* 135.0804 and *m/z* 213.0857 were designated a quantifier and qualifier ions (Table 5). These are generally the most intense MC fragments – product ions attributed to Adda and Mdha structural moieties, characteristic of MCs (and nodularins) [119,142,209]. (HR)MS/MS scans of each analyte's quantifier ion in HESI+ mode was used to achieve greater detection limits. The MS/MS quantifier ion peak areas were selected to quantitate targeted analytes with available reference standards (Table 5). In the case of MC-RR the double-protonated ion $[M+2H]^{2+}$ at *m/z* 519.7915 was predominant, due to the protonated arginine residues at the X and Z positions, thus it was targeted for MS/MS. All (HR)MS/MS calibration curves had good linearity with R^2 values > 0.996. MC WR used the qualifier ion of *m/z* 159.0908 corresponding to the Trp residue at X position of the molecule. [Dha^7]MC LR used *m/z* 107.0855 as the qualifier ion due to the absence of the Mdha

in its peptidyl backbone. For FA A, a quantifier ion at *m/z* 84.0810 and qualifier ion at *m/z* 114.0547 were used, related to the diagnostic D-Lys moiety of the AP group. CP 982 and MG FR1 were not detected in the fifty-five bloom samples. Of the eleven MC certified reference materials used for quantification only seven were detected; MG FR1, CP 982, MC LF and MC YR were not detected. MC LR and FA A were most commonly detected, found in 67% (37/55) and 53% (29/55) of cABs (Table 9). MCs were found in 76% (42/55) of the analyzed cABs (Table 9). Other selected MC congener were detected in the following percentage of cAB samples: MC LA 49% (27), MC RR 31% (17), [Dha^{7}]MC LR 29% (16), MC LY 7% (4), [D-Asp^{3}]MC-LY 5% (3), MC WR 2% (1; Table 9).

Table 9. Concentrations MCs and FA A displayed as µg per g of dry cellular biomass (µg/g d.w.)

			Concentration, (µg/g d.w.)								
No.	**Watercourse**	**Date**	**FA A**	**MC RR**	**MC LR**	**[Dha^{7}] MC LR**	**MC WR**	**MC LA**	**MC LY**	**[D-Asp^{3}] MC LY**	**Total MC**
5	Bagot Long Lake	2016-11-16	2	–	–	<1	–	–	–	–	<1
6	Bagot Long Lake	2016-11-16	3	–	<1	–	–	1	–	–	2
7	Bagot Long Lake	2016-11-16	<1	–	–	–	–	–	–	–	–
8	Bagot Long Lake	2016-11-16	1	–	–	<1	–	–	–	–	<1
9	Bagot Long Lake	2016-11-16	2	–	–	–	–	–	–	–	–
15	Bay of Quntie	2016-08-23	<LOQ	2	13	<1	25	194	–	–	236
1	Constance Lake	2017-05-25	–	–	4	–	–	25	–	–	28
3	Constance Lake	2017-06-07	–	–	77	3	–	1097	–	–	1177
4	Constance Lake	2017-06-07	–	–	1	<1	–	57	–	–	58
2	Constance Lake	2017-08-24	–	–	3	–	–	–	–	–	3
40	C-shore	2009-10-08	–	–	–	<1	–	–	–	<1	<1
24	Lac Breton	2011-07-04	–	–	1	–	–	106	–	–	108
10	Lac Breton	2017-05-21	–	–	2	–	–	53	–	–	55
11	Lac Breton	2017-05-21	–	–	3	–	–	53	–	–	57
54	Lac Breton	2019-08-13	–	–	8	<1	–	–	–	–	9
55	Lac Breton	2019-08-13	–	–	10	<1	–	–	–	–	11
12	Lac Breton	2019-09-02	–	–	–	–	–	–	–	–	–
41	Lac Heney	2009-06-11	–	–	–	–	–	–	–	–	–
49	Lac Heney	2009-06-11	–	–	–	–	–	–	–	–	–
51	Lac Heney	2009-06-11	–	–	–	–	–	–	–	–	–

43	Lac Heney	2009-06-30	249	<LOQ	13	–	–	85	–	–	98
20	Lac Meech	2016-11-29	<LOQ	–	–	–	–	–	–	–	–
21	Lac Meech	2016-11-29	–	–	<1	–	–	–	–	–	<1
22	Lac Meech	2016-11-29	<LOQ	–	–	–	–	–	–	–	–
25	Lac Meech	2016-11-29	<LOQ	–	–	–	–	–	–	–	–
26	Lac Meech	2016-11-29	–	–	–	–	–	–	–	–	–
14	Lac Meech	2019-06-17	7	–	1	–	–	6	–	<1	7
13	Lac Philippe	2019-06-19	<1	–	–	–	–	–	–	–	–
45	Lake X	2013-06-10	10	<1	9	<1	–	271	9	–	289
28	Lake X	2013-06-18	20	<1	6	<1	–	199	–	–	205
29	Lake X	2013-06-18	24	<1	5	–	–	251	<1	–	257
42	Lake X	2013-06-18	26	<LOQ	1	–	–	9	–	–	10
52	Lake X	2013-06-18	19	<LOQ	2	<1	–	13	3	–	18
27	Lake X	2013-07-09	12	<LOQ	4	–	–	13	12	–	29
32	Lake X	2013-07-16	–	–	–	–	–	–	–	–	–
33	Lake X	2013-07-16	–	–	–	–	–	–	–	–	–
35	Lake X	2013-07-16	<1	–	<1	–	–	–	–	–	<1
39	Lake X	2013-07-16	4	<LOQ	2	–	–	–	–	–	2
44	Lake X	2013-07-16	<1	–	<1	–	–	–	–	–	<1
50	Lake X	2013-07-16	–	–	<1	–	–	–	–	–	<1
30	Lake X	2013-10-10	6	<1	3	<1	–	483	–	–	487
31	Lake X	2013-10-10	1	<1	7	–	–	252	–	–	259
34	Lake X	2013-10-10	<LOQ	<LOQ	<1	–	–	10	–	–	10
36	Lake X	2013-10-10	5	<1	<1	<1	–	285	–	–	286
37	Lake X	2013-10-10	9	<1	11	<1	–	508	–	–	520
53	Lake X	2013-10-10	13	1	18	<1	–	545	–	–	566
23	Lake X	2013-07-09	1	<LOQ	1	–	–	–	–	–	1
18	Lake Nipissing	2016-08-07	–	–	<1	–	–	3	–	–	3
19	Lake Nipissing	2016-08-07	–	–	<1	–	–	–	–	–	<1
38	McNally Bay	2014-08-14	–	–	<1	–	–	15	–	–	15
48	McNally Bay	2014-08-14	<1	–	–	–	–	10	–	–	10
16	Mill Pond	2016-07-12	–	–	<1	–	–	44	–	–	44
47	Rice Lake	2013-08-22	–	<LOQ	4	<1	–	74	–	–	78
17	South Nation River	2016-07-12	–	–	<1	–	–	–	–	–	<1
46	Stand Out	2009-08-27	–	–	–	–	–	–	–	<1	<1
55	*M. a.* CPCC 300		–	–	3100	150	–	–	–	–	3250
56	*M. a.* CPCC 632		1099	–	–	–	–	–	–	–	–

% of cABs detected	53%	31%	67%	29%	2%	49%	7%	5%	76%
No. of cABs detected	29	17	37	16	1	27	4	3	42
Maximum concentration	249	2	77	3	25	1097	12	<1	1177
Average concentration	22	2	8	3	25	173	8	<1	154
Minimum concentration	<LOQ	<LOQ	<1	<1	25	1	<1	<1	<1

– = Not detected below LOD
<LOQ = a detection below a quantifiable level
* = Targeted analytes
n.d. = not determined

MC LR was the most common MC detected, in thirty-seven samples. However, MC LA concentrations were consistently one order of magnitude higher than MC LR when detected simultaneously (Table 9). The highest MC LA concentration found was 1,097 µg/g in a 2017 Constance Lake cAB and the highest value for MC LR (e.g., 77 µg/g) was within the same sample (Table 9). Overall, MC LA was the dominant congener in the twenty-seven cABs it was detected. Whereas, MC LR, [Dha^{7}]MC LR and [D-Asp^{3}]MC LY were the dominant congeners in ten, three and one cAB(s) respectively. [Dha^{7}]MC LR and D-Asp^{3}]MC LY never exceeded concentrations of 0.75 µg/g and their concentrations remained low as dominant congener (Table 9). The least common MC was MC WR found in the single Bay of Quinte sample from Lake Ontario, with MC LA as the dominant congener (Table 9). MCs were the least diverse CNP group consisting of seven congeners detected. Furthermore, only one MC congener was detected without a comparable reference standard, [D-Asp^{3}]MC LY. To provide a relative comparison of concentrations across CNP groups, values are reported in µg per gram of extracted cellular biomass.

2.4.3.1. Semi-quantitation of bloom cyanopeptides

Semi-targeted calibration used the most intense precursor ion in the full MS scan of each CNPs groups reference standard. For the cyanopeptide groups AP, MG and CP these were, FA A *m/z* 867.4400 $[M+H]^{+}$, MG FR1 *m/z* 728.4220 $[M+H]^{+}$ and CP 982 *m/z* 965.4770 $[M-H_2O+H]^{+}$. The precursor ion $[M+H]^{+}$ was frequently used for semi-quantitation of variants with exception of the more intense $[M+2H]^{2+}$ for two doubly charged APs and the $[M+H-H_2O]^{+}$ of the CP group. Total

concentrations of each CNP group quantified are listed in Table 10. CPs had the largest group total concentrations ranging from <1 – 15,598 µg/g (Table 10). The greatest semi-quantitative

Table 10. Total concentrations (µg/g) of targeted MCs and semi-quantitated APs, CPs, and MGs displayed with the number of congeners quantitated for the respective CNP group.

			Total concentration (µg/g d.w.)				No. of congeners			
	Quantification method		Targeted	Semi-targeted						
No.	Watercourse	Date	MC	AP	CP	MG	MC	AP	CP	MG
5	Bagot Long Lake	2016-11-16	<1	2	83	–	1	5	2	–
6	Bagot Long Lake	2016-11-16	2	3	10	–	2	4	4	–
7	Bagot Long Lake	2016-11-16	–	1	53	–	–	5	5	–
8	Bagot Long Lake	2016-11-16	<1	1	11	–	1	4	1	–
9	Bagot Long Lake	2016-11-16	–	2	31	<LOQ	–	5	5	1
15	Bay of Quinte	2016-08-23	236	2	4	<LOQ	5	3	1	1
1	Constance Lake	2017-05-25	28	–	609	<LOQ	2	–	8	1
3	Constance Lake	2017-06-07	1177	–	18	<LOQ	3	–	2	1
4	Constance Lake	2017-06-07	58	–	43	<LOQ	3	–	2	2
2	Constance Lake	2017-08-24	3	–	460	–	1	–	7	–
40	C-shore	2009-10-08	<1	<1	12	29	2	4	3	1
24	Lac Breton	2011-07-04	108	24	8	<LOQ	2	1	2	1
10	Lac Breton	2017-05-21	55	–	19	–	2	–	4	–
11	Lac Breton	2017-05-21	57	–	7	–	2	–	2	–
54	Lac Breton	2019-08-13	9	<1	1009	24	2	1	18	16
55	Lac Breton	2019-08-13	11	–	2790	123	2	–	22	16
12	Lac Breton	2019-09-02	–	<1	17	–	–	1	1	–
41	Lac Heney	2009-06-11	–	–	537	1	–	–	15	10
49	Lac Heney	2009-06-11	–	–	–	<LOQ	–	–	–	2
51	Lac Heney	2009-06-11	–	<1	20	–	–	3	1	–
43	Lac Heney	2009-06-30	98	4221	77	–	3	24	1	–
20	Lac Meech	2016-11-29	–	<1	–	–	–	3	–	–
21	Lac Meech	2016-11-29	<1	<1	–	<LOQ	1	2	–	1
22	Lac Meech	2016-11-29	–	<1	9	–	–	3	3	–
25	Lac Meech	2016-11-29	–	<1	453	1	–	2	9	3
26	Lac Meech	2016-11-29	–	<1	54	<LOQ	–	1	1	1

14	Lac Meech	2019-06-17	7	9	399	3	3	5	10	1
13	Lac Philippe	2019-06-19	–	7	41	–	–	5	1	–
45	Lake X	2013-06-10	289	27	15598	1755	5	19	16	11
28	Lake X	2013-06-18	205	613	20	–	4	28	4	–
29	Lake X	2013-06-18	257	679	884	5	4	28	18	12
42	Lake X	2013-06-18	10	532	<1	<1	3	24	1	1
52	Lake X	2013-06-18	18	143	273	<LOQ	5	24	11	4
27	Lake X	2013-07-09	29	186	3	<LOQ	4	25	1	1
32	Lake X	2013-07-16	–	<1	4257	153	–	11	24	17
33	Lake X	2013-07-16	–	76	7392	71	–	19	24	15
35	Lake X	2013-07-16	<1	2	58	3	1	16	7	10
39	Lake X	2013-07-16	2	76	2335	47	2	21	24	15
44	Lake X	2013-07-16	<1	21	1587	35	1	18	18	16
50	Lake X	2013-07-16	<1	19	30	<LOQ	1	16	5	1
30	Lake X	2013-10-10	487	140	3754	82	4	23	25	15
31	Lake X	2013-10-10	259	55	3336	137	3	20	24	17
34	Lake X	2013-10-10	10	464	27	54	3	26	6	4
36	Lake X	2013-10-10	286	77	6046	218	4	22	26	16
37	Lake X	2013-10-10	520	139	111	66	4	23	15	5
53	Lake X	2013-10-10	566	261	43	<1	4	25	1	1
23	Lake X	2013-07-09	1	51	7	<LOQ	2	15	2	1
18	Lake Nipissing	2016-08-07	3	<1	14	–	2	1	1	–
19	Lake Nipissing	2016-08-07	<1	–	7	–	1	–	2	–
38	McNally Bay	2014-08-14	15	<1	4042	15	2	2	23	13
48	McNally Bay	2014-08-14	10	2	–	247	1	14	–	2
16	Mill Pond	2016-07-12	44	<1	–	–	2	1	–	–
47	Rice Lake	2013-08-22	78	<1	1034	64	4	4	18	16
17	South Nation R.	2016-07-12	<1	–	88	–	1	–	3	–
46	Stand Out	2009-08-27	<1	52	276	6	1	1	17	5
	M. aeruginosa CPCC 300		<1	–	4800	n.d	2	–	13	n.d.
	M. aeruginosa CPCC 632		2	1130	65000	n.d.	–	5	7	n.d
% of cABs detected			76%	80%	91%	67%				
No. of cABs detected			42	44	50	37				

Maximum concentration	1177	4221	15598	1755
Average concentration	–	263	<1	143
Minimum concentration	<1	<1	<LOQ	<LOQ

– = Not detected <LOD
<LOQ = a detection below a quantifiable level
* = Targeted analytes
n.d. = not determined

concentrations of all CNPs were detected in Lake X, Lac Heney and Lac Brenton, in this order (Table 10). MCs APs and MGs has similar total concentration ranges with <1-1200 µg/g, excluding the Lac Heney #43 cAB sample (Table 10). The Lac Heney #43 cAB had concentration values an order of magnitude greater than other cABs. 117 CNPs were identified throughout the fifty-five bloom samples using metabolomic LC-(HR)MS/MS approaches. Eighty-six analytes were determined semi-quantitatively in comparison to the seven analytes were quantitated using targeted methods. Two lab cultured strains of *M. aeruginosa* (CPCC 300 and CPCC 632) were quantified with the same analytical method. McDonald (2021) [142] previously identified the metabolome for all reported APs, CPs and MGs produced by *M. aeruginosa* CPCC 300 and CPCC 632. The magnitude at which lab cultured strains produced CNPs was similar to several observed in cABs extracts (Table 9 and 10). APs and CPs were the most diverse CNP groups detected in this study with thirty-three and thirty-two congeners respectively (Table 10).

CP variants were the most commonly detected CNP group, found in 91% (50/55) of cABs sampled (Table 10). The percent occurrence of other semi-quantitated groups is was 80% (44/55) for APs and 67% (37/55) for MGs (Table 10). CP concentrations > 300 µg/g were observed for MP MZ845, CP S, CP 1017 (1), CP 954, CP 875, NP BN920, CP 1051 (C), CP 1051 (A), CP 1017 (2) and CP 1143 (Table S1). Total CPs concentrations were greater than the other quantitated CNP groups (Table 10). CP 847, NP BN920, CP 954 and MP MZ845 were detected most frequently: 69% (38/55), 47% (26/55), 45% (25/55) and 44% (24/55) of cABs, respectively. Three Lac Breton cABs (Lac Breton 5, Lac, Breton SS 1) had CP 1143 concentrations above 100 µg/g (Table S1). MP MZ845 was the dominant CP in fourteen cABs, within Lake X, McNally's Bay, Lac Heney. The

potentially new, CP 847 was the dominant CP in thirteen cABs found in Bagot Long Lake, Lac Breton, Mill Pond, Lac Meech and Lac Heney. Additionally, CP 847 was the most common individual CP but had only a single detection above 100 µg/g. No CPs were detected in the Lake Nipissing, Stand out and Rice Lake cABs. NP BN920 was a commonly detected CP, in 47% (26/55) of cABS, seven with concentrations > 100 µg/g. CP 1017 (1), Mp MZ845 and CP S were detected in much larger concentrations for the group (> 1000 µg/g) but occurred in no more than 18 – 44% of samples (Table S1). Of all CPs, MP MZ845, CP S, CP 954 and CP 1051 C were the most ubiquitous and abundant in studied cABs. Interestingly, the less common CP 1143 was exclusive to Lac Breton and a single Constance Lake sample. CP 1143 was the dominant CP congener for all cABS in which it was detected. The three Lac Breton samples containing CP 1143 had consistent concentrations > 150 µg/g, greater than that of [total MCs] for these samples (Table S1). *M. aeruginosa* CPCC 632 produced CPs in large concentrations (> 20,000 µg/g), similar to concentrations observed in the Lac Heney No. 43 cAB (e.g., ~15,598 µg/g). In summary, CP 1143, MP MZ845, CP S and CP 954 possessed the largest concentrations and most frequent occurrence. In addition, concentrations of these congeners exceeded that of MCs, more frequently at higher concentrations (Table S1). Interestingly, MP MZ845, CP S and CP 954 frequently co-occurred in cABs and contributed largely to the total CP content (Table S1). CP 1143 was limited primarily to Lac Breton cABs however, it was consistently detected in concentration >100 µg/g when total MCs concentrations were ≤ 10 µg/g (Table S1).

APs and MCs had similar concentration ranges (~10 – 700 µg/g; Table 10). The Lac Heney No. 43 cAB was an exception, having a [total AP] of 4,200 µg/g, approximately ten times greater than any other cAB (Table 10). *M. aeruginosa* CPCC 632 had a total AP concentration of 1,100 µg/g, CPCC 300 produced no APs (Table 10). Individual AP concentrations remained ≥ 100µg/g, with exception of the Lac Heney No. 43 cAB (Table 10). AP B and FA A were the most commonly detected AP found in 56% and 53% of cABs. All but one analyzed cAB had FA A concentrations

< 30 µg/g (Table S1). AP G_{Itou}, AP MM913, AP 915, AP H, AP 906b, AP KB899 had the greatest concentrations of the group, in this order (Table S1). AP G-Itou was the dominant congener in fifteen cABs found in McNally's Bay, Lake X and Lac Heney. Constance Lake, Lac Breton, Lac Meech, South Nation River, Rice Lake and Lake Nipissing had no detectable APs or low concentrations (<1 µg/g; Table 10). FA A and AP MM913 were the dominant congeners in six and four cABs, respectively (Table S1). Four cABs had either AP A, AP B or AP D as the dominant congener, however [total AP] values never exceeded 25 µg/g for these instances. Total AP concentrations exceeded total MCs in twenty-seven instances and eleven blooms had [total APs] five time greater than [total MCs] (Table 10). Five blooms had total AP concentrations ~10x higher than that of [total MC] (Table S1). In contrast, [total MCs] exceeded APs in twenty-three blooms, with nine of twenty-three being ~10x higher than total APs (Table 10). Overall, FA A, AP G_{Itou}, AP MM913 were the most ubiquitous and abundant in studied cABs (Table S1).

MGs were less frequently detected in comparison to other studied groups, detected in 67% (37/55) of cABS (Table 10). Excluding MG 532 and MG FR12, all semi-quantitated MG values remained below 25 µg/g for all cABs. A single Lack Land sample had a MG FR12 value of 87 µg/g. MG 532 had several detections > 50 µg/g and a value of 1,600 µg/g in the Lac Heney #43 sample (Table S1). In addition, MG 532, was the most common MG detected, found in 42% of analyzed cABs (Table S1). MG 532 was most ubiquitous variant of the group and possessed similar concentrations to most abundant CPs and APs (Table S1).

2.5. DISCUSSION

2.6. Discussion Multivariate Analysis

The diverse and complex CNP mixtures of cABs translates to unique chemical profiles of individual blooms. The PCA analysis of the ninety-five dominant CNPs visualized the variation among cABs CNP profiles. The loading plot distinguished specific CNP group features

contributing to cAB profile variation. Furthermore, the non-parametric KW test identified the specific CNP features which are non-significant in driving the overall variation between watercourses. Non-significance is an indication of ubiquity throughout analyzed cABs implying, less influence on profile variations. All but one MC was non-significant indicating their ubiquity throughout the blooms sampled. Overall MCs and CPs were the most ubiquitous CNP groups followed by AS > MG > AP > CB, in that order. Within the PCA Lake X displayed the largest variations among its CNP profiles. Lake X is a privately-owned man-made watercourse with a history of chemical treatments and removal of aquatic vegetation (ionized mineral matrix Cu^{2+} and/or diquat dibromide). The historical management of the waterbody has favored cyanobacteria growth dominance and the diverse chemical profiles produced by it's resident cyanobacteria spp. are an anticipated result of this [210]. The McNally's Bay samples were from an Upper Rideau Lake bloom that drew 2014 media attention [211]. Lake Heney experienced blooms throughout the 2000s as a result of P loading from fish farming operations in the Lake, subsequent iron chloride treatments have observably reduced cABs [212,213]. McNally's Bay 1 and Lac Heney #43 grouped with several of the Lake X samples along the bottom of the PCA. Lac Heney #43 shared 6 ASs, 12CBs, 15 CPs, 9 MGs, 23 APs with the Lake X cABs whereas, McNally's Bay 1 shared 2 ASs, 6 CBs, 5 CPs, 1 MG, 14 APs. Both the Lac Heney #43 and McNally's Bay 1 cABs are similar to some Lake X samples but lack the extreme group diversity observed in others. Lac Heney profiles share AS 654, MG 556 and several APs which are significant in driving the variation. In contrast, McNally's Bay does not share these and therefore is located further left of the y-axis on the PCA. Three Lac Breton group separately to the upper left of the PCA due to the presence of CP 1109 and to a lesser extent CP 1143. CP 1109 is exclusive to the three Lac Breton cABs and is significant in driving variation of CNP profiles (pval = 0.031). CP 1143 is shared with the Constance Lake (Dock) cAB and consequently is non-significant, as such. The Lac Breton CNP profiles are less diverse than those of Lake X, Lac Heney and McNally's Bay. Lac Breton was the easternmost lake sampled and much smaller than Lac Heney or McNally's Bay. However, the

Lac Heney and McNally's Bay cABs produced more diverse profiles of CNPs when compared to Lac Breton. The variability in metabolite profiles of cABs is inherently related to genetic composition of the resident cyanobacteria communities. The slight contrast in CNP profiles of McNally's Bay, Lac Heney and Lake X suggests the genotypic compositions of these blooms are similar. Interestingly, the majority of watercourses lack the CNP diversity observed within cABs pertaining to these three watercourses. The more diverse profiles observed in McNally's Bay, Lac Heney and Lake X are a hypothesized result from anthropogenic factors (i.e., nutrient loading and chemical treatments). Anthropogenic influences, local climate and hydrological variations from watercourse-to-watercourse can be expected to cause the majority of CNP profile variations observed in the PCA. Lake X cABs are diverse due to the lake's shallow morphology and lack of macrophyte competition (due to chemical treatment) for nutrients and light, allowing cyanobacteria to dominate [210]. Lac Heney and McNally's Bay are much larger lakes and are not chemically treated therefore, their cABs do not reflect the prolific CNP diversities observed in Lake X. However, both these watercourses have a history of anthropogenic nutrient loading from residential and commercial sources. Moreover, the Rideau Lakes have recently experienced cABs earlier and more commonly during the spring-summer seasons if conditions are favorable, a possible consequence of a changing climate [211,214]. Overall, results indicate that Lac Breton cABs are much less diverse in their genotypic composition in comparison to those from Lac Heney, Lake X or McNally's Bay. The loadings plot and KW test distinguished the most common CNPs in studied cABs (Fig. 21). 87% of MCs were found to be the most commonly detected group followed by 66% of CPs. Many cyanotoxin studies report MCs as the dominant metabolite in freshwater cHABs. These results generally agree with the results of the few similar studies that explore environmental concentrations of CNPs [137,143,200]. However, gaps in the literature still exist as there are few studies which analyze several CNP groups concurrently. MC RR was the only congener significant in driving profile variation (p-val = 0.009), it was exclusive to Lake X, Bay of Quinte, Lac Heney and Rice Lake cABs. CPs were the second most ubiquitous class with the

most common congeners being, the new CP 805 (pval=0.69), CP 1051 C (pval=0.61), CP 886 (pval=0.58). Although, APs produced the majority of variation between watercourses, two new APs and AP A were found to be quite common throughout cABs, AP 706 (pval=0.91) and AP 923 (pval=0.88). ASs, CBs and APs were all less common throughout samples therefore the majority of congeners for these groups were significant in driving the inter-watercourse variation. Cyanobactins were the least common group suggesting they are the most variable group among watercourse cAB profiles. Commonly detected CNPs were explored further with semi-quantitation.

2.6.1. Discussion Molecular Network

The combination of multivariate analyses and GNPS MN for LC-(HR)MS/MS metabolomics provides a robust approach to identify compound groups and dereplicate reported of natural products. GNPS MN groups structurally related chemicals using their diagnostic fragments that generate common product ions in MS/MS. The Generated MN contained fourteen clusters identifying the CNPs groups of interest. The diversity of cyanobacterial communities among studied cABs resulted in the production of a wide array of CNPs observed in the MN. The resulting MN displays clusters of annotated CNPs which share contain common structural features. MN analyses are useful for CNPs produced from NRPS, NRPS-PKS and RiPP pathways, incorporating non-proteinaceous and proteins residues. Most CNP groups share diagnostic structural features including partial amino acid sequences or non-proteinaceous amino acids. Conveniently, these diagnostic structural features generate unique product ions in $MS^{/}$and generate cluster groupings in MN analyses. The generation of multiple network clusters for a single group, provides indication of increased structural diversity. Multiple MN clusters pertaining to a single group represent greater shifts in MS/MS spectra reflecting a higher degree of variation within chemical structures. Therefore, the clustering of a single chemical group is indicative of structural variability within the group. This is relevant as the CNPs possess great amount of

structural variability – challenging to annotate with LC-MS data alone. Seeding MNs provides with known data a place to start with unknown congener annotation, as it gives a gauge of structural similarity of the metabolites to previously identified compounds and allows for in-depth subclassification of cyanopeptides. The molecular network and the multivariate analyses show that chemical profiles of the studied cABs are unique. However, there significant structural overlap between CNPs profiles from different watercourses. Hence, MN exemplifies both the metabolomic similarities of CNP biosynthetic pathways and the vast structural variation which arises due to nature of these pathways. This is visualized within the MN subclassifications of the CP group by partial amino acid sequences. For example, *NMe*Phe (or -*NMe*Hty) containing cyanopeptolins with either a Lxx-Ahp or Val-Ahp produced cluster #8 whereas *NMe*Tyr (or *NMe*Tyr-Cl) containing cyanopeptolins with Phe-Ahp or Lxx-Ahp produced cluster #8; CP 847 contained the unique sequence *NMe*Tyr-Val-Ahp and did not cluster within the MN (Fig. 17). Similarly, the same is observed with the AP group as Arg-CO containing APs produce cluster #4 and APs lacking Arg within their structure produced cluster #3 (Fig. 18). As such, single residue variations within a peptide sequence can be indicated by MS/MS and provide further structural classification. An advantageous characteristic, as minor changes in enzyme-substrate specificity of NRPS pathways can selectivity produce of hundreds of congener variations for each metabolite produced by the NRPS and combined PKS pathways [58,60,61,66]. RiPP synthesized peptides (i.e., CBs and MVs) possess greater increased structural variation as the identity, number, and location of PTMs is independent of the core peptide synthesis. Furthermore, many of these PTM enzymes show high plasticity and substrate tolerance. This biosynthesis variability is apparent in the MN cluster produced by cyanobactins which produced three separate cluster groupings.

2.6.2. LC-(HR)MS/MS Identification

The applied reversed-phase HPLC method optimized for CNPs resolved analytes effectively over a 10.5 min solvent program. The RTs of most CNPs groups possessed consistent ranges, with greater vakues observed for MC, CP and CB variants. The RT ranges observed for these groups are a result of their combinatory PKs-NRPs and RiPP synthetic pathways, increasing structural variation and concurrently the range in amphiphilic properties. The target and semi-targeted LC-(HR)MS/MS methods applied for the metabolomic study identified 117 CNPs. Seven MC congers and FA A were quantified with targeted LC-(HR)MS/MS analysis. Eighty-six CNPs pertaining to the groups of AP, CP and MG were determined by semi-quantitative methods. Overall CPs were more frequently detected than MCs, the two most common CNP group. APs and CPs displayed a similar diversity in identified group variants. APs and CPs did occur at greater concentrations than MCs in select cABs however, the opposite was too observed. A single congener from the MG group, MG 532 occurred at similar frequencies and concentrations as MCs, APs, CPs.

2.6.2.1. Targeted Quantitation

Targeted analytes were quantified using an associated reference material. Seven MCs, MG FR1, CP 982 and FA A were quantified by targeted methods. CP 982 nor MG FR1 were detected in the fifty-five bloom samples. MC LR and FA A were most commonly detected, found in 67% (37/55) and 53% (29/55) cABs, respectively. MCs were found in 76% (42/55) of the analyzed cABs. Both CPs (91%; 50/55) and APs (80%; 44/55) were more commonly detected that MCs. Although, MC LR was the most common MC detected, MC LA concentrations were consistently one order of magnitude higher when co-occurring. Overall, MC LA was the dominant congener within the twenty-seven cABs it was detected. MCs mixers are known to be highly variable between cABs representative of cyanobacteria species diversity and toxic and non-toxic traits can be specific to the stain level. The genotypic composition of cABs is largely influenced indirectly by local climate and geography. The observed variability in MC dominance agrees with other studies analyzing MCs in north American freshwaters. Previous studies of Canadian lakes with

similar geography found MC LR and MC LA to be the predominant congeners [215,216]. The results support this as MC LR and MC LA were most dominant. MC LA dominance has been shown to favor smaller freshwater lakes based on local climate and geography [215]. MC LA was the dominant congener in the majority of samples. It is known that different MC congeners are produced by distinct cyanobacterial species or strains. Therefore, MC congener dominance is related to bloom genotypic variation and highly influenced by environmental variables. The variation in MC congener dominance is an inherent result of the resident genera and the environmental conditions during bloom formation. This study observed MC LA the dominant congener in the twenty-seven cABs whereas MC LR was only dominant in four Lake X cABs. A global analysis of MC congeners by Taranu et al. (2019) revealed that MC LA dominance was more prevalent in North and South American cABs [215]. The highest MC LA concentration found was 1,100 µg/g in a 2017 Constance Lake cAB and the highest value for MC LR (77 µg/g) was within the same sample. The least common MC was MC WR found in the single Bay of Quinte sample from Lake Ontario, with MC LA as the dominant congener. [Dha^{7}]MC LR and D-Asp^{3}]MC LY were the dominant congeners in three cABs however, they never exceeded concentrations of 0.75 µg/g and its concentration remained low when detected with other congeners. The cABs analyzed exhibited a large variability in their MC profiles a likely result of the differences in geographic and hydrological conditions in addition, to watercourse morphology. Moreover, cABs from the same watercourse but sampled at different times or locations had highly varying profiles indicating the large temporal and spatial variability in cAB composition.

2.6.2.2. Semi-quantitation

The semi-quantitative approach used a single standard of the CNP group. The accuracy of a semi-quantitative approach is limited in comparison to targeted methods. However, semi-quantitation is favorable when reference standards are not available to achieve a complete quantitative analysis. Therefore, a semi-quantitative approach can fill in gaps, providing an

estimation of concentration where no alternative is available. APs, CPs and MGs are reported in metabolomic analyses of cyanobacteria extracts however, their concentrations and potential toxicological effects are generally unknown [142,143,190,200]. Here a semi-quantitative approach was used to quantify select CNP groups to better examine the toxigenic potential of blooms. CPs and APs were the most commonly detected groups (91% and 80%). CPs had larges range in concentrations (<1 – 15,598 µg/g) and possessed total concentrations greater than the other quantitated group. The groups of AP (80%), MC (76%) and MG (67%) were less common than CPs, occurring in similar percentages of cABs (Table 10). To assess the concentrations of bloom semi-quantified values, two lab cultured strains of *M. aeruginosa* (CPCC 300 and CPCC 632) were quantified with the same method for comparison. A [total CP] value of 65,000 µg/g was observed for *M. aeruginosa* CPCC 632 and 4,700 µg/g for *M. aeruginosa* CPCC 300 (Table 10). A 2013 Lake X bloom had a [total CP] value of 15,598 µg/g however, the majority of blooms contained total CPs in the ≈ 4.0 – 7000 µg/g range. The semi-quantitated values of cABs and cultured strains fall within the same order of magnitude, providing consistency to these results. This agrees with results from studies quantitating CPs from bloom extracts. Neumann et al. (2000) estimated CP concentration to compose 0.01% of dry bloom biomass, greater than that of co-occurring MCs [152]

APs and MCs, were the second and third most common CNP groups detected, sharing similar concentration ranges (≈ 10 – 700 µg/g). A single 2009 Lac Heney cAB was an exception, with a [total AP] of 4,200 µg/g, approximately ten times greater than any other blooms sampled. The same trend was observed for CPs and MGs for this Lac Heney cAB. APs were the third most abundant and ubiquitous CNP group behind CPs and MCs and possessed the greatest overall group diversity with thirty-three congeners detected. FA A was the most ubiquitous variant, detected in 53% of cABs however, all but one had concentrations below 30 µg/g. The majority of individual AP concentrations remained below 100 µg/g, with exception of the single Lac Heney

sample. *M. aeruginosa* CPCC 632 had a total AP concentration of 1,100 µg/g, CPCC 300 produced no APs (Table 10). AP G-Itou, AP MM913, AP 915, AP H, AP 906b, AP KB899 had the greatest concentrations of the group, in this order. Total AP concentrations exceeded total MCs in twenty-seven instances and eleven blooms had [total APs] five time greater than [total MCs]. Five blooms had total AP concentrations ≈10x higher than that of [total MC] (Table 10). In contrast, [total MCs] exceeded APs in twenty-three blooms and nine blooms had [total MC] ten times greater than total APs (Table 10).

Semi-quantitated CNPs were evaluated by both their concentrations and ubiquity throughout studied cABs and compared to that of MCs. Select CPs, APs and MGs were evaluated based on this criterion to indicate specific CNPs that possess the greatest toxigenic potential for this study. Of studied MCs, MC LA was most abundant in cABs and used for comparison of semi-quantitated CNPs. MC LA possessed maximum contraction of 1097 µg/g and an average value of 173 µg/g. However, MC LA was > 250 µg/g in nine cABs. Because of the disproportionate representation of CPs and APs in Lake X an arbitrary concentration 250 µg/g in one or more cABs was used. For CPs a value of 300 µg/g in two or more cABs was used as they had larger concentration ranges than APs. The four CPs, CP 1143, MP MZ845, CP s and CP 954 were determined to be the most abundant and ubiquitous congeners of the group. These CPs displayed an occurrence of 30% or greater in studied cABs and possessed individual concentrations > 300 µg/g. In addition, the average concentrations of these congeners was > 150 µg/g. CP 1143 occurred in less than 30% of cABs but was selected based on its disproportionate occurrence in Lake X and Lac Heney, comparative to other CPs. Additionally, CP 1143 concentration were 100-fold greater than total MCs in Lac Breton. FA A, AP G_{Itou}, AP MM913 and AP 915 were designated as the most abundant and ubiquitous congeners of the group. These variants possessed occurrences > 35% of cABs and were detected in concentrations ≥ 250 µg/g. Average concentrations of these congers was ≥ 50 µg/g, with exception of FA A as it possessed low concentrations. However, FA

A was selected due to its greater ubiquity relative to all other studied APs. A single MG, MG 532 met the concentration and occurrence criteria. MG 532 possessed two concentrations > 200 µg/g in two cABs and a maximum value of 1642 µg/g. In summary, the CP group is the most prevalent and at higher concentrations and CNPs of this group are recommended for prioritization in future studies. The AP group was found in a similar occurrence to CPs however, total concentrations were consistently lower. Based on evaluated concentrations and percent occurrence for this study, nine individual CNPs are recommended for prioritization if future eco-toxicological studies. These include MP MZ 845, CP S, AP G_{Itou}, AP MM 913, CP 954, AP 915, CP1143, FA A and MG 532. CNPs are listed in order of there toxigenic potential indicated by abundance and ubiquity.

2.7. CONCLUSION

LC-(HR)MS/MS metabolomics approaches were applied to identify 117 CNPs and evaluate the CNP profiles of cABs from fifteen watercourses across Eastern Ontario and Western Quebec Lakes. Targeted and semi-targeted mass-spectrometry based metabolomic techniques were used to decipher the CNP profiles of fifty-five cABs. CNP profiles were interpreted and individual CNPs identified using a seeded GNPS MN, DFF and multivariate statistics. The most common CNP groups detected in order were CP, MC, AP and MG. A targeted method was used to quantitate seven MCs congeners and FA A. Concentrations of eighty-six identified CPs, APs and MGs were determined by semi-quantitative methods. MC LR was the most abundant conger detected; however, MC LA was consistently observed in greater concentrations when co-occurring. In addition, MC LA possessed the greatest concentration of all MCs studied. Of all semi-quantitated CNPs, CPs contained the greatest range in concentration and the largest maximum values. Based on CNP occurrence and concentrations relative to that of other groups, the CP group possesses the greatest toxigenic potential is recommend for prioritization in future studies. For this study nine of the most ubiquitous and abundant CNP are highlighted for consideration in future metabolomics studies exploring CNP mixtures. These include four CPs:

CP 1143, MP MZ845, CP s and CP 954; four APs: FA A, AP G_{Itou}, AP MM913 and AP 915; The individual MG 532.

CHAPTER 3: Microcystin Method Development and Multi-proxy Evaluation of Rideau Lake Sediment Archives

3.1. INTRODUCTION

The presence of toxic bloom-forming cyanobacteria is increasing on a global scale [32,39,217]. Nutrient enrichment from anthropogenic sources is a major cyano-bloom driver and has been well documented in large Canadian freshwater lakes such as the Lake Erie (ON), The Rideau Lakes (ON) and, Lake Winnipeg (MA) [32,34,39,211]. In addition, climate change has increased water surface temperatures to favor cyanobacteria growth [18,38,39]. Warmer temperatures and nutrient enrichment from anthropogenic sources (i.e., fertilizers, livestock animal waste and untreated wastewater) have been linked to prolific cyanobacteria blooms, enhancing both their incidence and intensity [18,39,217]. Canadian freshwater lakes are experiencing a similar trend. However, similar trends in bloom events have been observed in small inland Canadian lakes which lack anthropogenic drivers of bloom occurrence. Raising the question if climate may play a greater role in driving bloom occurrence for minimally impacted lakes. One of the most pervasive bloom-forming cyanobacteria genera in freshwater lakes are *Microcystis* spp. [32,116,138]. Many *Microcystis* strains can produce the hepatotoxic microcystins (MC), a group of >250 chemically stable non-ribosomal peptide hepatotoxins and possible human carcinogens (IARC Group 2B) [67,114,138]. The vast array of cyanotoxin groups produced and the variance in toxin production between individual cyanobacteria strains makes it difficult to predict bloom toxicity. The main challenge for studying MCs other cyanotoxin groups is the large structural diversity within the group. MCs require costly

instrumentation to effectively separate MC congeners and provide accurate quantitation of each. Current knowledge lacks the occurrence and composition of historical MCs from freshwater cyano-blooms. Few studies examine the historical occurrence of freshwater cyanobacteria in natural archives. Previous studies that do have had little focus on identifying cyanotoxins in sediments [218,219]. Recent progress was made by Zastepa et al. (2015) to extract MCs from sediments using an ASE base extraction method [220]. Traditionally, microfossils, pigments and DNA analyses have been used to reconstruct historical trends in cyanobacterial assemblages in sediment cores [218]. However, MCs are stable cyclic heptapeptide and pertain a greater chemical stability than that of with DNA or pigments. Therefore, MCs have the potential to be a robust paleolimnological proxy of historical cyanobacteria growth. Here, a method was developed and validated for the extraction and quantification of four MCs (MC LR, -[Dha^7]LR, -RR and -LA) congers from freshwater lake sediment cores. The presented method was further applied with two lake sediment cores in the Rideau Canal system.

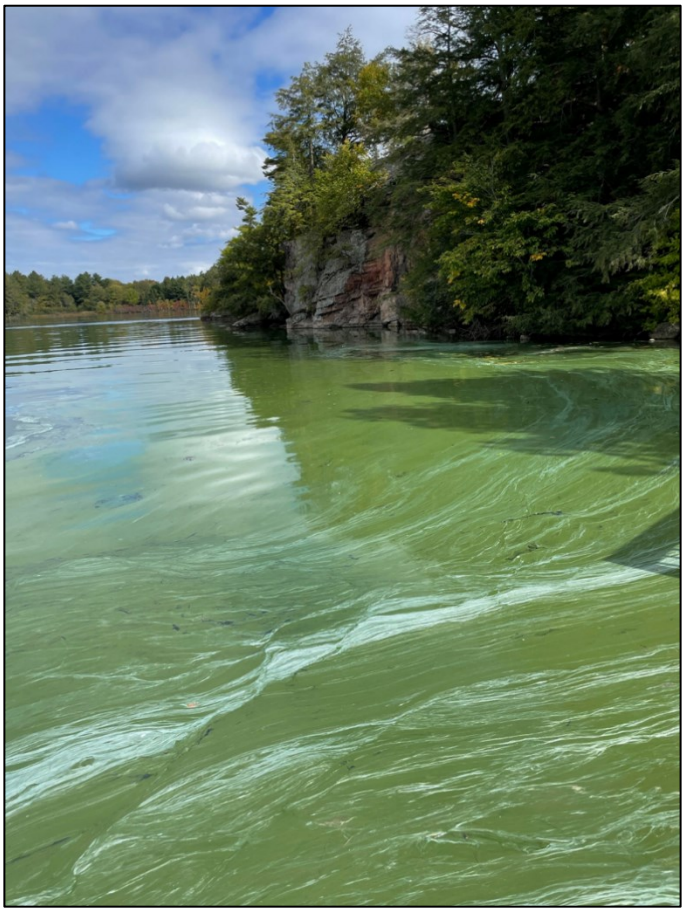

Figure 22. Photograph of a Dog Lake cHAB that occurred in the summer of 2012 (credit: anonymous Dog Lake resident)

The Rideau Canal was constructed between 1826 and 1831 to create a navigable waterway from Lake Ontario to the Ottawa River. It is the oldest continuously operated canal in North America. Several previous studies have employed paleolimnological methods to lake sediment archives within the Rideau Canal system for evaluation of historical conditions [221,222]. A wide range of

indicators are routinely employed in paleoenvironmental studies to determine anthropogenic changes, hydrological conditions and land-use impacts. Frequently, geochemical and organic matter analyses are applied as paleolimnological proxies [223–225]. Within the last few decades, X-ray fluorescence (XRF) core scanners have become increasingly utilized in for geochemical analyses in paleoenvironmental research. New-age Itrax-XRF core scanners are capable of providing high-resolution imagery, X-ray fluorescence (XRF), and magnetic susceptibility data at 0.1 mm resolutions [226–229]. The high resolution provided by Itrax-XRF can enhanced interpretation of climate signals at resolutions difficult to achieve by other traditional analysis. The lack of sample preparation, minimal analysis time, high resolution, and non-destructive nature of Itrax-XRF analyses make it an efficient and effective choice for whole sediment core analyses [227,228]. Itrax-XRF analyses provide semi-quantitative data, a limitation due to the analysis of non-dry sediment at high resolution however, drying of samples prior to analysis can assist this variability. For the current study, two sediment cores were collected from two lakes (Dog Lake and Colonel By Lake) within the southern portion of the Rideau system. Both Lake sediment cores were analyzed by Itrax-XRF and for Chlorophyll-a (Chl-a). Geochemical data and chl-a were applied as paleolimnological indicators of historical system conditions. In addition, the developed method for MC quantitation in lake sediments was applied to evaluate method applicability with other proxies inferring paleolimnological condition.

3.2. METHODS

3.2.1. Study Site:

Dog and Colonel By Lake are holomictic freshwater inland lakes situated north-east of the Kingston Ontario, (≈9 and ≈21km, respectively) and south of Battersea (Fig. 23). The two Lakes are located in the Cataraqui watershed occupying the southern portion of the Rideau Canal

system. Both Lakes were created in 1831 when Rideau Canal construction was completed, and subsequent flooding inundated a 22 km^2 portion of the lower Cataraqui watershed.

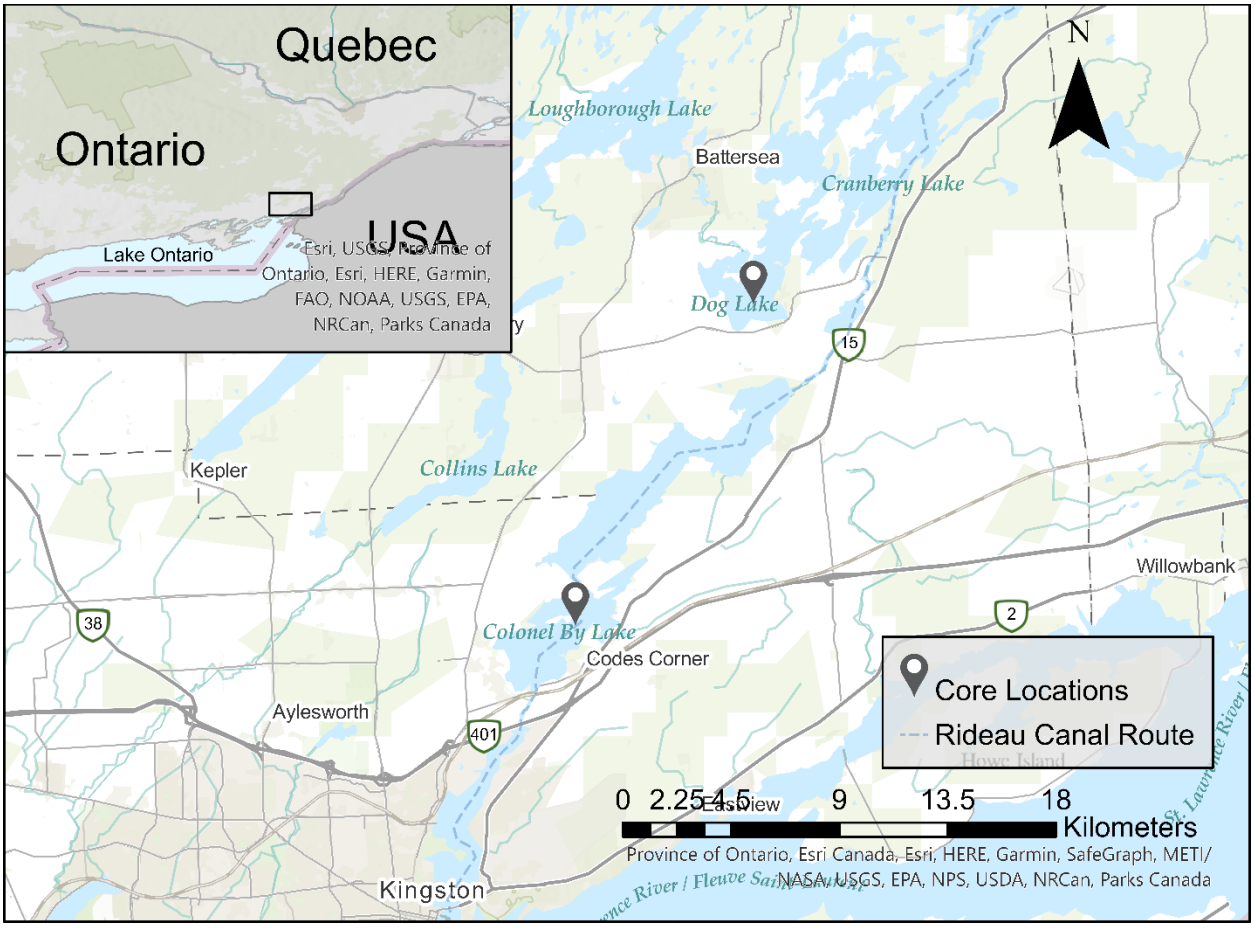

Figure 23. Arc®GIS Pro generated map, indicating locations of Colonel By and Dog Lake relative to the Rideau Canal waterway. Collection sites of the lake sediment cores are annotated with a location pin.

Dog Lake encompasses a surface area of 9.64 km^2 with a shoreline circumference of 59km [230]. It varies in depth; the average depth is 5.79 m and the maximum depth is 50 m; northern portions of the lake contain deeper waters (10-50 m) whereas the southern basin is shallow (<3 m). Colonel By Lake is smaller possessing a 5.29 km^2 surface area with a shoreline circumference of 27.8 km; waters enter the lake through the River Styx and outflows are regulated by a set of locks at southern end, Kingston Mills [231]. The Lake is generally shallow with an average depth of 4.0 m and a maximum depth of 6.1 m. Both systems are eutrophic in status with average TP values >0.02 mg/L [230,232]. The northern portions of Dog Lake appear to have lower nutrient values,

associated with deeper waters [230]. Both these lakes reside on the Frontenac axis – an exposed Precambrian rock which connects the Canadian shield (west) and the Adirondack mountains (east). The surficial bedrock geology of both lakes is generally composed of Precambrian metasedimentary rocks and limestones deposited by marine sedimentation during the Paleozoic [221,222,233,234].

3.2.2. Sample collection:

Individual Lake sediment cores were collected from Dog Lake and Colonel By Lake near Kingston, Ontario, Canada. The Dog Lake core was collected from a central location of the southern basin (Fig. 23). The Colonel By Lake core was collected within the main canal channel (historically the Cataraqui river channel between Caseys Island and the west shoreline (Fig. 23). Cores were collected using a surface core sampler (UWITECH surface corer, USC 06000). Each core was transported indoors, and sub-sampled at 0.5 cm depth intervals using a vertical extruder. Sub samples were collected into pre-labeled 18 oz. Whirl-Pak® Bags and transported in a cooler with ice and placed in freezer storage at -20°C, prior to sample preparation and instrumental analysis.

3.2.3. Microcystin (MC) Analysis

3.2.3.1. Reagents and Standards

Microcystin certified reference materials of MC LR (CAS: 101043-37-2), [Dha^7]MC LR (CAS: 120011-66-7), MC LA (CAS: 96180-79-9), NOD R (CAS: 118399-22-7) and MC RR (CAS: 111755-37-4) were procured from the National Research Council of Canada's (NRC) Metrology Research Centre (Halifax, NS). Analyte and internal standard spiking solutions were prepared in 50% aqueous methanol (ddH_2O and HPLC grade methanol, Sigma-Aldrich)

3.2.3.2. Experimental: MC extraction

Each sub-sample of the sediment core was homogenized and a 10 g aliquot was removed for extraction. Each 10 g aliquot was halved and split between two 50 mL centrifuge tubes (polypropylene, FroggaBio Inc.), with 5 g (± 0.1) in each tube. Both half samples were centrifuged (10,000 rpm, 4°C, 25 min). Each tube containing the 5 g was spiked with 5 µL of 1 ug/mL NOD-R standard and mixed by vortex. 20 mL of 75% aqueous (ddH_2O) methanol (HPLC grade) was added to each tube. Each sample was vortexed then sonicated for 30 s and placed on an orbital shaker (VWR® Advanced 3500 Orbital Shaker) for extraction at 320 rpm for 30 mins. Half samples were centrifuged again (10,000 rpm, 15-20 min, 4°C), and the resulting supernatant methanol extract of each half sample being decanted into a single 500 mL round-bottom flask. Again, 20 mL of 75% aqueous methanol was added to each tube, and the extraction process repeated a second time. The resulting methanolic extracts of each individual sample was combined into the same 500 mL round-bottom flask. Methanol was removed from extracts by rotovap and replaced by an equal amount of ddH_2O (~30 mL).

Solid phase extraction (SPE) was used to increase microcystin recovery from extracts. SPE was preformed using Hydrophilic-Lipophilic-Balanced (HLB) SPE cartridges (Waters, Oasis HLB 150 mg/6cc/30µm). Cartridges were primed using 10 mL of HPLC grade methanol followed by 10 mL of ddH_2O. Extracts were added to the cartridges at a flow rate of ≈4 mL/min. Cartridges were washed with 12 mL of 20% aqueous methanol to reduce matrix effects. Each cartridge was left to dry prior to elution. Bound solids were eluted from the cartridge using 6 mL methanol, to 14 mL glass amber vials (Type I, Class B amber borosilicate glass, Fisher Scientific). The resulting methanolic extracts were gently dried under a stream of nitrogen gas (4.8HP) at 35°C. Dried extracts were then reconstituted with 1.5 mL methanol and transferred to 2 mL amber HPLC vials (Agilent). The extracts were again dried by nitrogen gas, then stored at -20 °C prior to LC-(HR)MS, MS/MS analysis.

3.2.3.3. Liquid chromatography (LC) Conditions

All LC-(HR)MS/MS analysis was performed using an Agilent 1290 HPLC coupled to a Q-Exactive Quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific). Samples were reconstituted in 1 mL of 90% aqueous methanol. 10 µL volumes were injected onto a C18 Eclipse Plus RRHD column (2.1 × 50 mm, 1.8 µm; Agilent Technologies) maintained at 35 °C and flow rate of 300 µL/min with LC-MS grade water with 0.1% formic acid (mobile phase A) and acetonitrile with 0.1% formic acid (mobile phase B). The 10.5 min solvent gradient program began with mobile phase A. mobile phase B was increased from 0% to 35% over 1.5 min and again to 45% over 2.5 min. Mobile phase B was increased to 100% B over 2 min, and held at 100% B for 3 min prior to returning to 0% B over 0.5 min and maintained for 1 min.

3.2.3.4. High-resolution mass spectrometry ((HR)MS) Conditions

All (HR)MS and (HR)MS/MS data was obtained using a Q-Exactive Quadrupole Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a HESI-II, heated electrospray ionization source (HESI). Analytes were ionized using $HESI^+$ with the following settings: capillary voltage, 3.9 kV; capillary temperature, 400 °C; sheath gas, 17 units; auxiliary gas, 8 units; probe heater temperature, 450 °C; S-Lens radio frequency level, 45.00. (HR)MS data was acquired using a targeted acquisition that included a full MS scan at 35,000 resolution, with a scan range of 150.0000 – 2000.0000 *m/z*; automatic gain control target, 1×10^6; and a maximum injection time of 128 ms. MCs and NOD were targeted from each full scan for MS/MS analysis using 1.2-Da isolation and a 0.8s retention time window. MS/MS scans used the following conditions: resolution, 17,500; automatic gain control target, 5×10^6; max intensity threshold, 64 ms. The retention time, precursor ion *m/z*, and optimized NCE used for each targeted analyte are listed Table 11.

Table 11. LC-(HR)MS, (HR)MS/MS parameters for targeted MCs and NOD

Name	Chemical formula	*m/z*	RT	NCE	Quantifier, Qualifier ion *m/z*
MC LR	$C_{49}H_{74}N_{10}O_{12}$	995.5560 $[M+H]^+$	3.55	45	135.0798, 213.0858
[Dha^7] MC LR	$C_{48}H_{72}N_{10}O_{12}$	981.5404 $[M+H]^+$	3.55	45	135.0798, 213.0858
MC RR	$C_{49}H_{75}N_{13}O_{12}$	519.7915 $[M+2H]^{2+}$,	2.89	38	135.0798, 213.0858
MC LA	$C_{46}H_{67}N_7O_{12}$	910.4921 $[M+H]^+$	5.19	41	127.0861, 213.0858
NOD R	$C_{41}H_{60}N_8O_{10}$	825.4505 $[M+H]^+$	3.20	38	135.0798, 227.1013

3.2.3.5. Method Performance

Blank sediment for use as a laboratory blank sample matrix (LBSM) that was obtained from the southeastern portion of Constance Lake, ON – a watercourse with a documented history of low MC concentrations [220,235]. Prior to experimental method validation, collected blank sediment was extracted and analyzed by LC-(HR)MS/MS to confirm the absence of MC congeners. A microcystin (MC) mix spiking solution of 1 µg/mL was prepared in 50% aqueous methanol from concentrated standards of each reference material (i.e., NOD R, [Dha^7]MC LR, MC LA, and MC RR). The MC spiking solution was stored at -20°C prior to sediment extractions.

Recovery efficiency (R_E): Laboratory fortified sample matrix (LFSM) samples were prepared by pre-spiking MCs to LBSM samples at a concentration of 10 ng/g d.w. Five LFSMs were prepared by spiking prior to extraction to simulate normal recovery. Five LBSMs were extracted, then spiked after extraction procedures to simulate 100% recovery, referred to as post-spike LFSMs. Pre-spike LFSMs were spiked post pore-water removal by centrifugation, then kept at room temperature and away from light for four hours prior to extraction to ensure sufficient absorption equilibrium of the spiked MC was reached [220]. Post-spike LFSMs were spiked after elution from SPE.

$$R_E = \frac{\text{Avg. Pre} - \text{spiked LFSM}}{\text{Avg. Post} - \text{spiked LFSM}}$$

Signal suppression/enhancement (SSE): Laboratory fortified blanks (LFB) were prepared by spiking 10 µL of the MC spiking solution to empty HPLC vials, letting them air dry prior to reconstituting in 1 mL of methanol.

$$\%SSE = \frac{\text{Avg. Post} - \text{spiked LFSM}}{\text{LFB}}$$

Third extraction recovery (R_3): A third sequential extraction was conducted with an additional 20 mL volume of 75% aqueous methanol and analyzed separately to evaluate its efficacy.

$$R_3 = \frac{\text{Avg. Third Extraction}}{\text{Avg. Post} - \text{spiked LFSM}}$$

3.2.3.6. Microcystin Quantitation

Peak area integration and quantification was performed using *Xcalibur™(v3.0.63)*. Automated peak integration was conducted using the Genesis peak detection algorithm, 7-point smoothing, enabled valley detection and 0.5 S/N threshold. Analyte peaks were designated by presence of a precursor ion (± 5 ppm), associated retention time (± 0.8 min) and detection in ≥ 7 scans. Analytes were quantified using their (HR)MS precursor ions which provided greater sensitivity. Linear calibration curves were generated from the associated chemical and internal standard, under identical analytical conditions using a 1/x weighting factor. An internal standard calibration curve containing concentrations of 0.1, 0.5, 1, 5, 10, 50, 100, 500, 1000 ng/g was prepared to quantitate MC analytes. Five replicates were prepared for all levels of the calibration curve and each spiked with 10 µL of a 1 µg/mL of internal standard (NOD). internal standard calibration curves were generated using the following equation:

$$\frac{peak\ area_{MC}}{peak\ area_{NOD_{IS}}} = slope\frac{[MC]}{[NOD_{IS}]} + intercept$$

The average dry weight (d.w.) of sediments analyzed in this study (1.26 ± 0.01 g) was used to express the instrumental limits in terms of mass. There exists no classical baseline noise level for data generated by the high-resolution mass spectrometers, hence the limits of detection (LOD) and quantification (LOQ) cannot be calculated based on the traditional methods. Therefore, the lowest concentration level of the calibration curve wherein five consecutive injections produced a detectable signal was defined as the limit of detection (LOD) [202]. The lowest concentration at which the peak area relative standard deviation (RSD) of five consecutive injections is below 20% was defined as the limit of quantification (LOQ) [202]. All limits of detection and analytical figures of merit are listed in Table 12.

Table 12. Method parameters for MC quantitation and instrumentation limits

Name	Method equation	Correlation Coefficient (R^2)	LOQ (ng/g)	LOQ %RSD	LOD (ng/g)
MC LR	y = 0.6193x + 0.0127	0.996	0.40	14	0.08
[Dha^7] MC LR	y = 0.5941x + 0.0023	0.996	0.40	11	0.08
MC RR	y = 2.5769x - 0.0968	0.997	0.40	6	0.08
MC LA	y = 0.4624x - 0.0231	0.997	0.08	4	0.08

3.2.4. Radiocarbon Dating

Three core depths, top (0.5 cm), middle (14.5 cm) and bottom (28 cm) were prepared for radiocarbon dating. Each homogenized bulk sediment sub-sample weight, approximately 1 g wet, was placed into 2 mL microcentrifuge tubes. Samples were prepared for accelerator mass spectrometer (AMS) analysis by triple acid wash following procedures outlined by Crann et al. (2016) [236]. Carbon radioisotopes were measured by 3MV AMS (High Voltage Engineering) at the University of Ottawa's A.E. Lalonde AMS facility. Age models for Colonel by and Dog Lake cores were developed using Bayesian Age Calibration (*Bacon*) package with *R(v4.1.3)* [237]. Radiocarbon dates were calibrated using the terrestrial radiocarbon curve IntCal13 [238].

3.2.5. ^{210}Pb Dating

Fifteen Colonel By Lake and eighteen Dog Lake sediment core subsamples, were prepared and analyzed at the Palaeoecological Environmental Assessment and Research Laboratory (PEARL) at Queen's University, Canada. Subsamples were freeze-dried, placed into gamma-tubes to a height of about 2.5 cm and sealed using epoxy over a silicone septum and left to reach equilibrium for 3 weeks. Using standard methodologies, subsamples were analyzed using an Ortec high-purity Germanium gamma spectrometer to measure the gamma activity of ^{210}Pb, ^{214}Pb, ^{214}Bi and ^{137}Cs. Lead dating relies on the assumption that a sediment core contains a consistent unmixed temporal structure where incorporated ^{210}Pb follows a natural order of accumulation, unaffected by redistribution. Calendar dates was generated using ^{210}Pb activities and a Constant Rate of Supply (CRS) model [239]. The CRS model assumes a constant rate of ^{210}Pb atmospheric deposition and a thus constant supply rate of unsupported ^{210}Pb. Unsupported ^{210}Pb of the initial activity (at time zero) is inversely proportional to the sedimentation rate, therefore increase in autochthonous erosion dilutes the supply of unsupported ^{210}Pb. The CRS ages-depth relationship is described by the following equation where, A_i = $^{210}Pb_{unsupported}$ (Bq/kg) accumulated below the depth interval corresponding to time (t_i); A_0 = total residual $^{210}Pb_{unsupported}$ (Bq/kg) in the entire core; λ = ^{210}Pb disintegration constant 0.03114 per year) [239,240].

$$A_i = A_0 \times e^{-\lambda t_i}$$

The sedimentation rate (r, g/cm^2) is given depth by the following formula where, C_i = $^{210}Pb_{unsupported}$ activity at t_i (Bq/kg) [239,240].

$$r = \lambda \frac{A_i}{C_i}$$

The age of a given sediment depth interval (t_i) is then given by:

$$t_i = \frac{1}{\lambda} ln \frac{A_0}{A_i}$$

Extrapolated curves for the CRS age models were generated in MS excel and by *R(v4.1.3)* with the *Bacon* package [237]. It is commonly advised that extrapolating CRS age models should be avoided as they can overestimate ages from CRS models in deeper sediment layers [241]. However, extrapolation was used here due to the lack of viable ^{14}C dates for extension of chronological dates to core bottom.

3.2.6. Chlorophyll-a (Chl-a) reflectance spectroscopy

Visible reflectance spectroscopy was used to reconstruct lake sediment Chl-a concentrations using methods outlined by Das et al. (2005) and Wolfe et al. (2006) [242,243]. Samples were prepared for spectroscopy by lyophilization and 125 µm sieve. Reflectance spectral measurements were performed with visible-near infrared reflectance spectroscopy using a spectroradiometer operating at a 400–2500 nm range. An average of ~30 scans were acquired per sample with a white reference panel was used as a comparable standard. A range of sedimentary Chl-a concentrations were measured HPLC by methods outline by Vinebrooke and Leavitt (1999) [244]. A subsequent calibration model was developed quantify reflectance peak area of the Chl-a at 650 and 700nm using HPLC-measured Chl-a concentrations.

3.2.7. CONISS analysis:

For stratigraphic interpretation of Itrax-XRF results, constrained incremental sum of squares (CONISS) cluster analysis was applied to (Ti, Zr, Sr, Fe, Rb, Mg, K, Ca CIR, S, P, Si, Cl) to provide delineation of major shifts in lake environmental conditions over time. Ratios of select elemental variables of environmental importance were used in a second CONISS analysis with MCs and

Chl-a to further assist in evaluating environmental change. Statistical analyses and figures were generated using *R (v. 4.1.3)* and RStudio including the packages *vegan* [245] and *rioja* [246].

3.2.8. Itrax-XRF Analysis

Dog and Colonel By Lake sediments from each sub-sampled depth interval were homogenized by hand and sequentially loaded to a Itrax-XRF sequential sample reservoirs (SSR; Fig. 24) [247]. Sediment subsamples were loaded to each compartmental reservoir of the SSR using a metal spatula. Loaded SSRs were left to dry and settle overnight at room temperature. Repeated loadings of SSRs occurred until dry subsample reservoirs were flush with the top surface of the SSR and no voids remained. All discrete subsamples for both cores were analyzed by XRF using an Itrax Cox Core scanner at McMaster University's core laboratory by methods outlined by Braden Gregory (2017) [228]. Raw Itrax-XRF data was deconvoluted of all measurements not corresponding to sample reservoirs of the SSR. This included removal of measurements of the SSR's acrylic edges and those collected within ≤ 1 mm of a reservoir's edge. Measurements collected ≤1 mm of the acrylic edge were removed to

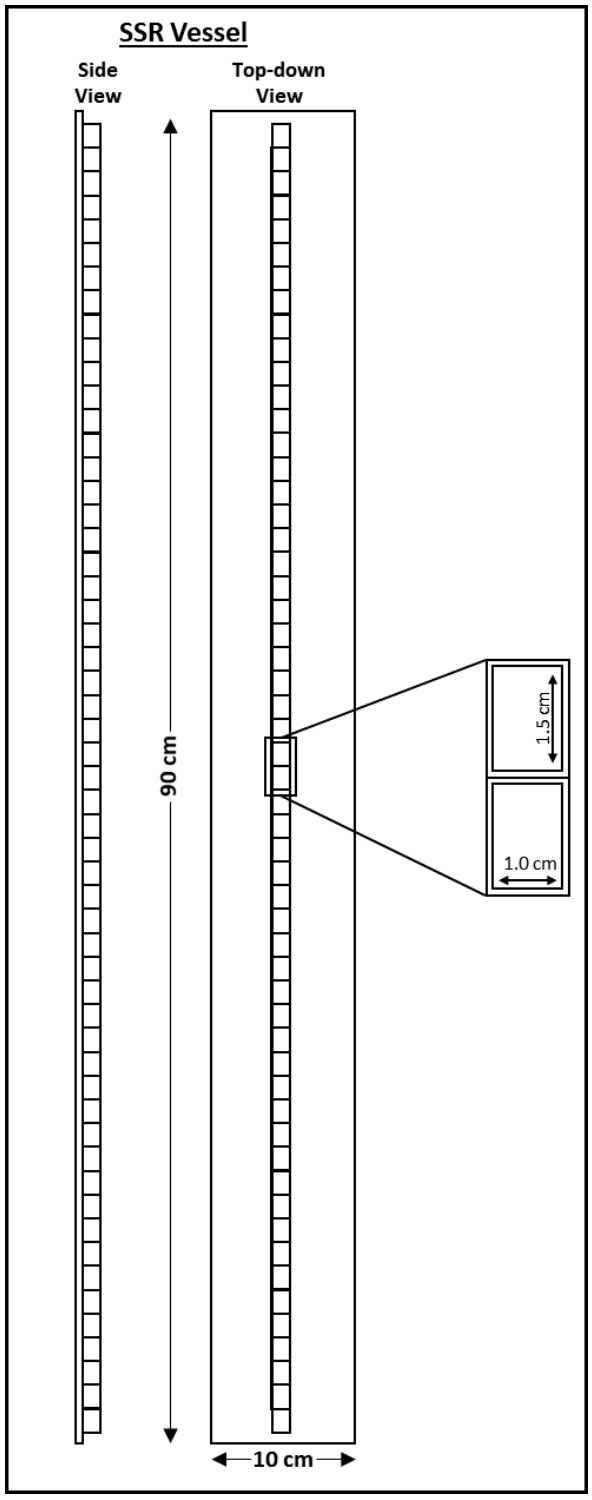

Figure 24. Schematic of the SSR vessel and reservoir dimensions used for Itrax-XRF analysis, adapted from Gregory et al. (2017) [247]

minimize potential boundary effects. Sample measurements were distinguished from measurements of acrylic border by observable shifts in total number of counts for a given interval. This included, major decreases in abundant element concentrations such as Fe, substantial decreases in the ratio of coherent/incoherent backscatter, and knowledge of SSR compartment well length and position [228,247,248]. Further normalization of the data to reduce matrix effects used the mean square error (MSE) and coherent/incoherent X-ray backscatter ratio (CIR). Values influence by voids and air pockets reflected MSE values ≥ 2 times the average standard deviation of elemental counts per second (cps). All XRF cps values were divided by the CIR for normalization. Ratioing elemental counts by CIR has been demonstrated as an estimated correction for water and organic matter content [247,248]. All Itrax-XRF results are expressed as unitless normalized cps. Several elemental ratios were determined (Ti/Ca, Fe/Ca, Br/Ti), selected as paleolimnological proxies reflective of both historical environmental processes occurring at the site. Itrax-XRF analyses are effective at providing a suite of elemental data for whole core sample in a timely manner at low-cost. However, limitation of Itrax-XRF analyses is the difficulty to achieve accurate analysis of lighter elements in comparison to more sensitive and robust methods (i.e., ICP-MS). All normalized Itrax-XRF results and ratios considered for the study are listed in tables S3 and S4.

3.3. RESULTS

3.3.1. Method Validation

A 75% aqueous methanol solution was selected as a suitable solvent for MCs due to its amphiphilic nature and provided good R_E values for selected MCs and NOD from spiking experiments. MC LR, [Dha^7]MC LR, MC LA and NOD all had higher R_E values, 90%, 100%, 80% and 80%, respectively. The trends in R_E values are similar to previous studies; MC RR possessed the greatest affinity for sediment whereas [Dha^7]MC LR and MC LR had the lowest affinity

[220,249,250]. Overall, the recoveries expressed in this study are similar to those found by Wu et al. (2012) and are notably 5 – 20% higher than other previous studies exploring MC extraction from sediments [220,249,250]. The %SSE for all analytes was between 93 – 110% (Table 13). The method reproducibility was defined as the percent relative standard deviation for five pre-spiked LBSFMs and evaluated at concentrations of 0.1 ng/g and 0.2 ng/g; both concentrations produced comparable values > 20% RSD (Table 13). The subsequent 3rd extraction of MCs (R_3) found that limited (<10%) MCs remained after the two initial extractions and therefore a 3rd extraction was not required. The precursor ion of MCs provided greater sensitivity than MS/MS quantifier ions and was therefore used to achieve lower detection limits for quantitation.

Table 13. Method validation parameters of targeted MCs and NOD

Compound	R_E (%)	Reproducibility 0.1ng/g (%RSD)	%SSE	Reproducibility 0.2 ng/g (%RSD)	R_3 (%)
MC RR	71	6	98	8	3
MC LR	90	15	93	5	nd
[Dha^7]MC LR	100	16	97	8	nd
MC LA	80	11	100	-	-
NOD	80	19	110	-	-

- = not preformed
nd = not detected
R_E = recovery efficiency
%SSE = signal suppression/enhancement
R_3 = 3rd extraction recovery

3.3.2. Microcystin Analysis

The developed analytical method was used to quantitate the selected MC congeners in sediment cores from both Lakes studied. Chromatographic conditions resulted in good separation of MC LA, MC RR and NOD. However, it did not effectively resolve MC LR and [Dha^7] MC LR. The precursor ion with the greatest intensity was selected for quantitation of analytes, predominantly the $[M+H]^+$ ion *m/z*. MC RR was an exception, with a more intense $[M+2H]^{2+}$ ion associated with its enhanced protonation susceptibility from two basic Arg groups within the molecule. LOQ and LOD values for MCs in lake sediment were determined in the low ng/g d.w. range, similar to

previous studies exploring the same concept [220]. Colonel By Lake sediment possessed a single MC LR detection below LOQ at a 2.5 cm depth and provided no other detectable MC concentrations (Fig. 25). [Dha^{7}]MC LR was not detected in either core. Dog Lake core samples contained detectable

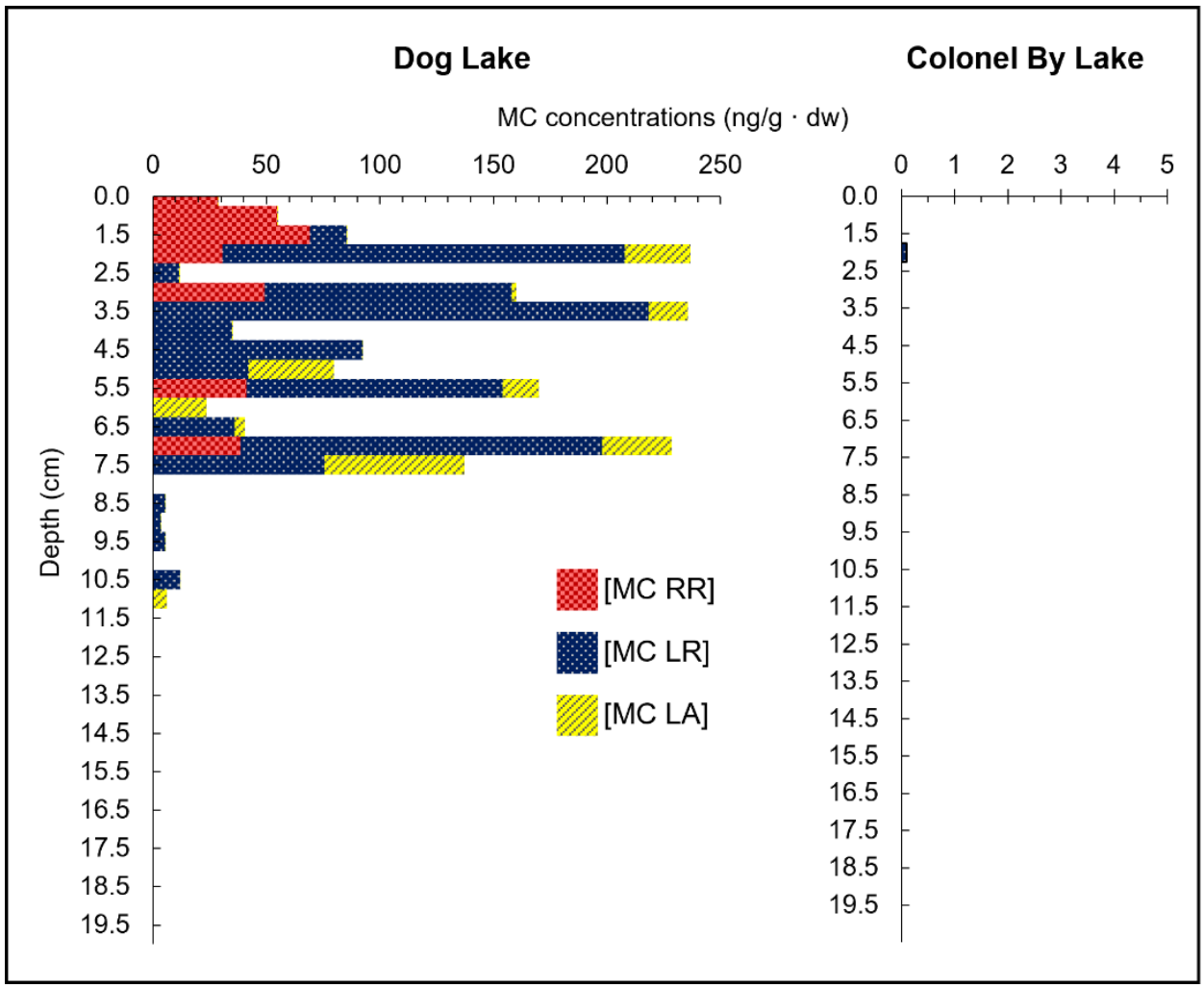

Figure 25. Down-core profiles of quantified MCs for Dog Lake (left) and Colonel By Lake (right).

concentrations of the three congeners, MC RR, MC LR and MC LA. MC LR was most abundant in Dog Lake sediment, followed by MC LA and MC RR (Fig. 25). No detectable concentrations of MCs were found below the core depth of 17.5 cm (Fig. 25). The majority of detectable MCs was concentrated in depths above 10.5 cm. Below this, MCs were detected at levels ≥ LOQ. MC LA was concentrated deeper in the core (5.0 – 11.0 cm) comparatively to MC RR (0.5 – 7.0 cm; Fig.

25). Several samples contained concentrations above the detection limit (LOD), but below the quantification limit (LOQ), therefore were enumerated by a placeholder value of 0.1 ng/g to provide visual representation. MC LR provided the highest concentrations in comparison to other congeners. The greatest concentration of MC LR was 237.5 ng/g, determined at a depth of 2.0 cm (Fig. 25). Overall, MCs were most abundant above a depth of 7.5 cm, whereas the opposite is observed deeper in the core.

3.3.3. Sediment Core Chronologies:

The resulting radiocarbon dates for both Dog and Colonel By lakes were much older than anticipated, shown in Table 14. It is suspected that the samples were hampered by the presence of old carbon in the system resulting in observably older ^{14}C dates. The presence of old carbon is thought to be associated with canal flooding and flushing, resulting in redistribution of older organic carbon deposits and/or the assimilation of dissolved inorganic carbon derived from carbonate bedrocks or other sources.

Table 14. Radiocarbon dating results from the Lalonde AMS facility (University of Ottawa)

Core	Depth (cm)	^{14}C yrs. BP	$F^{14}C$	Date (AD/BC)
Colonel By	0.5	5 ± 31	0.9935 ± 0.0038	1955 AD
	14.5	120 ± 32	0.9852 ± 0.0039	1830 AD
	29	565 ± 35	0.9321 ± 0.0041	1385 AD
Dog	0.5	943 ± 25	0.8892 ± 0.0028	1007 AD
	14	1125 ± 25	0.8693 ± 0.0027	825 AD
	27.5	4860 ± 27	0.5461 ± 0.0018	2730 BC

Age-depth modeling using Bayesian age calibration (*Bacon*) of the data with *R(v4.1.3)* provided chronologies that spanned ~4000- and ~600-year periods for Dog and Colonel By Lakes respectively. The larger date range observed for Dog Lake is suspected to be associated with pre-canal wetland conditions and the heightened accumulation of older residual organic carbon in greater concentrations. The disturbance of Rideau Canal flooding likely resuspended and

redistributed residual organic carbon throughout the system. Comparatively, Colonel By Lake existed as a main channel of the Cataraqui river prior to Rideau Canal and lacked the high organic matter content that existed with Dog Lake's pre-canal wetland conditions. This suggests that old carbon within Colonel By Lake is derived from dissolved inorganic carbon sources or organic carbon redistributed from elsewhere in the system. In summation, no viable ^{14}C chronological model for either lake could be determined, therefore a ^{14}C chronology was abandoned (Fig. 26).

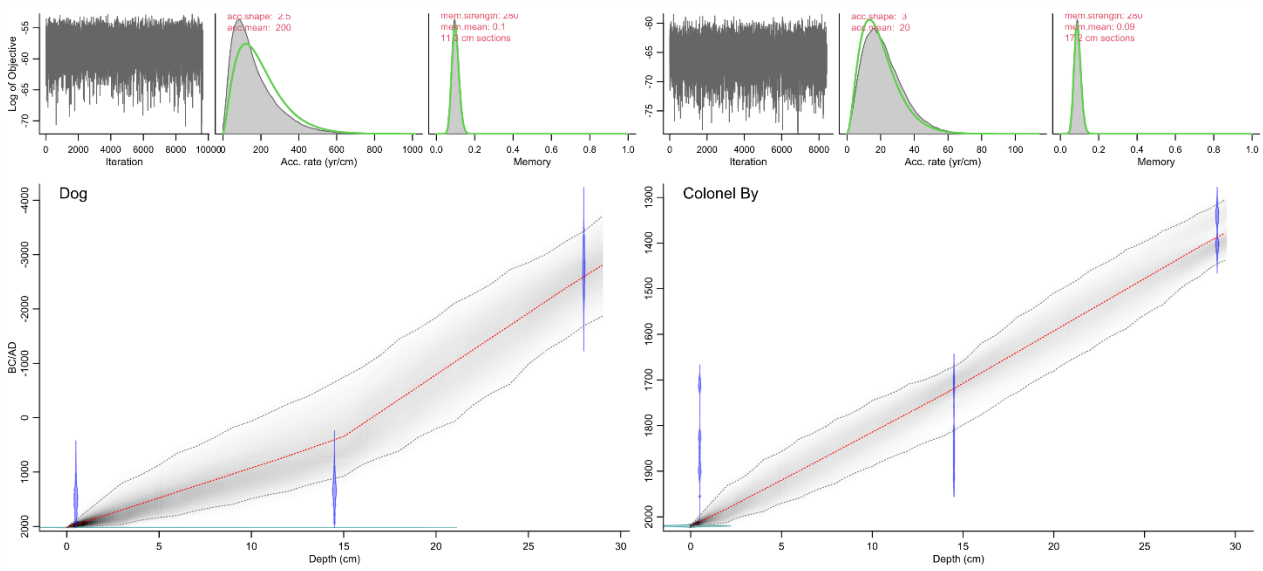

Figure 26. ^{14}C Age-depth models generated in R with a date of 2020 CE (-70 yr. BP) representing core surface. Radiocarbon dates are shown in blue, and model age in grey – darker areas suggest greater probabilities of age at a given depth and average is indicated by red line.

A ^{210}Pb chronology for Dog and colonel By Lakes was established using a Constant Rate of Supply (CRS) model [239]. The generated ^{210}Pb dates for Colonel By Lake spanned a range of 1901 – 2020 from core surface to bottom and ^{210}Pb excess activity never decreased below supported levels (Fig. 27). Colonel By CRS dates possessed larger errors (>5 yrs.) below a depth of 20.5cm ^{210}Pb excess activity in Dog Lake decreased below supported levels beyond a depth of 18.5cm (Fig. 27). At depth of 16.5 – 18.5 cm the Dog Lake CRS dates possessed errors >5 yrs. Dates were generated for depths below 18.5 cm (Dog) and 28.5 cm (Colonel By) using polynomial extrapolation of the CRS age models (Fig. 27).

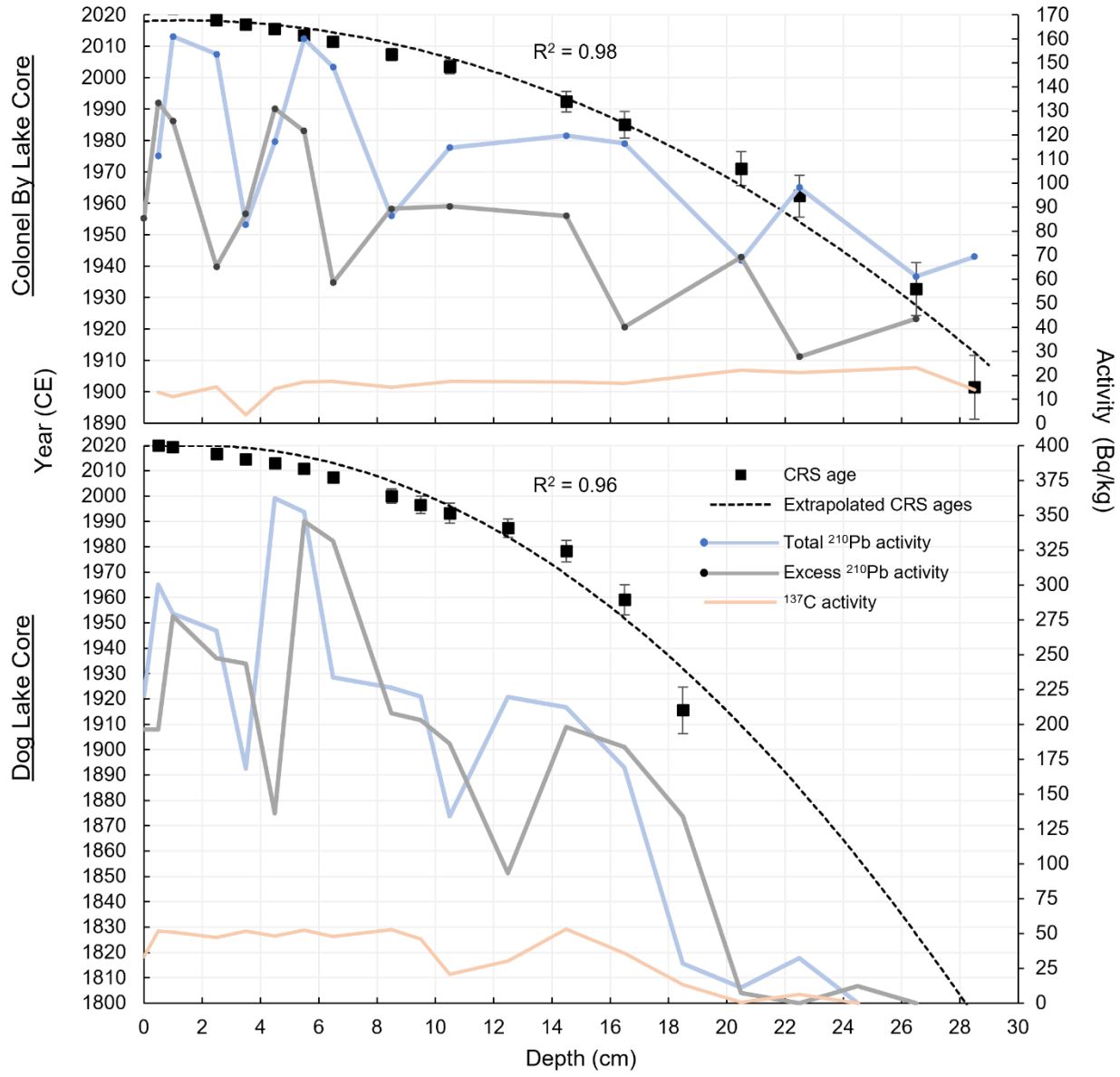

Figure 27. CRS ages and extrapolated ages displayed with 137Cs and 210Pb activities for Colonel By (top) and Dog Lake (bottom) sediment cores.

Resulting ^{137}Cs activities provided one broad spike in Dog Lake core at a 14.5 cm depth and roughly aligns with 1970s given by CRS dates (Fig. 27). Largely, the ^{137}Cs down-core activity of Dog and Colonel By provided no distinctive peaks indicative of specific events during 1950-1980s nuclear testing. Extrapolated dates for Dog Lake provided the estimated date of Rideau Canal flooding (~1831) at a depth of 25.75 cm. The estimated date aligns with observable shifts in the Itrax-XRF geochemical data and is though to be representative of such (Fig. 31). The partial

validation of Dog Lake CRS dates with the single ^{137}Cs peak and shifts in elemental count data indicates the age model is sufficient for estimation of age ranges but cannot provide accurate indication of exact dates. The Colonel By age-model developed contains a large amount of variation due to the lack of validation with anthropogenic radioisotopes and should be interpreted with caution. To confirm the extrapolation performed in excel the CRS dates for both cores were extrapolated using *R(v4.1.3)* and *Bacon* (Fig. 28). The CRS Age-depth models generated in R

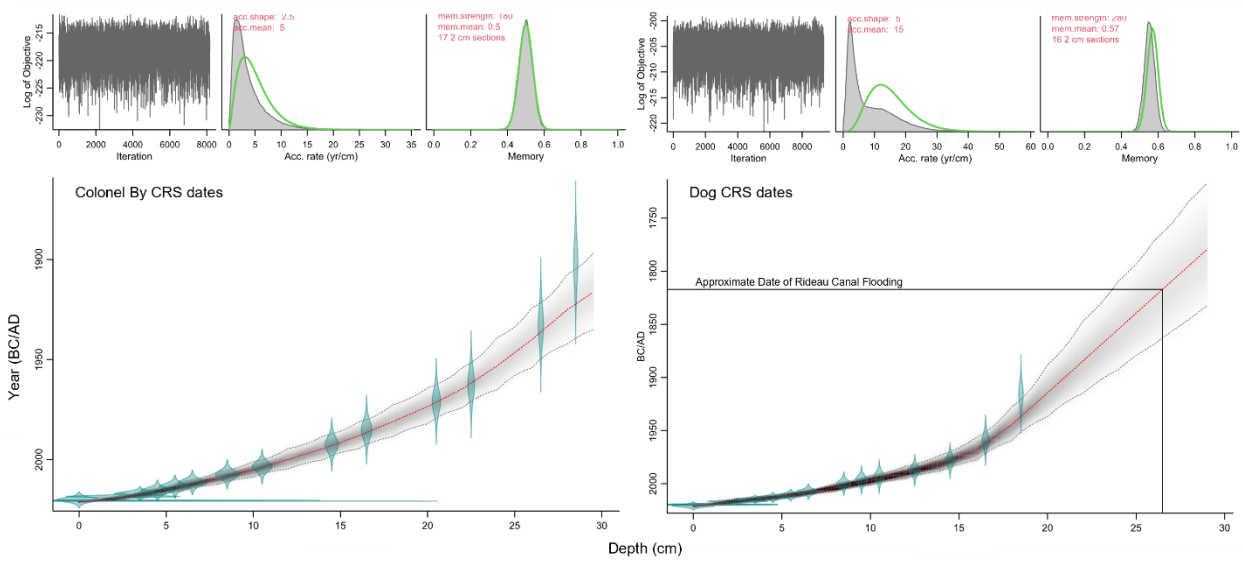

Figure 28. CRS Age-depth extrapolations generated in R. CRS dates and errors are shown in blue, and model age in grey – darker areas suggest greater probabilities of age at a given depth and average is indicated by red line.

Provided similar date ranges to those observed in Figure 27. For example, both extrapolations placed the event of Rideau Canal completion and flooding at a depth of 25.5-26.5 cm (Fig. 27 and 28). The extrapolated model generated in excel was considered for further use as it correlated best with shifts in geochemical count data (Fig. 31).

3.3.4. Itrax-XRF

Overall, both lake sediment cores display temporal and spatial variations in their elemental distributions throughout the study period (Fig. 28 and 29). Elements of interest were further grouped into lithologically derived (Ti, Fe, Zr, Rb, Sr, Mg) and biologically or anthropogenically

influenced elements (K, S, Si, Ca, P, Cl) and CIR. Two CONISS cluster analysis were preformed on each core. The preliminary CONISS analysis was performed solely on elemental data of each core. The subsequent CONISS analysis performed, further assessed the elemental down-core stratigraphy with concentrations of MCs and Chl-a. This included, select individual elements (Mg, P, S), elemental ratios proxies (Ti/K, Si/Ti, Zr/Rb, Ca/Sr, Ca/Ti) MC and Chl-a concentrations.

3.3.5. Colonel By Lake: Down Core Multiproxy Trends

The elemental data generated from Itrax-XRF analysis displayed no distinct down core trends for Colonel By lake. The CONISS cluster analysis of thirteen elements and the CIR (Fig. 29) displayed two distinct zones (CB-1 and CB-2), differentiated at 15.75 cm (~1989; Fig. 29). The elements of Ti (titanium), Zr (Zircon), Sr (Strontium), Fe (iron), Rb (Rubidium) and Mg (Magnesium) showed little variation between these zones. The CIR, K (Potassium) and Ca (Calcium) followed similar trends and displayed an observable increase from CB-1 to CB-2 and a subsequent decline near the cores surface. Cl remained consistently low down-core, spiking at depths above 5 cm in CB-2 – postulated to be of anthropogenic origin (Fig. 29). Similarly, an increasing trend was observed in the biogenic elements of Si (Silicon) and P within CB-2 – shared with other biogenic proxies (Chl-a, Si/Ti, Ca, CIR; Fig. 20 and 30). S (Sulfur), mainly displayed an increasing values within CB-2 and declined the surface, a unique trend (Fig. 29). Three elemental ratios (Si/Ti, Ti/K, Zr/Rb), Mg and P were selected for further CONISS analysis with Chlorophyll-a (Chl-a) and MC data (Fig. 30). The second CONISS analysis provided identical zone to those observed in Figure 29. Mg is a proxy of hydrological condition, increasing at times of negative water balance (low water level) [251]. Mg is typically mobilized by carbonate weathering and endogenic carbonate compounds contain different amounts of Mg, inducing

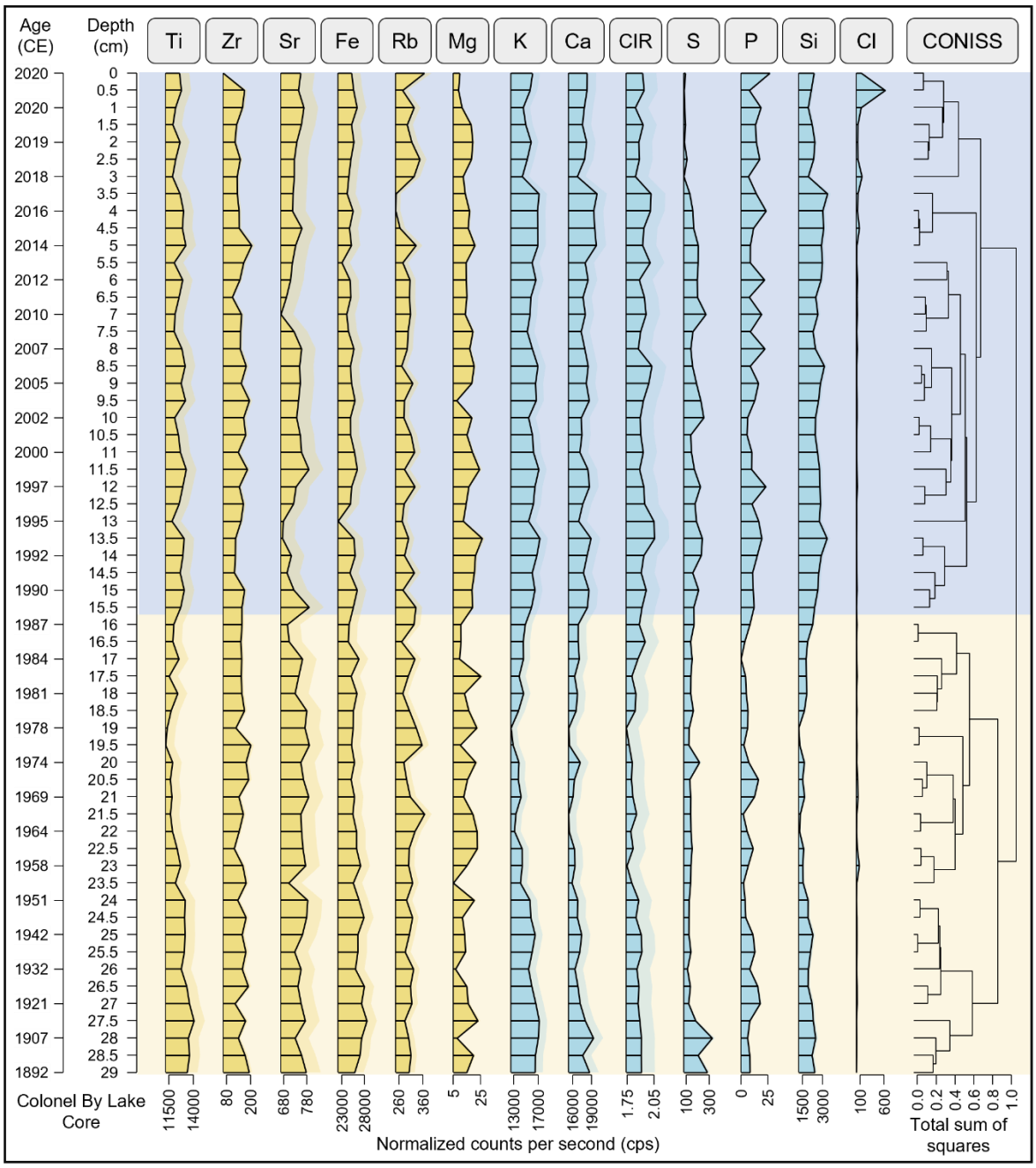

Figure 29. Colonel By Lake down-core stratigraphic profiles of Itrax-XRF elemental count data. CONISS cluster analysis (right) displays two significant zones (CB-1 and CB-2) annotated by colour.

variation during different hydrologic periods [223,251,252]. Zr/Rb ratio is a proxy for grain size as Zr and Rb are generally concentrated in different grain-size fractions with lower values indicating finer sedimentation [223,224]. Both Mg and Zr/Rb show little change between zones and consistently fluctuates throughout the core (Fig. 30). The lack of change between zones and constant

fluctuation of these proxies down-core is thought to be a result of Colonel By Lake's constant change in water level. The Lake's water level is regulated by the locks at its southern end and undergoes a ≈0.5 m change lake water level annually.

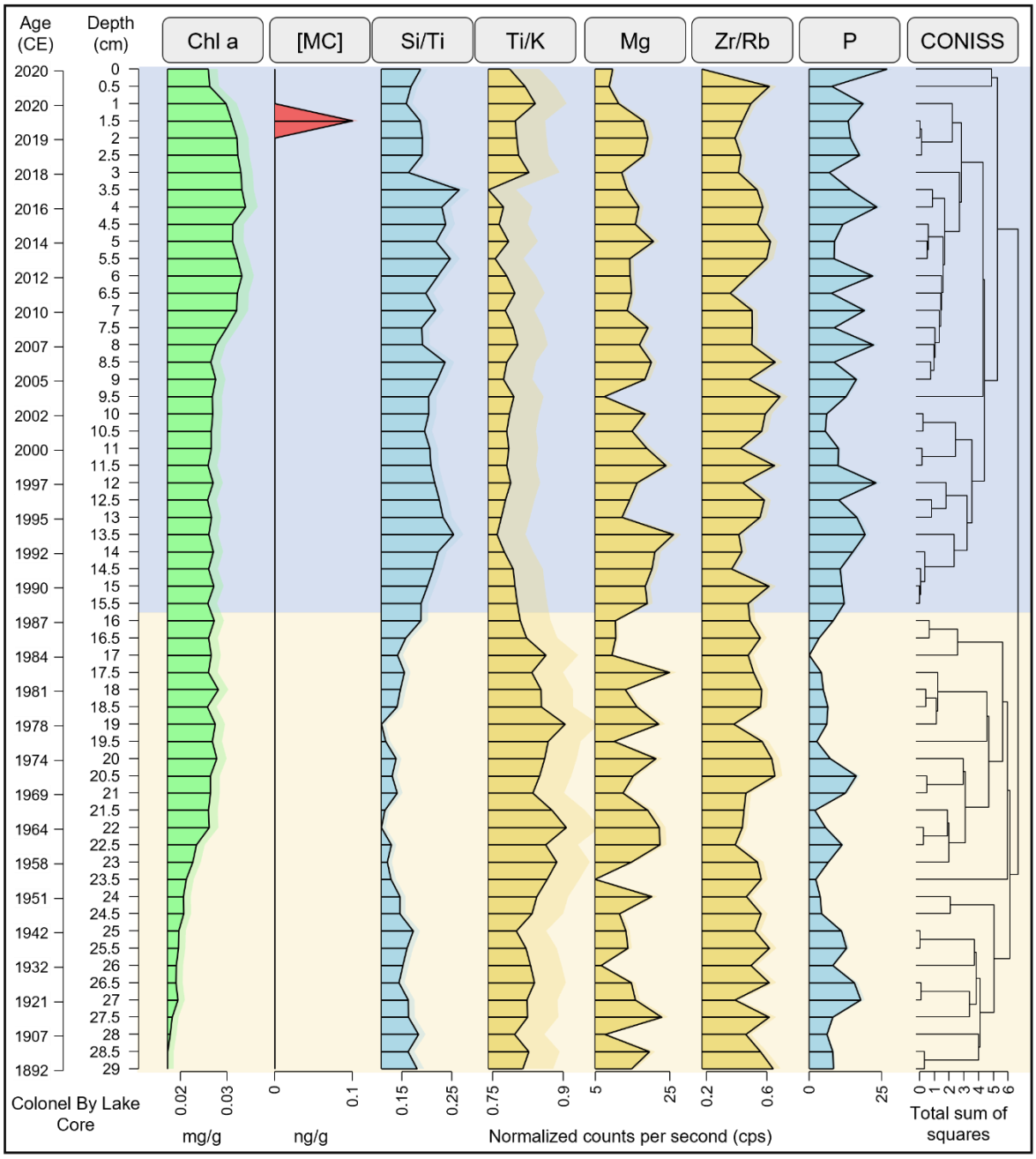

Figure 30. Colonel By Lake down-core stratigraphic profiles of Itrax-XRF elemental count data, Chl-a and MC concentrations. CONISS cluster analysis (right) displays two significant zones (CB-1 and CB-2) annotated by colour.

K is mainly considered a proxy for soil weathering (clays) and Ti an indicator of clastic sedimentation [253,254]. Typically, the Ti/K ratio provides an indication of the erosional influx driven by runoff from precipitation – with higher values indicating precipitation runoff [224,253,254]. Si/Ti is typically used as a biological proxy indicative siliceous productivity and exhibits a lagged response to event of nutrient deposition (i.e., rainfall, flooding, anthropogenic activity) [223]. P abundance correlated well with Si/Ti and other biogenic proxies and was regarded for this reason. It should be noted that post-depositional mobility of P is common and should be interpreted with caution [255]. CB-2 shifts of Ti/K and Mg aligned with Si/Ti and P; therefore, it was selected as an indicator of nutrients (Fig. 30). Within CB-1 P remained consistently lower than CB-2, where there is an observable increase and several spikes from the 15 cm (~1990) onward (Fig. 30). The increasing trend from CB-1 to -2 is matched by Si/Ti and Chl-a (Fig. 30). The increase in P loads is likely associated with increased primary productivity indicated by Si/Ti and Chl-a in CB-2. Ti/K displays an inverse trend to P and Si/Ti, decreasing from CB-1 to -2. The decrease in Ti/K can be attributed to less soil weathering in the lake catchment over time. Chl-a displays an increase in CB-1 that remains constant in CB-2 until ~2007 increasing again to ~2020 (Fig. 30). The increase in Chl-a aligns with Si/Ti and P trends indicating a increase in primary productivity of the Lake post ~1990 (Fig. 30). MCs only had one detection near the cores surface at the mid point of ~2019 (Fig. 30).

3.3.6. Dog Lake: Down Core Multiproxy Trends

The Dog Lake Itrax-XRF elemental data generated displayed an observable shift in Ti, Sr, Fe, K, Ca, CIR, S, and Si at depths of 25.5cm and 19.75cm (~1828 and ~1918). This observable shift is annotated in Figure 31. The CONISS cluster analysis of thirteen elements and the CIR (Fig. 31) provided two significant zones DG-1 and DG-2, the boundary of which aligns closely with the second shift in elemental counts (Fig. 31). Similar to Colonel By Lake, Mg undergoes constant fluctuation down-core is thought to be a result of annual variation In lake water level (Fig. 31). The

S (sulfur) displayed a drastic decrease in cps towards the cores surface with declines at both 26 cm and 19.5 cm (Fig. 31).

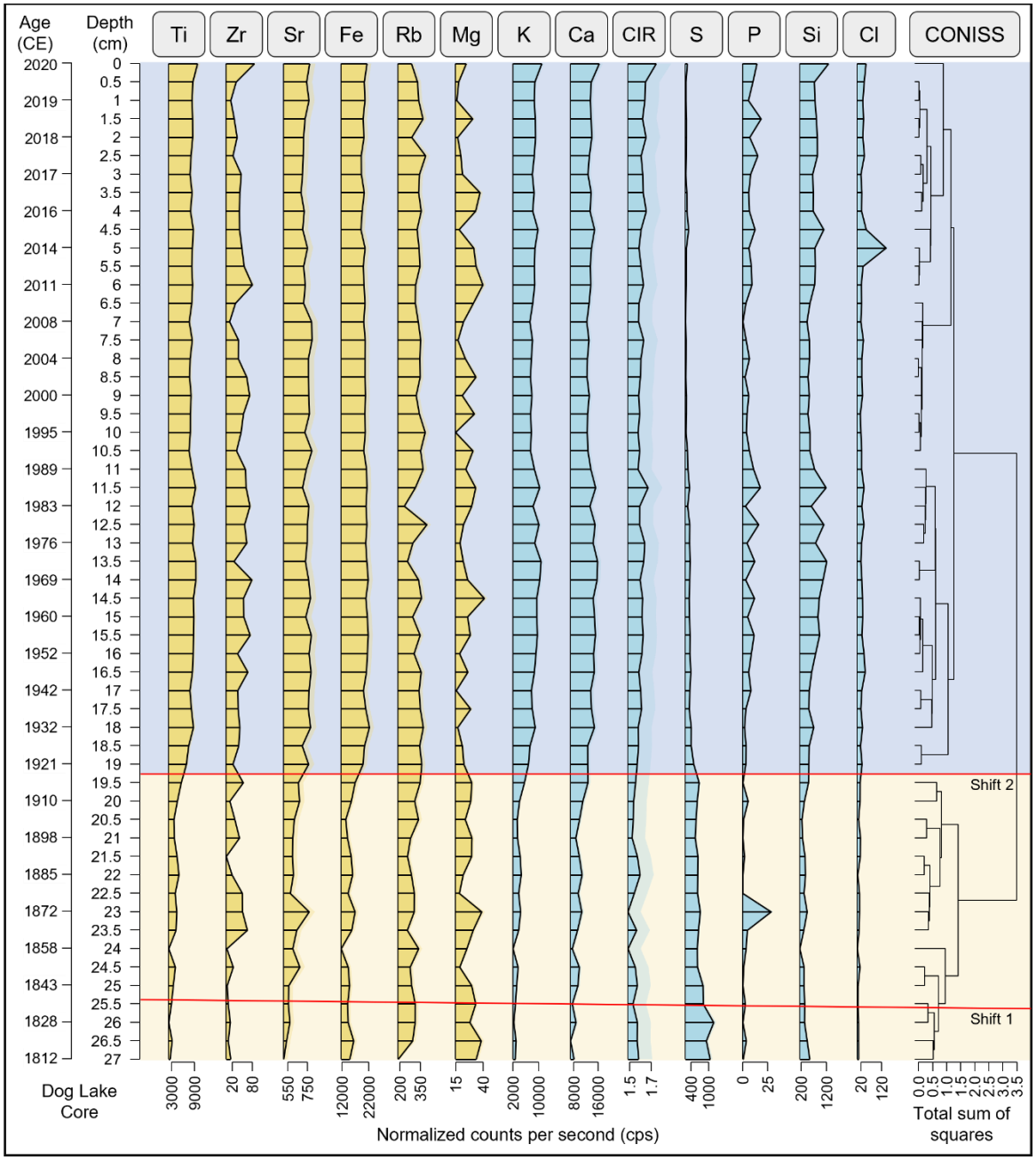

Figure 31. Dog Lake down-core stratigraphic profiles of Itrax-XRF elemental count data. CONISS cluster analysis (right) displays two significant zones (DG-1 and DG-2) annotated by colour. Shifts in elemental count data are annotated.

Prior to establishment of the Rideau Canal the deeper northern portions of Dog Lake were headwaters of the historical Cataraqui river [233,234]. The southern basin of Dog Lake existed as a

pre-canal wetland adjacent to the Cataraqui River to which it drained [233,234]. This is supported by trend observed with S, steeply decreasing from 27 cm – 25 cm (~1812 to 1835) and a gradual decline thereafter (Fig. 31). Freshwater wetlands comparative to Lake sediments, typically possess greater S content due to anoxic (reducing) conditions, microbial recycling, and organosulfur compounds [256,257]. Sulfides produced or deposited within wetland sediments can undergo oxidization to sulfate by biological or chemical reactions both of which can be microbially disproportioned. The production of elemental sulfur is substantially faster than its consumption, generally accumulating in wetland sediments [256,257]. The heightened S content at the bottom of the core and its abrupt dissipation, is indicative of the change to the pre-existing wetland system upon canal flooding. Overall, DG-1 is associated with system change due to canal flooding and DG-2 displays equilibration of the newly formed lake environment, the observed shifts in elemental counts align with boundaries of both these zones (Fig. 31).

To further assess the downcore trends in Dog Lake, seven elemental Itrax-XRF proxies (Si/Ti, Mg, Ca/SR, P, Ca/Ti, S, K/Ti) indicative of the canal system change were selected for CONISS cluster analysis with Chl-a and MC concentrations (Fig. 32). Ca/Ti ratio and Chl-a share this same trend as S, with high abundance at core bottom, a steep decline to 25 cm (~1840), a gradual decline to 19.5 cm (~1918) and continuing at a relatively constant values to the surface (Fig. 32). The Ca/Ti ratio is useful for assessing relative changes in biogenic versus lithogenic sedimentation and recording carbonate content, generally applied as a paleoproductivity proxy. Here, the low values of Ca/Ti localized deeper in the core reflect a disproportionate sedimentation in biogenic material (Fig. 32). Furthermore, Adaïmé et al. (2022) observed a high correlation between the Ca/Ti and organic matter content [225]. An increase in both biological sedimentation and high organic matter content is facilitated by productive, anoxic and hydraulically saturated conditions, representative of the pre-canal wetland. Furthermore, the presence of old carbon within ^{14}C dated core samples suggests an accumulation of organic matter and its resuspension and distribution – likely associated with the disturbance of canal flooding. These results support

historical records that document the presence of extensive wetlands in the area that is now Dog Lakes southern basin [233,234].

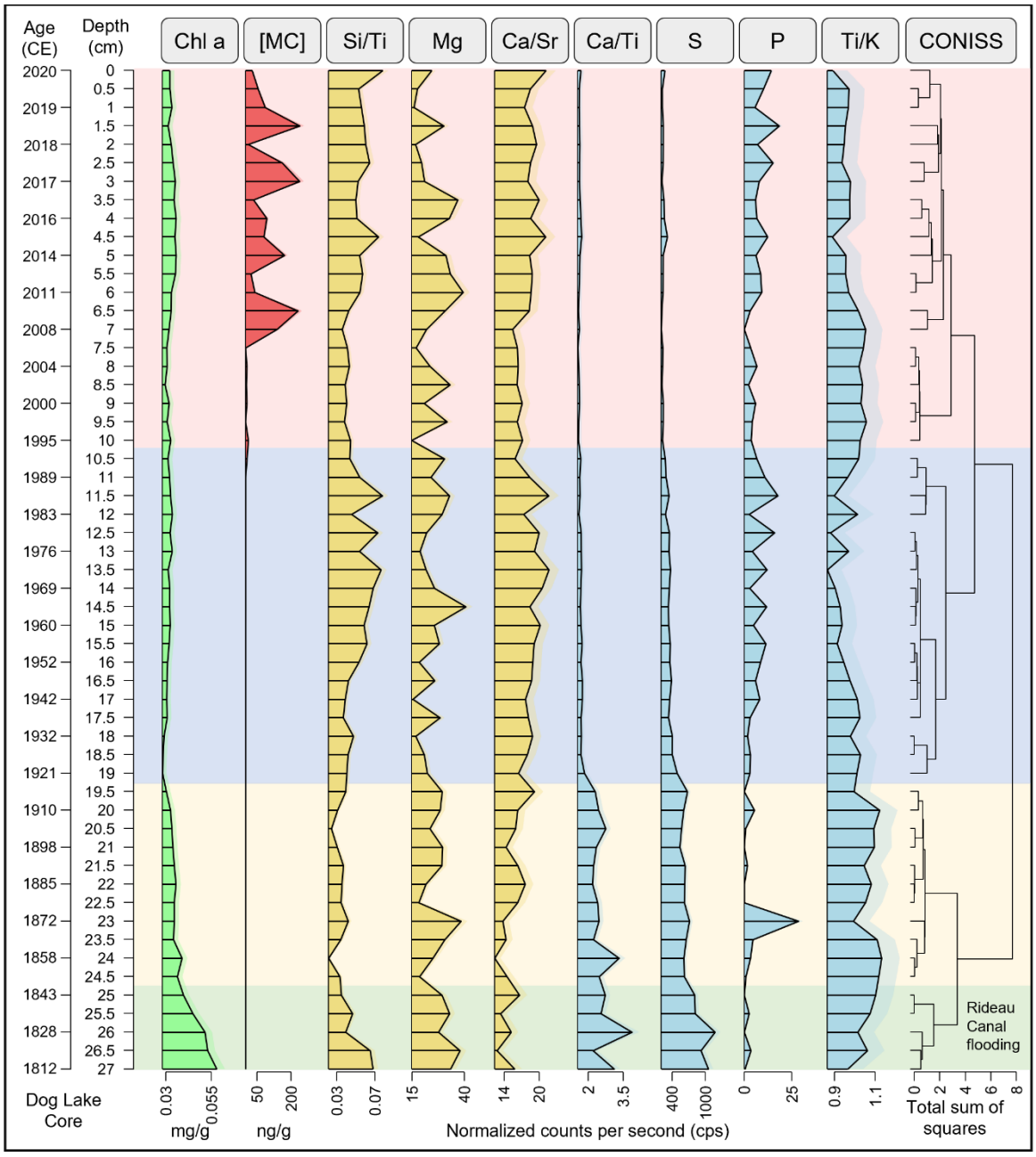

Figure 32. Dog Lake down-core stratigraphic profiles of Itrax-XRF elemental count data, Chl-a and MC concentrations. CONISS cluster analysis (right) displays four significant zones (DG-1A, DG-1B, DG-2A and DG-2B) annotated by colour.

Sr is fixed by calcifying organisms at same time as Ca and so can be used as marker of biogenic origin. The Ca/Sr ratio typically indicates Ca sourced from biogenic $CaCO_3$. Throughout the core the Ca/Sr ratio gradually increases to the cores surface, spiking at 12cm (Fig. 32). This trend is likely result of in-situ precipitation of carbonates induced by system flooding or to a lesser extent, calcifying organisms occupying the system. The lithogenic elements of Ti, Fe, Zr, Rb and Sr display an increase post canal flooding until 1930; thereafter reaching constant rate, continuing to the cores surface (Fig. 32). This same trend is observed with the biologically influenced elements of Ca, K, Si and P (Fig. 32). During pre-existing wetland conditions transport and deposition of these elements is thought to be non-uniform, reflecting lower values due to a lack of hydraulic connectivity and potential buffering from aquatic macrophytes and riparian vegetation. The creation of southern basin Dog Lake in 1831 is thought to be associated with the observable shift elemental counts as allochthonous deposition from overland runoff. Both the flooding of the basin and clear-cutting of the surrounding are feature that enhance overland runoff and increase allochthonous deposition to the newly formed Lake environment. The subsequent cluster analysis of Si/Ti, Mg, Ca/SR, P, Ca/Ti, S, K/Ti, MCs and Chl-a displayed four significant zones indicated by a broken stick model (DG-1A, DG-1B, DG-2A, DG-2B; Fig. 32). The four zones are each subdivision of the two original DG zones determined by CONISS analysis of elemental data (Fig. 32). The first zone DG-1A extends from the 27.0 to 24.75 cm (~1812 – 1850) and represents the period of canal construction, flooding (Fig. 32). In DG-1A values of S, Ca/Ti, Si/Ti and Chl-a decrease drastically throughout, with the steepest declines after the inferred date of canal construction (~1831; Fig. 32). Ti/K displays an inverse trend increasing drastically post canal-flooding (Fig. 32). P does not correlate with other trends but remains low throughout zone DG-1A (Fig. 32). DG-1B represents the disturbance to the systems post canal flooding and the subsequent system change as a result, extending from 24.75 cm – 19.25 cm (~1851 – 1918; Fig. 32). This zone represents the environmental disturbance brought by flooding and the subsequent of change conditions to reach new equilibrium. This is observed within the profiles of S, Ca/Ti and

Chl-a which display steady downward trends, steeply decreasing towards the DG-1B/DG-2A boundary (Fig. 32). Within DG-1B P displays two small spikes and one large spike but overall, remained low. Si/Ti and Ca/Sr both display an initial decrease then increase throughout DG-1B (Fig. 32). The Ti/K within this zone increases, remaining consistently high before tailing off at the upper boundary (Fig. 32). Here, the increase in Ti/K is matched by the dates of canal flooding providing indication of hydrological disturbance to the system (Fig. 32). DG-2A delineates the establishment of the newly formed lake environment, 19.25 cm – 10.25 cm (~1918 – 1994; Fig. 32). Within this zone the biological productivity proxies of Ca/Ti, S, Chl-a stabilize, at a near constant value. There is also a steady observable increase in Si/Ti and P within this zone (Fig. 32). The increase in P and Si/Ti aligns with a small increasing Chl-a, suggesting an overall increase in primary productivity throughout DG-2A (Fig. 32). Additionally, the lack of correlation between Ti/K and Si/Ti implies that increased productivity is derived from internal-loadings or anthropogenic origins. The heightened value of Ca/Sr suggests that Ca is derived from weathering in the catchment rather than in-situ precipitation. However, the lack of correlation of Ca/Sr with Ti/K suggests that Ca is mainly derived from in-situ precipitation of carbonate minerals. DG-2B denotes the most recent period of lake, occupying a depth of 10.25 – 0 cm (~1994 – 2020; Fig. 32). DG-2B possesses an observable increase in proxies of primary productivity (Chl-a, Si/Ti, P). Within DG-2B biological productivity proxies of Ca/Ti, S, and Chl-a possess similar values to the previous zone DG-2A (Fig. 32). The increase in Chl-a correlates with Si/Ti, P and Ca/Sr displaying noticeable trend beyond a depth of 7 cm. Interestingly, this aligns with MCs concentrations – detected in substantial concentration above depth of 7.5 cm (~2006). Ti/K in DG-2B increases initially declining at 4 cm depth increasing to the cores surface thereafter (Fig. 32). The lack of correlation between Ti/K, the Ca/Sr and other biological productivity proxies provides no clear indication that primary productivity is associated with loadings from precipitation runoff. Freshwater bloom forming cyanobacteria genre favor water temperatures above 20°C. Increased temperatures favor a shift in phytoplankton communities towards cyanobacteria dominance.

Therefore, it is suspected that increased nutrients at the cores surface and heightened surface water temperatures from climate change are driving the near surface occurrence of MCs. The role of temperature is further supported by the lack of MCs observed greater core depths.

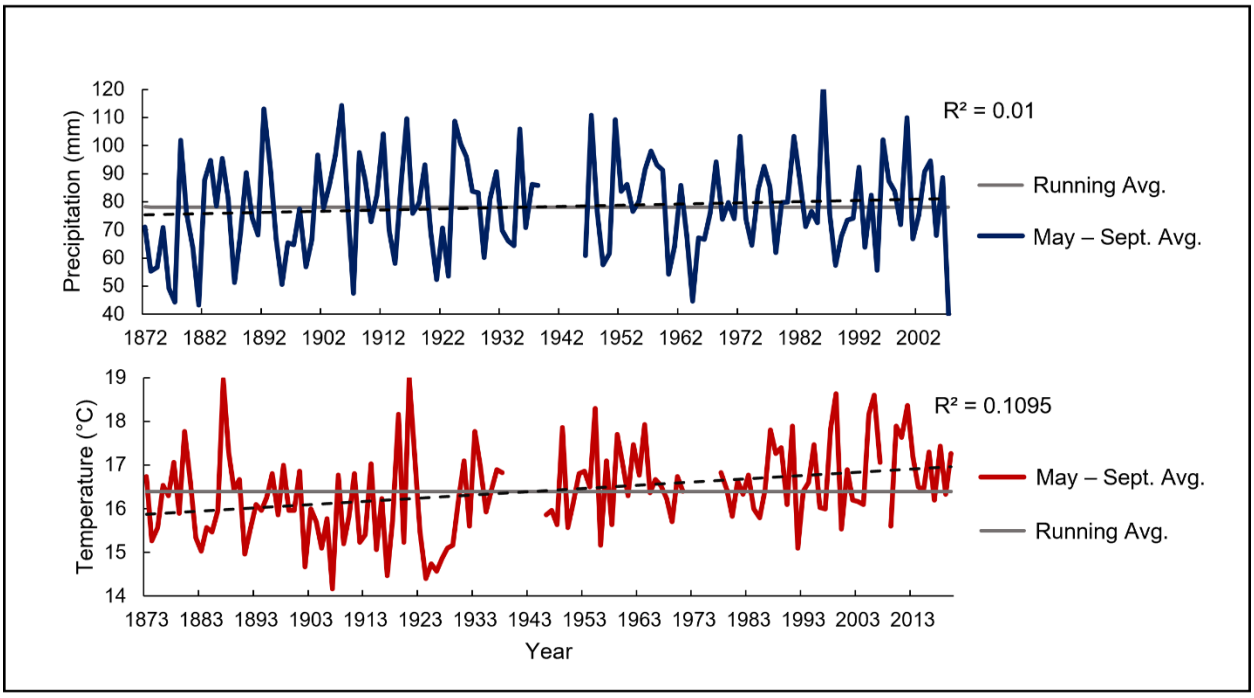

Figure 33. Monthly adjusted and homogenized precipitation and temperature data for the Kingston weather station.

Monthly adjusted and homogenized climate data was procured from the Canadian National Climate Data and Information Archive to evaluate the increasing trend in MC core profiles near core surface. The monthly precipitation and temperature data were averaged for May-September to encompass the seasonal period in which cyanobacteria blooms form. Both temperature and precipitation display weak increasing trend extending from 1800s to 2000s (Fig. 33). Temperature possesses a steeper increasing trend in comparison to precipitation data (Fig. 33).

3.4. DISCUSSION:

3.4.1. MC Method

The developed analytical method for the extraction and quantitation of MCs in sediments yielded exceptional recoveries and batch reproducibility with limited matrix effects. MC RR possessed the lowest R_E with a recovery of 71%, a result of its higher affinity for negatively charged sediment particles due to the molecules positively charged Arg groups. The %SSE for all analytes was between 93 – 110% implying minimal influence of matrix effects for the developed method. The selected solvent ratio proved to be effective at extracting the four selected MC congeners. The amphiphilic nature of 3:1 methanol and water renders it an optimal choice for MC congeners with varying polar and non-polar properties. The process of extraction used for this method is unique to previous studies in the application of sequential extractions with vortex, sonication and orbital shaker and is thought to potentially enhance MC recovery.

The optimized LC-(HR)MS method for MC provided accurate quantitation of values in the low ng/g d.w. (sediment dry weight) range – well within the range of environmentally relevant concentrations. However, this study predominantly focused on more polar MC congers and further validation is required for extension of the method to MCs of a non-polar nature (i.e., MC LF, MC LY, MC WR). The method achieved extraction and quantitation of MCs present in sediment core subsamples. The developed method was applied on two Rideau Lake (Dog Lake and Colonel By Lake) sediment cores in conjunction with Itrax-XRF and Chl-a analyses. Positive MC detections were observed for each core and the combination of MC profiles, geochemical counts and Chl-a were successful in inferring temporal variations of system conditions.

Colonel By provided a single MC detection below LOQ. The Colonel By core differs from Dog Lake as it was sampled in the Rideau Canal channel, passing directly through the lake, regulated by the locks at Kingston Mills. The lake's connectivity to the Rideau Canal system decreases the amount of water residence time and increases system flushing and surface flows. It is

hypothesized that these features and boat traffic reduce the deposition of MCs to the main channel of Colonel By Lake. This is supported by algal bloom reporting from MECP (Ministry of the Environment, Conservation and Parks) that has confirmed cHABs in Colonel By Lake for the years of 2010, 2011, 2012, 2016, 2017, 2020 and 2021. Dog Lake contained detectable MC concentrations above 11.0 cm (~1989) with elevated concentrations above 7.5 cm (~2006). This is supported by MCEP reports of cHAB occurrence in Dog Lake for the years of 2010, 2012, 2016, 2017, 2018, 2019, 2020, 2021. Below an approximate date of 2006 MC concentration in Dog Lake are ≤LOQ. The shallow depths and low surface velocities of Dog Lakes southern basin provides optimal conditions for deposition and storage of MCs. Future applications of this analytical method should select coring sites base on favorable conditions of MC deposition and storage. The lack of MC detection in deeper sediment layers requires further investigation, here it is postulated that Dog Lake did not historically experience MC producing blooms prior to the 1990s. To further validate the method and provide insight into potential in-situ degradation of MCs it is recommend that the method be applied to a lacustrine system with a document history of MC producing cHABs throughout the 1900s.

3.4.2. Paleolimnological Sediment Core Proxies

The results of the multi-proxy paleolimnological evaluation of two Rideau Lake system sediment cores provided significant stratigraphic zones indicative of environmental change. Colonel By Lake provided the two zones of CB-1 and CB-2 from a chronology spanning ~1892 – 2020 (Fig. 30). Dog Lake possessed four zones from a chronology spanning ~1812 – 2020, encompassing Rideau Canal construction (Fig. 32).

3.4.3. Colonel By Lake

The CRS ages generated for Colonel By Lake place the bottom of the core ~60 yrs. after Rideau Canal construction. No observably distinct down core trends were observed in the raw elemental

count data for Colonel By Lake. Two statistically significant zones were determined by cluster analysis (CB-1 and -2). The CB-1 and -2 boundary appeared to indicate shift denoted by an increase in biogenic proxies and a decrease in Ti/K (Fig. 30). The increase in biological activity within this zone displayed no correlation with runoff and catchment erosion (Ti/K). However, the increases in biogenic proxies aligned with that of P and therefore are potentially nutrient driven. Due to the lack of correlation with runoff, it is hypothesized that temperature may to play a role in driving this biogenic shift. Near the core surface, Chl-a and P increase and a decrease in Si/Ti is observed (Fig. 30). There is a single <LOQ detection of MCs near the core surface suggests the occurrence of cyanobacteria (Fig. 30). These trends potentially indicate a shift in the microalgae community towards non-siliceous cyanobacteria as Si/Ti indicates productivity associated with siliceous microalgae. This is further supported by the consistently higher values of Chl-a, P and temperature (Fig. 30 and 33). Recent cyanobacterial blooms throughout the 2010s were confirmed in Colonel By Lake by the MECP. However, there is not sufficient MC data to draw clear assertion of cyanobacteria dominance near the cores surface. The lack of detectable MCs in Colonel By Lake sediment is a likely result of the coring location – within the main channel of the Rideau Canal route (Fig. 23). Both the occurrence of blooms and the deposition of MCs is thought to be hindered by increased surface water flows and boat traffic at this location. Hence, future studies should optimize coring sites based on properties of bloom formation and MC deposition.

3.4.4. Dog Lake

The extrapolated CRS ages for the Dog Lake core provided a bottom date of ~1812, prior to canal construction. The second CONISS analysis displayed four statically significant zones that correspond, DG-1A: canal flooding, DG-1B: post canal disturbance, DG-2A: new lake equilibration and DG-2B: current lake conditions. DG-1A is indicative of the pre-canal wetland that existed prior to canal flooding that formed Dog Lake. This is denoted by higher values in the biological proxies

of Ca/Ti, Si/Ti, S, and Chl-a. Additionally, P is low within this zone indicative of nutrient deficient conditions associated with anoxic wetland soils. The biogenic proxies display rapid decreases upon the estimated date of canal flooding in DG-1A. Further indication of system disturbance is provided by a drastic increase in Ti/K that continues into DG-1B. Overall, DG-1A is defined by a rapid decrease in biological activity and increase in erosional influx that aligns with canal flooding from ~1812 to 1855. DG-1B possesses high runoff proxy values throughout and shows a continued gradual decrease in primary productivity proxies. The reductions in productivity and increase in catchment erosion influx are representative of the hydrological disturbance induced by canal flooding and the residual effect on the system. The gradual reduction in primary productivity is thought to largely be induced by habitat change caused by flooding – expressed by the planktic and macrophyte communities. Large hydraulic disturbances are known to alter species assemblages specifically in environmentally sensitive organisms (i.e., benthic invertebrates, phytoplankton and submergent macrophytes). Throughout ~1848 – 1915, it is suspected that flooding of the pre-canal wetland reduced resident macrophyte and phytoplankton communities and this is reflected in the biogenic proxies. The reduction in biogenic proxies aligns with a noticeable shift of heightened counts for the lithogenic elements of Fe, Ti, K and Ca. The increasing lithogenic trend associates with catchment weathering and can be attributed to the transition from a flooded wetland to a lake environment, as an increase in erosional flux over time. DG-2A identifies the establishment of the lake environment formed by canal flooding. This is best indicated by increases in biogenic proxies from DG-1B to -2A indicative of shift in biological conditions related to species assemblages – representative of a shift from wetland to lacustrine organisms. The Ti/K decreases rapidly in this zone suggesting decreases in precipitation-based influx. However, it is more likely a lagged response from flushing of clastic materials during canal-flooding. The most recent zone, DG-2B exhibits an increase in biogenic proxies and P above 7.5 cm depth (~2006). This trend does not appear to align with Ti/K (even with consideration of a lagged response) although, precipitation-based runoff does increase from DG-2A to DG-2B. The

lack of alignment between Ti/K and P suggest a nutrient influx from anthropogenic sources. Within DG-2A, MCs are detected and present in higher concentrations above 7.5 cm – sharing this trend with Chl-a. The increase in biogenic proxies and the presence of MCs above 7.5 cm indicates a potential increase in primary productivity and cyanobacteria abundance. The increase in P counts above this depth supports an increasing biological trend related to nutrients. It is hypothesized that temperature may play a role in cyanobacteria abundance as their occurrence is not coherent with erosional nutrient fluxes. Monthly temperature and precipitation data for the Kingston region display weak increasing trend from 1800s to the 2000s. Temperature displayed steeper trend from the 1800s to the 2010s. It is known that elevated water temperatures increase cyanobacteria growth rates, favoring their dominance relative to eukaryotic phytoplankton [18,258]. The results suggest that both elevated nutrients and warming temperatures near the cores surface have increased cyanobacteria growth rates and subsequently elevated MC concentrations. Overall, both primary productivity and cyanobacteria occurrence appears related to anthropogenic nutrient inputs.

3.5. CONCLUSION:

The developed analytical method for the extraction and quantitation of MCs in sediments with a NOD internal standard provided quantitation of MC LR, [Dha^7]MC LR, MC RR and MC LA to the low ng/g level. Sequential extraction of MCs by sonication, vortex and orbital shaker enabled effective recoveries of extracellular MCs and NOD in lake sediments. The LC-(HR)MS analytical method optimized for four MCs and NOD had negligible matrix effects and yielded accurate quantitation of select MCs in two Rideau Lake sediment cores. A greater abundance of MCs was detected in Dog Lake sediment than those of Colonel By. It is suspected that the core sampling locations largely contributed to the variation in MC profiles between Lakes. The efficacy of the method on more non-polar MC congeners requires further evaluation. Overall, the developed analytical method is well suited to quantitate MC LR, MC RR, MC LR and [Dha^7]MC LR in

freshwater sediment for the purpose paleolimnological investigations. Evaluation of historical conditions with geochemical proxies, Chl-a and MC concentration provided down-core trends for both systems. The multiproxy approach was effective in identifying major environmental shifts from canal flooding to current times, an approximate 200-year span. Dog Lake possessed four significant zones extending from pre-canal construction to current dates. DG-1A denoted pre-canal conditions and the disturbance of canal flooding in 1831. The second, DG-1B displayed the gradual change of environmental conditions upon flooding, equilibrating to the new lake environment. DG-2A denotes the dissipation of wetland conditions and establishment newly formed Lake environment with proxies reflecting lacustrine conditions. The final and most interesting zone, DG-2B displays an increase in biological productivity, nutrients, and MC concentrations. Colonel By Lake had two distinct zones delineated at ~1988 (15.75 cm). Post 1988 an increase in nutrient and biological productivities is observed. Although only a single MC detection was found for Colonel By, it was detected near the cores surface, similar to those of Dog Lake. The lack of down-core occurrence and correlation with erosional influx proxies suggested that MC (or cyanobacteria) occurrence is influenced by anthropogenic nutrient loading. Furthermore, climate warming appears to be a potential compounding factor in driving cyanobacteria occurrence. Average month temperatures from May to September for the study region provided an increasing trend from the 1800s. Therefore, it is suspected that the near surface trends observed with MCs is influenced by both anthropogenic nutrient loadings and elevated surface water temperatures.

REFRENCES

1. Huisman, J. *et al.* Cyanobacterial blooms. *Nat Rev Microbiol* **16**, 471–483 (2018).
2. Whitton, B. A. Diversity, Ecology, and Taxonomy of the Cyanobacteria. *Photosynthetic Prokaryotes* 1–51 (1992) doi:10.1007/978-1-4757-1332-9_1.

3. Carr, N. G. and Whitton, B. A. *The Biology of Cyanobacteria*. (Univ of California Press; Blackwell Scientific Publications, 1982).

4. Sánchez-Baracaldo, P. & Cardona, T. On the origin of oxygenic photosynthesis and Cyanobacteria. *New Phytologist* **225**, 1440–1446 (2020).

5. Vermaas, W. F. Photosynthesis and Respiration in Cyanobacteria. in *eLS* (Wiley, 2001). doi:10.1038/npg.els.0001670.

6. Oren, A. Cyanobacteria: biology, ecology and evolution. in *Cyanobacteria* 1–20 (John Wiley & Sons, Ltd, 2013). doi:10.1002/9781118402238.ch1.

7. Glazer, A. N. Structure and molecular organization of the photosynthetic accessory pigments of cyanobacteria and red algae. *Mol Cell Biochem* **18**, 125–140 (1977).

8. Stadnichuk, I. N., Krasilnikov, P. M. & Zlenko, D. v. Cyanobacterial phycobilisomes and phycobiliproteins. *Microbiology (Russian Federation)* Preprint at https://doi.org/10.1134/S0026261715020150 (2015).

9. Shively, J. M. *et al.* Intracellular Structures of Prokaryotes: Inclusions, Compartments and Assemblages. *Encyclopedia of Microbiology* 404–424 (2009) doi:10.1016/B978-012373944-5.00048-1.

10. Tomitani, A. *et al.* Chlorophyll b and phycobilins in the common ancestor of cyanobacteria and chloroplasts. *Nature 1999 400:6740* **400**, 159–162 (1999).

11. Gaysina, L. A., Saraf, A. & Singh, P. Cyanobacteria in Diverse Habitats. *Cyanobacteria: From Basic Science to Applications* 1–28 (2019) doi:10.1016/B978-0-12-814667-5.00001-5.

12. Martínez-Francés, E. & Escudero-Oñate, C. Cyanobacteria and Microalgae in the Production of Valuable Bioactive Compounds. *Microalgal Biotechnology* (2018) doi:10.5772/INTECHOPEN.74043.

13. Kultschar, B. & Llewellyn, C. Secondary Metabolites in Cyanobacteria. *Secondary Metabolites - Sources and Applications* (2018) doi:10.5772/INTECHOPEN.75648.

14. Bullerjahn, G. S. & Post, A. F. Physiology and molecular biology of aquatic cyanobacteria. *Front Microbiol* **5**, 359 (2014).

15. Planavsky, N. J. *et al.* Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. (2014) doi:10.1038/NGEO2122.

16. Schopf, T. J. M. Permo-Triassic Extinctions: Relation to Sea-Floor Spreading. *https://doi.org/10.1086/627955* **82**, 129–143 (2015).

17. Litchman, E. *et al.* Linking traits to species diversity and community structure in phytoplankton. *Fifty years after the '"Homage to Santa Rosalia"': Old and new paradigms on biodiversity in aquatic ecosystems* 15–28 (2010) doi:10.1007/978-90-481-9908-2_3.

18. Carey, C. C., Ibelings, B. W., Hoffmann, E. P., Hamilton, D. P. & Brookes, J. D. Eco-physiological adaptations that favour freshwater cyanobacteria in a changing climate. *Water Res* **46**, 1394–1407 (2012).

19. Jöhnk, K. D. *et al.* Summer heatwaves promote blooms of harmful cyanobacteria. *Glob Chang Biol* **14**, 495–512 (2008).

20. Kromkamp, J. New Zealand Journal of Marine and Freshwater Research Formation and functional significance of storage products in cyanobacteria Formation and functional significance of storage products in cyanobacteria. *N Z J Mar Freshwater Res* **21**, 457–465 (1987).

21. Paerl, H. W. & Otten, T. G. Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls. *Microb Ecol* **65**, 995–1010 (2013).

22. Kumar, K., Mella-Herrera, R. A. & Golden, J. W. Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol* **2**, a000315 (2010).

23. Zurawell, R. W., Chen, H., Burke, J. M. & Prepas, E. E. Hepatotoxic Cyanobacteria: A Review of the Biological Importance of Microcystins in Freshwater Environments. *http://dx.doi.org/10.1080/10937400590889412* **8**, 1–37 (2011).

24. Garg, R. & Maldener, I. The Formation of Spore-Like Akinetes: A Survival Strategy of Filamentous Cyanobacteria. *Microb Physiol* **31**, 296–305 (2021).

25. Wood, S. A., Jentzsch, K., Rueckert, A., Hamilton, D. P. & Cary, S. C. Hindcasting cyanobacterial communities in Lake Okaro with germination experiments and genetic analyses. *FEMS Microbiol Ecol* **67**, 252–260 (2009).

26. YAMAMOTO, Y. EFFECT OF SOME PHYSICAL AND CHEMICAL FACTORS ON THE GERMINATION OF AKINETES OF ANABAENA CYLINDRICA. *J Gen Appl Microbiol* **22**, 311–323 (1976).

27. Passarge, J., Hol, S., Escher, M. & Huisman, J. Competition for Light and Nutrients: Stable Coexistence, Alternative Stable States, or Competitive Exclusion? *Ecol Monogr* **76**, 57–72 (2006).

28. Huisman, J., Jonker, R. R., Zonneveld, C. & Weissing, F. J. Competition for Light Between Phytoplankton Species: Experimental Tests of Mechanistic Theory. *Ecology* **80**, 211–222 (1999).

29. Reynolds, C. S. *The Ecology of Phytoplankton*. (Cambridge University Press, 2006). doi:10.1017/CBO9780511542145.

30. Ibelings, B. W., Kroon, B. M. A. & Mur, L. R. Acclimation of photosystem II in a cyanobacterium and a eukaryotic green alga to high and fluctuating photosynthetic photon flux densities, simulating light regimes induced by mixing in lakes. *New Phytologist* **128**, 407–424 (1994).

31. Reynolds, C. S. Relationships among the biological properties, distribution, and regulation of production by planktonic cyanobacteria. *Toxicity Assessment* **4**, 229–255 (1989).

32. Pick, F. R. Blooming algae: a Canadian perspective on the rise of toxic cyanobacteria. *Canadian Journal of Fisheries and Aquatic Sciences* **73**, 1149–1158 (2016).

33. Michigan Sea Grant. Harmful Algal Blooms | Michigan Sea Grant. https://www.michiganseagrant.org/topics/coastal-hazards-and-safety/harmful-algal-blooms/ (2022).

34. Smith, D. R., King, K. W. & Williams, M. R. What is causing the harmful algal blooms in Lake Erie? *J Soil Water Conserv* **70**, 27A-29A (2015).

35. Makarewicz, J. C. Phytoplankton Biomass and Species Composition In Lake Erie, 1970 to 1987. *J Great Lakes Res* **19**, 258–274 (1993).

36. Conroy, J. D. *et al.* Temporal Trends in Lake Erie Plankton Biomass: Roles of External Phosphorus Loading and Dreissenid Mussels. *J Great Lakes Res* **31**, 89–110 (2005).

37. Moore, R. Lake Erie's Toxic Algae Breaks Records in 2015 . *NRDC* https://www.nrdc.org/experts/rob-moore/lake-eries-toxic-algae-breaks-records-2015 (2015).

38. Paerl, H. W. & Paul, V. J. Climate change: Links to global expansion of harmful cyanobacteria. *Water Res* **46**, 1349–1363 (2012).

39. Paerl, H. W. & Huisman, J. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. *Environ Microbiol Rep* **1**, 27–37 (2009).

40. Elliott, J. A. Is the future blue-green? A review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. *Water Res* **46**, 1364–1371 (2012).

41. Downing, J. A., Watson, S. B. & McCauley, E. Predicting Cyanobacteria dominance in lakes. *https://doi.org/10.1139/f01-143* **58**, 1905–1908 (2011).

42. Hauser, C. Algae Bloom in Lake Superior Raises Worries on Climate Change and Tourism. *The New York Times* (2018).

43. Sharon Oosthoek. Lake Superior Summer: Blue-green algal blooms come to a lake once believed immune. *Great Lakes Now* (2021).

44. Hilborn, E. D. & Beasley, V. R. One Health and Cyanobacteria in Freshwater Systems: Animal Illnesses and Deaths Are Sentinel Events for Human Health Risks. *Toxins 2015, Vol. 7, Pages 1374-1395* **7**, 1374–1395 (2015).

45. WHO. Cyanobacterial toxins: anatoxin-a and analogues Background document for development of. (2020).

46. BBC. Botswana: Mystery elephant deaths caused by cyanobacteria - BBC News. *BBC News: Africa* (2020).

47. Mosunova, O., Navarro-Muñoz, J. C. & Collemare, J. The biosynthesis of fungal secondary metabolites: From fundamentals to biotechnological applications. *Encyclopedia of Mycology* 458–476 (2021) doi:10.1016/B978-0-12-809633-8.21072-8.

48. Thirumurugan, D., Cholarajan, A., Raja, S. S. S. & Vijayakumar, R. An Introductory Chapter: Secondary Metabolites. *Secondary Metabolites - Sources and Applications* (2018) doi:10.5772/INTECHOPEN.79766.

49. Peregrín-Alvarez, J. M., Sanford, C. & Parkinson, J. The conservation and evolutionary modularity of metabolism. *Genome Biol* **10**, R63 (2009).

50. Sanchez, S. & Demain, A. L. Secondary Metabolites. *Comprehensive Biotechnology, Second Edition* **1**, 155–167 (2011).

51. Adrio, J. L. & Demain, A. L. Microbial Enzymes: Tools for Biotechnological Processes. *Biomolecules 2014, Vol. 4, Pages 117-139* **4**, 117–139 (2014).

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

53. Jones, M. R. *et al.* Comprehensive database of secondary metabolites from cyanobacteria. *bioRxiv* 2020.04.16.038703 (2020) doi:10.1101/2020.04.16.038703.

54. Welker, M. Cyanobacterial Hepatotoxins: Chemistry, Biosynthesis, and Occurrence. in *Seafood and Freshwater Toxins* (ed. Botana, L. M.) 843–862 (CRC Press, 2008). doi:10.1201/9781420007541-64.

55. Niina Leikoski. Cyanobactins - ribosomally synthesized and post-translationally modified peptides produced by cyanobacteria. (2013).

56. Zhang, Y., Chen, M., Bruner, S. D. & Ding, Y. Heterologous production of microbial ribosomally synthesized and post-translationally modified peptides. *Front Microbiol* **9**, 1801 (2018).

57. Kehr, J.-C., Picchi, D. G. & Dittmann, E. Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes. *Beilstein J. Org. Chem* **7**, 1622–1635 (2011).

58. Marahiel, M. A. Working outside the protein-synthesis rules: Insights into non-ribosomal peptide synthesis. *Journal of Peptide Science* **15**, 799–807 (2009).

59. Neilan, B. A. *et al.* Nonribosomal peptide synthesis and toxigenicity of cyanobacteria. *J Bacteriol* **181**, 4089–4097 (1999).

60. Strieker, M., Tanović, A. & Marahiel, M. A. Nonribosomal peptide synthetases: Structures and dynamics. *Curr Opin Struct Biol* **20**, 234–240 (2010).

61. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angewandte Chemie International Edition* **56**, 3770–3821 (2017).

62. Meyer, S. *et al.* Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. *Cell Chem Biol* **23**, 462–471 (2016).

63. Tillett, D. *et al.* Structural organization of microcystin biosynthesis in Microcystis aeruginosa PCC7806: an integrated peptide–polyketide synthetase system. *Chem Biol* **7**, 753–764 (2000).

64. Jenke-Kodama, H., Sandmann, A., Müller, R. & Dittmann, E. Evolutionary Implications of Bacterial Polyketide Synthases. doi:10.1093/molbev/msi193.

65. Chan, Y. A., Podevels, A. M., Kevany, B. M. & Thomas, M. G. Biosynthesis of Polyketide Synthase Extender Units. *Nat Prod Rep* **26**, 90 (2009).

66. Staunton, J. & Weissman, K. J. Polyketide biosynthesis: a millennium review. (2001) doi:10.1039/a909079g.

67. Pearson, L., Mihali, T., Moffitt, M., Kellmann, R. & Neilan, B. On the Chemistry, Toxicology and Genetics of the Cyanobacterial Toxins, Microcystin, Nodularin, Saxitoxin and Cylindrospermopsin. *Marine Drugs 2010, Vol. 8, Pages 1650-1680* **8**, 1650–1680 (2010).

68. Dittmann, E., Neilan, B., B. T. Molecular biology of peptide and polyketide biosynthesis in cyanobacteria. *Appl Microbiol Biotechnol* **57**, 467–473 (2001).

69. Ortega, M. A. & van der Donk, W. A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products. *Cell Chem Biol* **23**, 31 (2016).

70. Leikoski, N. *et al.* Genome Mining Expands the Chemical Diversity of the Cyanobactin Family to Include Highly Modified Linear Peptides. *Chem Biol* **20**, 1033–1043 (2013).

71. Zhang, Y. *et al.* A distributive peptide cyclase processes multiple microviridin core peptides within a single polypeptide substrate. *Nat Commun* **9**, 1780 (2018).

72. Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108–160 (2012).

73. Mattila, A. Expanding the genetic and chemical biosynthetic diversity of cyanobactin family of natural products. (University of Helsinki, 2020).

74. Le, T. & van der Donk, W. A. Mechanisms and Evolution of Diversity-Generating RiPP Biosynthesis. *Trends Chem* **3**, 266–278 (2021).

75. Hudson, G. A. & Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Curr Opin Microbiol* **45**, 61–69 (2018).

76. Martins, J. & Vasconcelos, V. Cyanobactins from Cyanobacteria: Current Genetic and Chemical State of Knowledge. *Mar Drugs* **13**, 6910 (2015).

77. Zong, W. *et al.* Molecular Mechanism for the Regulation of Microcystin Toxicity to Protein Phosphatase 1 by Glutathione Conjugation Pathway. *Biomed Res Int* **2017**, (2017).

78. Campos, A. & Vasconcelos, V. Molecular Mechanisms of Microcystin Toxicity in Animal Cells. *International Journal of Molecular Sciences 2010, Vol. 11, Pages 268-287* **11**, 268–287 (2010).

79. Metcalf, J. S. & Codd, G. A. Cyanotoxins. *Ecology of Cyanobacteria II: Their Diversity in Space and Time* **9789400738**, 651–675 (2012).

80. Christensen, V. G. & Khan, E. Freshwater neurotoxins and concerns for human, animal, and ecosystem health: A review of anatoxin-a and saxitoxin. *Science of The Total Environment* **736**, 139515 (2020).

81. WHO. *Toxic Cyanobacteria in Water*. (CRC Press, 2021). doi:10.1201/9781003081449.

82. Cox, P. A. BMAA, Neurodegeneration, and Neuroprotection. doi:10.1007/s12640-020-00303-z/Published.

83. Plato, Chris. C. *et al.* ALS and PDC of Guam. *Neurology* **58**, 765–773 (2002).

84. Plato, C. C. *et al.* Amyotrophic Lateral Sclerosis and Parkinsonism-Dementia Complex of Guam: Changing Incidence Rates during the Past 60 Years. *Am J Epidemiol* **157**, 149–157 (2003).

85. Vega, A., Bell, E. A. & Nunn, P. B. The preparation of l- and d-α-amino-β-methylaminopropionic acids and the identification of the compound isolated from Cycas circinalis as the l-isomer. *Phytochemistry* **7**, 1885–1887 (1968).

86. Spencer, P. S. *et al.* Guam Amyotrophic Lateral Sclerosis-Parkinsonism-Dementia Linked to a Plant Excitant Neurotoxin. *Science (1979)* **237**, 517–522 (1987).

87. Spencer, P. S. Hypothesis: Etiologic and Molecular Mechanistic Leads for Sporadic Neurodegenerative Diseases Based on Experience With Western Pacific ALS/PDC. *Front Neurol* **0**, 754 (2019).

88. Borenstein, A. R. *et al.* Cycad exposure and risk of dementia, MCI, and PDC in the Chamorro population of Guam. *Neurology* **68**, 1764–1771 (2007).

89. Bell, E. A. The discovery of BMAA, and examples of biomagnification and protein incorporation involving other non-protein amino acids. *http://dx.doi.org/10.3109/17482960903268700* **10**, 21–25 (2009).

90. Davis, D. A. *et al.* Cyanobacterial neurotoxin BMAA and brain pathology in stranded dolphins. *PLoS One* **14**, e0213346 (2019).

91. Glover, W. B., Baker, T. C., Murch, S. J. & Brown, P. N. Determination of β-N-methylamino-L-alanine, N-(2-aminoethyl)glycine, and 2,4-diaminobutyric acid in Food Products Containing Cyanobacteria by Ultra-Performance Liquid Chromatography and Tandem Mass Spectrometry: Single-Laboratory Validation. *J AOAC Int* **98**, 1559–1565 (2015).

92. Schneider, T., Simpson, C., Desai, P., Tucker, M. & Lobner, D. Neurotoxicity of isomers of the environmental toxin L-BMAA. *Toxicon* **184**, 175–179 (2020).

93. Ohtani, I., Moore, R. E. & Runnegar, M. T. C. Cylindrospermopsin: a potent hepatotoxin from the blue-green alga Cylindrospermopsis raciborskii. *J Am Chem Soc* **114**, 7941–7942 (1992).

94. Byth, S. Palm Island Mystery Disease. *Medical Journal of Australia* **2**, 40–42 (1980).

95. Kinnear, S. Cylindrospermopsin: A Decade of Progress on Bioaccumulation Research. *Marine Drugs 2010, Vol. 8, Pages 542-564* **8**, 542–564 (2010).

96. Norris, R. L. G. *et al.* Distribution of 14C cylindrospermopsin in vivo in the mouse. *Environ Toxicol* **16**, 498–505 (2001).

97. Yang, Y., Yu, G., Chen, Y., Jia, N. & Li, R. Four decades of progress in cylindrospermopsin research: The ins and outs of a potent cyanotoxin. *J Hazard Mater* **406**, 124653 (2021).

98. Humpage, A. R. & Falconer, I. R. Oral toxicity of the cyanobacterial toxin cylindrospermopsin in male Swiss albino mice: Determination of no observed adverse effect level for deriving a drinking water guideline value. *Environ Toxicol* **18**, 94–103 (2003).

99. Osswald, J., Rellán, S., Gago, A. & Vasconcelos, V. Toxicology and detection methods of the alkaloid neurotoxin produced by cyanobacteria, anatoxin-a. *Environ Int* **33**, 1070–1089 (2007).

100. Campbell, H. F., Edwards, O. E., D Ralph Kolt, A. N., F Campbell, R. Y. & Kolt, R. Synthesis of nor-anatoxin-a and anatoxin-a. *https://doi.org/10.1139/v77-190* **55**, 1372–1379 (1977).

101. Méjean, A., Paci, G., Gautier, V. & Ploux, O. Biosynthesis of anatoxin-a and analogues (anatoxins) in cyanobacteria. *Toxicon* **91**, 15–22 (2014).

102. Gorhamt, P. R. TOXIC WATERBLOOMS OF BLUE-GREEN ALGAE'. *Can Vet J* **1**, (1960).

103. Huber, C. S. The crystal structure and absolute configuration of 2,9-diacetyl-9-azabicyclo[4,2,1]non-2,3-ene. *Acta Crystallographica Section B* **28**, 2577–2582 (1972).

104. Carmichael, W. W., Biggs, D. F. & Gorham, P. R. Toxicology and Pharmacological Action of Anabaena flos-aquae Toxin. *Science (1979)* **187**, 542–544 (1975).

105. Puddick, J. *et al.* Acute toxicity of dihydroanatoxin-a from Microcoleus autumnalis in comparison to anatoxin-a. *Chemosphere* **263**, (2021).

106. Beach, D. G., Rafuse, C., Melanson, J. E. & McCarron, P. Rapid quantitative screening of cyanobacteria for production of anatoxins using direct analysis in real time high-resolution mass spectrometry. *Rapid Communications in Mass Spectrometry* **35**, e8940 (2021).

107. Kao, C. Y. Paralytic Shellfish Poisonings. in *Algal Toxins in Seafood and Drinking Water* (Academic Press Limited, 1993).

108. Sommer, H. The occurrence of the paralytic shell-fish poison in the common sand crab. *Science (1979)* **76**, 574–575 (1932).

109. Assun ão, J., Catarina Guedes, A. & Xavier Malcata, F. Biotechnological and pharmacological applications of biotoxins and other bioactive molecules from dinoflagellates. *Mar Drugs* **15**, (2017).

110. Wiese, M., D'Agostino, P. M., Mihali, T. K., Moffitt, M. C. & Neilan, B. A. Neurotoxic Alkaloids: Saxitoxin and Its Analogs. *Marine Drugs 2010, Vol. 8, Pages 2185-2211* **8**, 2185–2211 (2010).

111. Christensen, V. G. & Khan, E. Freshwater neurotoxins and concerns for human, animal, and ecosystem health: A review of anatoxin-a and saxitoxin. *Science of The Total Environment* **736**, 139515 (2020).

112. EFSA. Marine biotoxins in shellfish – Saxitoxin group. *EFSA Journal* **7**, (2009).

113. Bouaïcha, N. *et al.* Structural Diversity, Characterization and Toxicology of Microcystins. *Toxins (Basel)* **11**, 714 (2019).

114. Harke, M. J. *et al.* A review of the global ecology, genomics, and biogeography of the toxic cyanobacterium, Microcystis spp. *Harmful Algae* **54**, 4–20 (2016).

115. Díez-Quijada, L., Prieto, A. I., Guzmán-Guillén, R., Jos, A. & Cameán, A. M. Occurrence and toxicity of microcystin congeners other than MC-LR and MC-RR: A review. *Food and Chemical Toxicology* **125**, 106–132 (2019).

116. Catherine, A., Bernard, C., Spoof, L. & Bruno, M. Microcystins and Nodularins. *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis* 107–126 (2017) doi:10.1002/9781119068761.CH11.

117. Chen, L., Chen, J., Zhang, X. & Xie, P. A review of reproductive toxicity of microcystins. *J Hazard Mater* **301**, 381–399 (2016).

118. Dawson, R. M. the toxicology of microcystins. *Toxicon* **36**, 953–962 (1998).

119. WHO. *Cyanobacterial toxins: microcystins*. (2020).

120. Bláha, L., Babica, P. & Maršálek, B. Toxins produced in cyanobacterial water blooms – toxicity and risks. *Interdiscip Toxicol* **2**, 36 (2009).

121. Swingle, M., Ni, L. & Honkanen, R. E. Small Molecule Inhibitors of Ser/thr Protein Phosphatases: Specificity, Use and Common Forms of Abuse. *Methods Mol Biol* **365**, 23 (2007).

122. Mitchell, J. Ingested nitrate and nitrite and cyanobacterial peptide toxins. *IARC Monogr Eval Carcinog Risks Hum* **94**, 464 (2010).

123. Heinze, R. Toxicity of the Cyanobacterial Toxin Microcystin-LR to Rats after 28 Days Intake with the Drinking Water. *Environ Toxicol* **14**, (1999).

124. Health Canada. *Guidelines for Canadian Drinking Water Quality: Guideline Technical Document – Cyanobacterial Toxins*. https://publications.gc.ca/site/eng/9.831707/publication.html (2017).

125. Massey, I. Y. *et al.* A Mini-Review on Detection Methods of Microcystins. *Toxins 2020, Vol. 12, Page 641* **12**, 641 (2020).

126. Huang, I. S. & Zimba, P. v. Cyanobacterial bioactive metabolites—A review of their chemistry and biology. *Harmful Algae* **86**, 139–209 (2019).

127. Spoof, L., Błaszczyk, A., Meriluoto, J., Cegłowska, M. & Mazur-Marzec, H. Structures and Activity of New Anabaenopeptins Produced by Baltic Sea Cyanobacteria. *Marine Drugs 2016, Vol. 14, Page 8* **14**, 8 (2015).

128. Harada, K. ichi *et al.* Two cyclic peptides, anabaenopeptins, a third group of bioactive compounds from the cyanobacteriumAnabaena flos-aquae NRC 525-17. *Tetrahedron Lett* **36**, 1511–1514 (1995).

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

130. Murakami, M., Suzuki, S., Itou, Y., Kodani, S. & Ishida, K. New Anabaenopeptins, Potent Carboxypeptidase-A Inhibitors from the Cyanobacterium Aphanizomenon flos-aquae. *J. Nat. Prod* **63**, 24 (2000).

131. Monteiro, P. R., do Amaral, S. C., Siqueira, A. S., Xavier, L. P. & Santos, A. V. Anabaenopeptins: What we know so far. *Toxins (Basel)* **13**, 522 (2021).

132. Schreuder, H. *et al.* Isolation, Co-Crystallization and Structure-Based Characterization of Anabaenopeptins as Highly Potent Inhibitors of Activated Thrombin Activatable Fibrinolysis Inhibitor (TAFIa). *Scientific Reports 2016 6:1* **6**, 1–9 (2016).

133. Zafrir-Ilan, E. & Carmeli, S. Eight novel serine proteases inhibitors from a water bloom of the cyanobacterium Microcystis sp. *Tetrahedron* **66**, 9194–9202 (2010).

134. Roegner, A. *et al.* Combined Danio rerio embryo morbidity, mortality and photomotor response assay: A tool for developmental risk assessment from chronic cyanoHAB exposure. (2019) doi:10.1016/j.scitotenv.2019.134210.

135. Lenz, K. A., Miller, T. R. & Ma, H. Anabaenopeptins and cyanopeptolins induce systemic toxicity effects in a model organism the nematode Caenorhabditis elegans. *Chemosphere* **214**, 60–69 (2019).

136. Pawlik-Skowrońska, B. & Bownik, A. Synergistic toxicity of some cyanobacterial oligopeptides to physiological activities of Daphnia magna (Crustacea). *Toxicon* **206**, 74–84 (2022).

137. 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 (Aarau)* **76**, 133 (2022).

138. Beversdorf, L. J., Weirich, C. A., Bartlett, S. L. & Miller, T. R. Variable Cyanobacterial Toxin and Metabolite Profiles across Six Eutrophic Lakes of Differing Physiochemical Characteristics. *Toxins 2017, Vol. 9, Page 62* **9**, 62 (2017).

139. Elkobi-Peer, S. & Carmeli, S. New Prenylated Aeruginosin, Microphycin, Anabaenopeptin and Micropeptin Analogues from a Microcystis Bloom Material Collected in Kibbutz Kfar Blum, Israel. *Marine Drugs 2015, Vol. 13, Pages 2347-2375* **13**, 2347–2375 (2015).

140. Cornel, M. *et al.* CYANOPEPTOLINS, NEW DEPSIPEPTIDES FROM THE CYANOBACTERIUM Microcystis sp. PCC 7806. *J Antibiot (Tokyo)* **46**, 1550–1556 (1993).

141. McDonald, K. *et al.* Diagnostic Fragmentation Filtering for Cyanopeptolin Detection. *Environ Toxicol Chem* **40**, 1087–1097 (2021).

142. McDonald, K. Mass Spectrometry based Metabolomics to Decipher Strain Specific Microcystis Cyanopeptide Profiles. (2021) doi:10.22215/ETD/2021-14435.

143. Kust, A. *et al.* Insight into Unprecedented Diversity of Cyanopeptides in Eutrophic Ponds Using an MS/MS Networking Approach. *Toxins (Basel)* **12**, 561 (2020).

144. Gademann, K., Portmann, C., Blom, J. F., Zeder, M. & Jüttner, F. Multiple Toxin Production in the Cyanobacterium Microcystis : Isolation of the Toxic Protease Inhibitor Cyanopeptolin 1020. *J Nat Prod* **73**, 980–984 (2010).

145. Mazur-Marzec, H. *et al.* Cyanopeptolins with Trypsin and Chymotrypsin Inhibitory Activity from the Cyanobacterium Nostoc edaphicum CCNP1411. *Mar Drugs* **16**, 1–19 (2018).

146. Linington, R. G. *et al.* Symplocamide A, a potent cytotoxin and chymotrypsin inhibitor from the marine cyanobacterium Symploca sp. *J Nat Prod* **71**, 22–27 (2008).

147. 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. doi:10.1007/s11356-019-04543-1.

148. Czarnecki, O., Henning, M., Lippert, I. & Welker, M. Identification of peptide metabolites of Microcystis (Cyanobacteria) that inhibit trypsin-like activity in planktonic herbivorous Daphnia (Cladocera). *Environ Microbiol* **8**, 77–87 (2006).

149. Beversdorf, L. J. *et al.* Analysis of cyanobacterial metabolites in surface and raw drinking waters reveals more than microcystin. (2018) doi:10.1016/j.watres.2018.04.032.

150. Weckesser, J., Martin, C. & Jakobi, C. Cyanopeptolins, depsipeptides from cyanobacteria. *Syst Appl Microbiol* **19**, 133–138 (1996).

151. Linington, R. G. *et al.* Symplocamide A, a Potent Cytotoxin and Chymotrypsin Inhibitor from the Marine Cyanobacterium Symploca sp. doi:10.1021/np070280x.

152. Neumann, U. *et al.* Co-Occurrence of Non-toxic (Cyanopeptolin) and Toxic (Microcystin) Peptides in a Bloom of Microcystis sp. from a Chilean Lake. *Syst Appl Microbiol* **23**, 191–197 (2000).

153. 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 Lett* **57**, 1801–1803 (2016).

154. Chlipala, G. E., Mo, S. & Orjala, J. Chemodiversity in Freshwater and Terrestrial Cyanobacteria – a Source for Drug Discovery. *Curr Drug Targets* **12**, 1654 (2011).

155. Ishida, K., Kato, T., Murakami, M., Watanabe, M. & Watanabe, M. F. Microginins, Zinc Metalloproteases Inhibitors from the Cyanobacterium Microcystis aeruginosa. *Tetrahedron* **56**, 8643–8656 (2000).

156. Bober, B. & Bialczyk, J. Determination of the toxicity of the freshwater cyanobacterium Woronichinia naegeliana (Unger) Elenkin. *J Appl Phycol* **29**, 1355–1362 (2017).

157. Fernandes, K. *et al.* Toxicity of Cyanopeptides from Two Microcystis Strains on Larval Development of Astyanax altiparanae. *Toxins (Basel)* **11**, 220 (2019).

158. Neumann, U., Forchert, A., Flury, T. & Weckesser, J. Microginin FR1, a linear peptide from a water bloom of Microcystis species. *FEMS Microbiol Lett* **153**, 475–478 (1997).

159. Ishida, K., Matsuda, H. & Murakami, M. Four new microginins, linear peptides from the cyanobacterium Microcystis aeruginosa. *Tetrahedron* **54**, 13475–13484 (1998).

160. Ferreira, G. M. *et al.* Inhibition of Porcine Aminopeptidase M (pAMP) by the Pentapeptide Microginins. *Molecules 2019, Vol. 24, Page 4369* **24**, 4369 (2019).

161. Nath Bagchi, S., Sondhia, S., Kumar Agrawal, M. & Banerjee, S. An angiotensin-converting enzyme-inhibitory metabolite with partial structure of microginin in a cyanobacterium Anabaena fertilissima CCC597, producing fibrinolytic protease. doi:10.1007/s10811-015-0541-5.

162. Paiva, F. C. R., Ferreira, G. M., Trossini, G. H. G. & Pinto, E. Identification, In Vitro Testing and Molecular Docking Studies of Microginins' Mechanism of Angiotensin-Converting Enzyme Inhibition. *Molecules* **22**, (2017).

163. Portmann, C. *et al.* Isolation of aerucyclamides C and D and structure revision of microcyclamide 7806A: Heterocyclic ribosomal peptides from Microcystis aeruginosa PCC 7806 and their antiparasite evaluation. *J Nat Prod* **71**, 1891–1896 (2008).

164. Welker, M., Maršálek, B., Šejnohová, L. & von Döhren, H. Detection and identification of oligopeptides in Microcystis (cyanobacteria) colonies: Toward an understanding of metabolic diversity. *Peptides (N.Y.)* **27**, 2090–2103 (2006).

165. Ishida, K. *et al.* Plasticity and Evolution of Aeruginosin Biosynthesis in Cyanobacteria. *Appl Environ Microbiol* **75**, 2017 (2009).

166. Kohler, E. *et al.* The toxicity and enzyme activity of a chlorine and sulfate containing aeruginosin isolated from a non-microcystin-producing Planktothrix strain. *Harmful Algae* **39**, 154–160 (2014).

167. Ersmark, K., del Valle, J. R. & Hanessian, S. Chemistry and Biology of the Aeruginosin Family of Serine Protease Inhibitors. *Angewandte Chemie International Edition* **47**, 1202–1223 (2008).

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

169. Hasan-Amer, R., Carmeli, S., Raymond, I. D. & Sackler, B. Inhibitors of Serine Proteases from a Microcystis sp. Bloom Material Collected from Timurim Reservoir, Israel. *Marine Drugs 2017, Vol. 15, Page 371* **15**, 371 (2017).

170. Radau, G. & Schermuly, S. New Cyanopeptide-Derived Low Molecular Weight Inhibitors of trypsin-like Serine Proteases. *Arch. Pharm. Pharm. Med. Chem* **336**, 300–309 (2003).

171. Ahmed, M. N. *et al.* Potent Inhibitor of Human Trypsins from the Aeruginosin Family of Natural Products. *Cite This: ACS Chem. Biol* **16**, 2537–2546 (2021).

172. Anas, A. R. J. *et al.* FVIIa-sTF and Thrombin Inhibitory Activities of Compounds Isolated from Microcystis aeruginosa K-139. *Marine Drugs 2017, Vol. 15, Page 275* **15**, 275 (2017).

173. do Amaral, S. C. *et al.* Current Knowledge on Microviridin from Cyanobacteria. *Marine Drugs 2021, Vol. 19, Page 17* **19**, 17 (2021).

174. Reshef, V. & Carmeli, S. New microviridins from a water bloom of the cyanobacterium Microcystis aeruginosa. *Tetrahedron* **62**, 7361–7369 (2006).

175. Ishitsuka, M. O., Kusumi, T., Kakisawa, H., Kaya, K. & Watanabe, M. M. Microviridin: A Novel Tricyclic Depsipeptide from the Toxic Cyanobacterium Microcystis viridis. *J Am Chem Soc* **112**, 8180–8182 (1990).

176. Ziemert, N., Ishida, K., Weiz, A., Hertweck, C. & Dittmann, E. Exploiting the Natural Diversity of Microviridin Gene Clusters for Discovery of Novel Tricyclic Depsipeptides. *Appl Environ Microbiol* **76**, 3568–3574 (2010).

177. Rohrlack, T. *et al.* Isolation, Characterization, and Quantitative Analysis of Microviridin J, a New Microcystis Metabolite Toxic to Daphnia. *Journal of Chemical Ecology 2003 29:8* **29**, 1757–1770 (2003).

178. Glish, G. L. & Vachet, R. W. THE BASICS OF MASS SPECTROMETRY IN THE TWENTY-FIRST CENTURY. **2**, (2003).

179. Harris, G. A., Nyadong, L. & Fernandez, F. M. Recent developments in ambient ionization techniques for analytical mass spectrometry. *Analyst* **133**, 1297–1301 (2008).

180. Nagana Gowda, G. A. & Djukovic, D. Overview of mass spectrometry-based metabolomics: Opportunities and challenges. *Methods in Molecular Biology* **1198**, 3–12 (2014).

181. Clish, C. B. Metabolomics: an emerging but powerful tool for precision medicine. *Cold Spring Harb Mol Case Stud* **1**, a000588 (2015).

182. Alvarez-Rivera, G., Ballesteros-Vivas, D., Parada-Alfonso, F., Ibañez, E. & Cifuentes, A. Recent applications of high resolution mass spectrometry for the characterization of plant natural products. *TrAC - Trends in Analytical Chemistry* **112**, 87–101 (2019).

183. Eliuk, S. & Makarov, A. Evolution of Orbitrap Mass Spectrometry Instrumentation. (2015) doi:10.1146/annurev-anchem-071114-040325.

184. Stock, N. L. Introducing Graduate Students to High-Resolution Mass Spectrometry (HRMS) Using a Hands-On Approach. *J. Chem. Educ* **94**, (2017).

185. Sleno, L. The use of mass defect in modern mass spectrometry. *Journal of Mass Spectrometry* **47**, 226–236 (2012).

186. Dettmer, K., Aronov, P. A. & Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom Rev* **26**, 51–78 (2007).

187. Liu, X. & Locasale, J. W. Metabolomics: A Primer. *Trends Biochem Sci* **42**, 274–284 (2017).

188. Bundy, J. G., Matthew, A. E., Davey, P. & Viant, M. R. Environmental metabolomics: a critical review and future perspectives. doi:10.1007/s11306-008-0152-0.

189. Beger, R. D. A Review of Applications of Metabolomics in Cancer. *Metabolites 2013, Vol. 3, Pages 552-574* **3**, 552–574 (2013).

190. Tiam, S. K. *et al.* Insights into the Diversity of Secondary Metabolites of Planktothrix Using a Biphasic Approach Combining Global Genomics and Metabolomics. *Toxins 2019, Vol. 11, Page 498* **11**, 498 (2019).

191. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2**, 559–572 (1901).

192. Hartmann, H. C. Climate change impacts on Laurentian Great Lakes levels. *Clim Change* **17**, 49–67 (1990).

193. GC. Great Lakes protection - Canada.ca. *Government of Canada* https://www.canada.ca/en/environment-climate-change/services/great-lakes-protection.html (2022).

194. GC. Water sources: lakes - Canada.ca. *Government of Canada* https://www.canada.ca/en/environment-climate-change/services/water-overview/sources/lakes.html (2022).

195. Mahdiyan, O., Filazzola, A., Molot, L. A., Gray, D. & Sharma, S. Drivers of water quality changes within the Laurentian Great Lakes region over the past 40 years. *Limnol Oceanogr* **66**, 237–254 (2021).

196. Adrian, R. *et al.* Lakes as sentinels of climate change. *Limnol Oceanogr* **54**, 2283–2297 (2009).

197. Huot, Y. *et al.* The NSERC Canadian Lake Pulse Network: A national assessment of lake health providing science for water management in a changing climate. *Science of the Total Environment* **695**, (2019).

198. Goverenment of Canada. Water and the environment - Canada.ca. *Environment and Natural Resources Canada* https://www.canada.ca/en/environment-climate-change/services/water-overview.html (2022).

199. Bernard, C. *et al.* Appendix 2: Cyanobacteria Associated With the Production of Cyanotoxins. *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis* 501–525 (2017) doi:10.1002/9781119068761.APP2.

200. Winnikoff, J. R., Glukhov, E., Watrous, J., Dorrestein, P. C. & Gerwick, W. H. Quantitative molecular networking to profile marine cyanobacterial metabolomes. *J Antibiot (Tokyo)* **67**, 105–112 (2014).

201. Thermo Fisher Scientific. Xcalibur™. Preprint at https://www.thermofisher.com/order/catalog/product/OPTON-30965 (2013).

202. Blackwell, B. A. *et al.* Assessment of deoxynivalenol and deoxynivalenol derivatives in Fusarium graminearum-inoculated Canadian maize inbreds. *Canadian Journal of Plant Pathology* (2022) doi:10.1080/07060661.2022.2032830/SUPPL_FILE/TCJP_A_2032830_SM2797.XLSX.

203. Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching and identification. *Anal Chem* **78**, 779–787 (2006).

204. Gift, N., Gormley, I. C. & Brennan, L. MetabolAnalyze: Probabilistic Latent Variable Models for Metabolomic Data. Preprint at https://cran.r-project.org/package=MetabolAnalyze (2019).

205. Lê, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J Stat Softw* **25**, 1–18 (2008).

206. Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 1–11 (2010).

207. Wang, M. *et al.* Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **34**, 828–837 (2016).

208. Zervou, S. K., Gkelis, S., Kaloudis, T., Hiskia, A. & Mazur-Marzec, H. New microginins from cyanobacteria of Greek freshwaters. *Chemosphere* **248**, (2020).

209. Chen, W., Li, L., Gan, N. & Song, L. Optimization of an effective extraction procedure for the analysis of microcystins in soils and lake sediments. *Environmental Pollution* **143**, 241–246 (2006).

210. Zastepa, A., Pick, F. R. & Blais, J. M. Fate and Persistence of Particulate and Dissolved Microcystin-LA from Microcystis Blooms. *Human and Ecological Risk Assessment: An International Journal* **20**, 1670–1686 (2014).

211. Butler, D. Blue-green algae in Upper Rideau Lake contain toxins, tests confirm. *Ottawa Citizen* 1 (2014).

212. Bell, K. REHABILITATING LAC HENEY. *Association for the Protection of Lac Heney Watershed* 1 https://bvlacheney.ca/history/articles/rehabilitation/ (2018).

213. Community. Lac Heney Water Quality. *Gatineau Lakes* 1 https://gatineaulakes.com/index.php/2017/06/28/101/ (2022).

214. Blue-green algae toxins found in Upper Rideau Lake | CBC News. https://www.cbc.ca/news/canada/ottawa/blue-green-algae-toxins-found-in-upper-rideau-lake-1.2736630.

215. Taranu, Z. E., Pick, F. R., Creed, I. F., Zastepa, A. & Watson, S. B. Meteorological and Nutrient Conditions Influence Microcystin Congeners in Freshwaters. *Toxins 2019, Vol. 11, Page 620* **11**, 620 (2019).

216. Dyble, J., Fahnenstiel, G. L., Litaker, R. W., Millie, D. F. & Tester, P. A. Microcystin concentrations and genetic diversity of Microcystis in the lower great lakes. *Environ Toxicol* **23**, 507–516 (2008).

217. Richardson, J. *et al.* Response of cyanobacteria and phytoplankton abundance to warming, extreme rainfall events and nutrient enrichment. *Glob Chang Biol* **25**, 3365–3380 (2019).

218. Pal, S., Gregory-Eaves, I. & Pick, F. R. Temporal trends in cyanobacteria revealed through DNA and pigment analyses of temperate lake sediment cores. *J Paleolimnol* **54**, 87–101 (2015).

219. Thesis, P. & Reuss, N. Sediment pigments as biomarkers of environmental change.

220. Zastepa, A., Pick, F. R., Blais, J. M. & Saleem, A. Analysis of intracellular and extracellular microcystin variants in sediments and pore waters by accelerated solvent extraction and high performance liquid chromatography-tandem mass spectrometry First analytical method for intracel-lular microcystins (MCs) in sediment. Includes a suite of variants (LR, 7dm. doi:10.1016/j.aca.2015.02.056.

221. Forrest, F., Reavie, E. D. & Smol, J. P. Comparing limnological changes associated with 19th century canal construction and other catchment disturbances in four lakes within the Rideau Canal system, Ontario, Canada. *J Limnol* **61**, 183–197 (2003).

222. Sonnenburg, E. P., Boyce, J. I. & Reinhardt, E. G. Multi-proxy paleoenvironmental record of colonial land-use change in the lower Rideau Canal system (Colonel By Lake), Ontario, Canada. *J Paleolimnol* **42**, 515–532 (2009).

223. *Micro-XRF Studies of Sediment Cores*. vol. 17 (Springer Netherlands, 2015).

224. Kylander, M. E., Ampel, L., Wohlfarth, B. & Veres, D. High-resolution X-ray fluorescence core scanning analysis of Les Echets (France) sedimentary sequence: New insights from chemical proxies. *J Quat Sci* **26**, 109–117 (2011).

225. Adaïmé, M. É., Pienitz, R., Legendre, P. & Antoniades, D. Postglacial palaeoenvironmental reconstruction of the Fury and Hecla Strait region (Nunavut) inferred from microfossils and geochemical proxies. *J Quat Sci* **37**, 944–958 (2022).

226. Richter, T. O. *et al.* The Avaatech XRF Core Scanner: technical description and applications to NE Atlantic sediments. *Geological Society, London, Special Publications* **267**, 39–50 (2006).

227. Croudace, I. W., Rindby, A. & Rothwell, R. G. ITRAX: Description and evaluation of a new multi-function X-ray core scanner. *Geol Soc Spec Publ* **267**, 51–63 (2006).

228. Gregory, B. R. B. Understanding the impact of millennial to sub-decadal climate and limnological change on the stability of arsenic in lacustrine sediments. (Carleton University, 2019).

229. Sakamoto, T. *et al.* Non-Destructive X-Ray Fluorescence (XRF) Core-Imaging Scanner, TATSCAN-F2. *Scientific Drilling* **2**, 37–39 (2006).

230. CRCA. *Dog Lake lake fact sheet*. https://www.crca.ca/wp-content/uploads/PDFs/LakeReports/2017-FactSheet-DogLake.pdf (2017).

231. CRCA. *Colonel By Lake lake fact sheet*. https://www.crca.ca/wp-content/uploads/PDFs/LakeReports/2017-FactSheet-ColonelByLake.pdf (2017).

232. Tian, H. (Allen). Assessing the Biotic and Abiotic Factors Influencing the Distribution and Intensity of Cyanobacterial Harmful Algal Blooms in Dog Lake, South Frontenac Using Uav Imaging and eDna - ProQuest. (Queens University , 2021).

233. Watson, K. *Engineered landscapes: the Rideau Canals' transformation of a wilderness waterway.* (2006).

234. Watson, K. *The Rideau route: exploring the pre-canal waterway*. (2007).

235. Tillmanns, A. R., Pick, F. R. & Aranda-Rodriguez, R. Sampling and analysis of microcystins: Implications for the development of standardized methods. *Environ Toxicol* **22**, 132–143 (2007).

236. Crann, C. A. *et al.* First Status Report on Radiocarbon Sample Preparation Techniques at the A.E. Lalonde AMS Laboratory (Ottawa, Canada). *Radiocarbon* **59**, 695–704 (2017).

237. Blaaw, M. & Chrsten, J. A. Age-Depth Modelling using Bayesian Statistics. Preprint at https://orcid.org/0000-0002-5680-1515 (2022).

238. Reimer, P. J. *et al.* IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0–50,000 Years cal BP. *Radiocarbon* **55**, 1869–1887 (2013).

239. Appleby, P. G. & Oldfieldz, F. The assessment of 210Pb data from sites with varying sediment accumulation rates. *Hydrobiologia 1983 103:1* **103**, 29–35 (1983).

240. Appleby, P. G. & Oldfield, F. The calculation of lead-210 dates assuming a constant rate of supply of unsupported 210Pb to the sediment. *Catena (Amst)* **5**, 1–8 (1978).

241. Baud, A. *et al.* A framework for 210 Pb model selection and its application to 37 cores from Eastern Canada to identify the dynamics and drivers of lake sedimentation rates. *Earth Surf Process Landf* **47**, 2518–2530 (2022).

242. Das, B., Vinebrooke, R. D., Sanchez-Azofeifa, A., Rivard, B. & Wolfe, A. P. Inferring sedimentary chlorophyll concentrations with reflectance spectroscopy: a novel approach to reconstructing historical changes in the trophic status of mountain lakes. *https://doi.org/10.1139/f05-016* **62**, 1067–1078 (2011).

243. Wolfe, A. P., Vinebrooke, R. D., Michelutti, N., Rivard, B. & Das, B. Experimental calibration of lake-sediment spectral reflectance to chlorophyll a concentrations: methodology and paleolimnological validation. *J Paleolimnol* **36**, 91–100 (2006).

244. Vinebrooke, R. D. & Leavitt, P. R. Phytobenthos and Phytoplankton as Potential Indicators of Climate Change in Mountain Lakes and Ponds: A HPLC-Based Pigment Approach. *J North Am Benthol Soc* **18**, 15–33 (1999).

245. Oksanen, J. ‘vegan’ Community Ecology Package. Preprint at (2022).

246. Juggins, S. ‘rioja’ Title Analysis of Quaternary Science Data. Preprint at http://www.staff.ncl.ac.uk/stephen.juggins/, (2020).

247. Gregory, B. R. B. *et al.* Sequential sample reservoirs for Itrax-XRF analysis of discrete samples. *J Paleolimnol* **57**, 287–293 (2017).

248. Gregory, B. R. B., Patterson, R. T., Reinhardt, E. G., Galloway, J. M. & Roe, H. M. An evaluation of methodologies for calibrating Itrax X-ray fluorescence counts with ICP-MS concentration data for discrete sediment samples. (2019) doi:10.1016/j.chemgeo.2019.05.008.

249. Chen, W., Li, L., Gan, N. & Song, L. Optimization of an effective extraction procedure for the analysis of microcystins in soils and lake sediments. doi:10.1016/j.envpol.2005.11.030.

250. Wu, X. *et al.* Optimal strategies for determination of free/extractable and total microcystins in lake sediment. *Anal Chim Acta* **709**, 66–72 (2012).

251. Roberts, N. *et al.* A tale of two lakes: a multi-proxy comparison of Lateglacial and Holocene environmental change in Cappadocia, Turkey. *J Quat Sci* **31**, 348–362 (2016).

252. Kober, B., Schwalb, A., Schettler, G. & Wessels, M. Constraints on paleowater dissolved loads and on catchment weathering over the past 16 ka from 87Sr/86Sr ratios and Ca/Mg/Sr chemistry of freshwater ostracode tests in sediments of Lake Constance, Central Europe. *Chem Geol* **240**, 361–376 (2007).

253. Evans, G. *et al.* A multi-proxy paleoenvironmental interpretation spanning the last glacial cycle (ca. 117 ± 8.5 ka BP) from a lake sediment stratigraphy from Lake Kai Iwi, Northland, New Zealand. *J Paleolimnol* **65**, 101–122 (2021).

254. Marshall, M. H. *et al.* Late Pleistocene and Holocene drought events at Lake Tana, the source of the Blue Nile. (2011) doi:10.1016/j.gloplacha.2011.06.004.

255. Ostrofsky, M. L. Differential post-depositional mobility of phosphorus species in lake sediments. *J Paleolimnol* **48**, 559–569 (2012).

256. Pester, M., Knorr, K. H., Friedrich, M. W., Wagner, M. & Loy, A. Sulfate-reducing microorganisms in wetlands - fameless actors in carbon cycling and climate change. *Front Microbiol* **3**, 72 (2012).

257. Guo, W., Cecchetti, A. R., Wen, Y., Zhou, Q. & Sedlak, D. L. Sulfur Cycle in a Wetland Microcosm: Extended 34S-Stable Isotope Analysis and Mass Balance. *Environ Sci Technol* **54**, 5498–5508 (2020).

258. Lürling, M., Mello, M. M., van Oosterhout, F., Domis, L. de S. & Marinho, M. M. Response of natural cyanobacteria and algae assemblages to a nutrient pulse and elevated temperature. *Front Microbiol* **9**, 1851 (2018).

4. Supplementary information

Table S1. Concentrations (µg/g d.w.) of MCs and semi-quantitated CNPs with the greatest occurrence and abundance

No.	Watercourse	Date	MC RR	MC LR	[Dha^{7}] MC LR	MC WR	MC LA	MC LY	[D-Asp^{3}] MC LY	CP 1143	MP MZ845	CP S	CP 954	FA A	AP G_{Itou}	AP MM913	AP 915	MG 532
5	Bagot Long Lake	2016-11-16	–	–	<1	–	–	–	–	–	10	–	–	2	–	–	–	–
6	Bagot Long Lake	2016-11-16	–	<1	–	–	1	–	–	–	–	–	–	3	–	–	–	–
7	Bagot Long Lake	2016-11-16	–	–	–	–	–	–	–	–	–	–	–	<1	–	–	–	–
8	Bagot Long Lake	2016-11-16	–	–	<1	–	–	–	–	–	–	–	–	1	–	–	–	–
9	Bagot Long Lake	2016-11-16	–	–	–	–	–	–	–	–	–	–	–	2	–	–	–	–
15	Bay of Quinte	2016-08-23	2	13	<1	25	194	–	–	–	–	–	47	<LOQ	–	–	–	–
1	Constance Lake	2017-05-25	–	4	–	–	25	–	–	9	–	–	–	–	–	–	–	–
3	Constance Lake	2017-06-07	–	77	3	–	1097	–	–	–	–	–	41	–	–	–	–	–
4	Constance Lake	2017-06-07	–	1	<1	–	57	–	–	–	–	–	5	–	–	–	–	–
2	Constance Lake	2017-08-24	–	3	–	–	–	–	–	–	–	–	–	–	–	–	–	<LOQ
40	C-shore	2009-10-08	–	–	<1	–	–	–	<1	–	–	–	–	–	–	–	–	<1
24	Lac Breton	2011-07-04	–	1	–	–	106	–	–	–	–	–	–	–	–	–	–	<LOQ
10	Lac Breton	2017-05-21	–	2	–	–	53	–	–	–	–	–	–	–	–	–	–	–
11	Lac Breton	2017-05-21	–	3	–	–	53	–	–	–	–	–	–	–	–	–	–	–
54	Lac Breton	2019-08-13	–	8	<1	–	–	–	–	306	–	–	5	–	–	–	–	<LOQ
55	Lac Breton	2019-08-13	–	10	<1	–	–	–	–	220	–	–	4	–	–	–	–	–
12	Lac Breton	2019-09-02	–	–	–	–	–	–	–	156	<1	–	2	–	–	–	–	–
41	Lac Heney	2009-06-11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
49	Lac Heney	2009-06-11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
51	Lac Heney	2009-06-11	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	<1
43	Lac Heney	2009-06-30	<LOQ	13	–	–	85	–	–	–	9793	3043	134	249	1040	658	542	1642
20	Lac Meech	2016-11-29	–	–	–	–	–	–	–	–	2	–	–	<LOQ	–	–	–	–
21	Lac Meech	2016-11-29	–	<1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
22	Lac Meech	2016-11-29	–	–	–	–	–	–	–	–	–	–	–	<LOQ	–	–	–	–

25	Lac Meech	2016-11-29	–	–	–	–	–	–	–	–	–	–	–	<LOQ	–	–	–	–
26	Lac Meech	2016-11-29	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
14	Lac Meech	2019-06-17	–	1	–	–	6	–	<1	–	–	–	–	7	–	–	–	–
13	Lac Philippe	2019-06-19	–	–	–	–	–	–	–	–	–	–	–	<1	–	–	–	–
45	Lake X	2013-06-10	<1	9	<1	–	271	9	–	–	510	388	35	10	<1	3	<1	<1
28	Lake X	2013-06-18	<1	6	<1	–	199	–	–	–	1318	1129	199	20	135	99	45	3
29	Lake X	2013-06-18	<1	5	–	–	251	<1	–	–	1104	1055	164	24	147	117	49	68
42	Lake X	2013-06-18	<LOQ	1	–	–	9	–	–	–	783	505	6	26	127	72	58	–
52	Lake X	2013-06-18	<LOQ	2	<1	–	13	3	–	–	410	240	<1	19	21	18	10	<LOQ
27	Lake X	2013-07-09	<LOQ	4	–	–	13	12	–	–	246	288	18	12	35	40	17	–
32	Lake X	2013-07-16	–	–	–	–	–	–	–	–	5	–	–	–	<1	<1	<1	54
33	Lake X	2013-07-16	–	–	–	–	–	–	–	–	32	11	–	–	15	16	7	<LOQ
35	Lake X	2013-07-16	–	<1	–	–	–	–	–	–	14	8	3	<1	2	<1	<1	62
39	Lake X	2013-07-16	<LOQ	2	–	–	–	–	–	–	266	139	5	4	11	15	6	–
44	Lake X	2013-07-16	–	<1	–	–	–	–	–	–	71	40	8	<1	5	5	2	3
50	Lake X	2013-07-16	–	<1	–	–	–	–	–	–	90	41	9	–	6	3	3	–
30	Lake X	2013-10-10	<1	3	<1	–	483	–	–	–	645	460	230	6	24	17	9	81
31	Lake X	2013-10-10	<1	7	–	–	252	–	–	–	435	328	309	1	10	9	1	50
34	Lake X	2013-10-10	<LOQ	<1	–	–	10	–	–	–	1428	851	480	<LOQ	102	70	55	5
36	Lake X	2013-10-10	<1	<1	<1	–	285	–	–	–	212	129	159	5	11	11	5	<1
37	Lake X	2013-10-10	<1	11	<1	–	508	–	–	–	395	267	135	9	19	21	10	13
53	Lake X	2013-10-10	1	18	<1	–	545	–	–	–	674	444	144	13	42	35	28	15
23	Lake X	2013-07-09	<LOQ	1	–	–	–	–	–	–	110	41	–	1	10	9	3	–
18	Lake Nipissing	2016-08-07	–	<1	–	–	3	–	–	–	–	–	–	–	–	–	–	–
19	Lake Nipissing	2016-08-07	–	<1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
38	McNally Bay	2014-08-14	–	<1	–	–	15	–	–	–	–	–	4	–	–	–	–	29
48	McNally Bay	2014-08-14	–	–	–	–	10	–	–	–	13	5	3	<1	2	<1	<1	–
16	Mill Pond	2016-07-12	–	<1	–	–	44	–	–	–	–	–	–	–	–	–	–	–
47	Rice Lake	2013-08-22	<LOQ	4	<1	–	74	–	–	–	–	–	–	–	<1	–	–	<LOQ
17	South Nation R.	2016-07-12	–	<1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
46	Stand Out	2009-08-27	–	–	–	–	–	–	<1	–	–	–	–	–	–	–	–	247
		% of cABs	31%	67%	29%	2%	49%	7%	5%	7%	44%	36%	45%	53%	40%	38%	38%	42%

No. of cABs	18	38	16	2	28	5	4	5	24	21	26	29	23	22	22	24
Average	2	8	3	25	173	8	<1	173	807	471	90	22	93	68	50	175
Maximum	2	77	<1	25	1097	12	<1	306	9793	3043	480	249	1040	658	542	1642
Minimum	<LOQ	<1	<1	25	1	<1	<1	2	<1	5	2	<LOQ	<1	<1	<1	<LOQ

Table S2. LC-(HR)MS/MS inclusion list of the 117 CNPs identified

No.	GNPS Cluster No.	Cyanopeptide	*m/z*	RT (min)	Chemical formula	Mass error (ppm)	Structural feature 1	Structural feature 2	Study
1	10	MC RR	519.7915 $[M+2H]^{2+}$	2.69	$C_{49}H_{75}N_{13}O_{11}$	0.19	Arg	Arg	Kusumi et al., 1987
2	10	MC LA	910.4921 $[M+H]^{+}$	5.28	$C_{46}H_{67}N_{7}O_{12}$	0.06	Leu	Arg	Botes et al. 1984
3	10	[Dha^{7}]MC LR	981.5404 $[M+H]^{+}$	3.34	$C_{48}H_{72}N_{10}O_{12}$	-1.21	Leu	Arg	Harada et al. 1991
4	10	[D-Asp^{3}]MC LY	988.5021 $[M+H]^{+}$	5.32	$C_{51}H_{70}N7O_{13}$	-0.62	Leu	Tyr	Miles et al. 2012
5	10	MC LR	995.5556 $[M+H]^{+}$	3.30	$C_{49}H_{74}N_{10}O_{12}$	0.86	Leu	Arg	Botes et al. 1985
6	10	MC LY	1002.5183 $[M+H]^{+}$	5.42	$C_{52}H_{71}N_{7}O_{13}$	0.04	Leu	Tyr	del Campo & Ouahid 2010
7	10	MC WR	1068.5515 $[M+H]^{+}$	3.58	$C_{54}H_{73}N_{11}O_{12}$	0.19	Trp	Arg	Namikoshi et al. 1992
8	1	AS 576	497.2281 $[M+H\text{-}SO_3]^{+}$	2.94	$C_{25}H_{30}N_5O_9S$	2.42	Hpla-Phe	Arg	This Study
9	1	AS 654	575.3560 $[M+H\text{-}SO_3]^{+}$	2.36	$C_{29}H_{45}O_9N_6S$	1.39	Hpla-Phe	Agma	This Study
10	1	AS 688	609.34058 $[M+H\text{-}SO3]^{+}$	2.45	$C_{32}H_{44}N_6O_9S$	1.78	Hpla-Phe	Agma	McDonald 2021
11	1	AS 609	610.3235 $[M+H]^{+}$	2.66	$C_{32}H_{43}N_5O_7$	-0.04	Hpla-Phe	Agma+2H	This Study
12	1	AS 703	625.3344 $[M+H\text{-}SO_3]^{+}$	2.38	$C_{32}H_{43}O_{10}N_6S$	-0.02	Hpla-Phe	Agma	This Study
13	1	AS 722	643.3007 $[M+H\text{-}SO_3]^{+}$	2.56	$C_{32}H_{43}N_6O_9ClS$	0.28	Hpla-Phe	Agma	McDonald 2021

14	1	AS 736	657.3264 $[M+H-SO_3]^+$	2.34	$C_{33}H_{40}N_{10}O_8S$	1.26	Hpla-Phe	Agma	This Study
15	2	AS 716	637.3106 $[M+H-SO_3]^+$	2.48	$C_{30}H_{45}N_6O_{10}ClS$	-0.77	Hpla-Phe	Argal	This Study
16	2	AS 750	671.2961 $[M+H-SO_3]^+$	2.56	$C_{33}H_{43}N_6O_{10}ClS$	0.94	Hpla-Phe	Argal	This Study
17	3	AP 940	471.2768 $[M+2H]^{2+}$	2.41	$C_{50}H_{72}O_6N_{10}S$	2.41	Arg-CO	Hty	This Study
18	3	AP 641	642.3724 $[M+H]^+$	2.57	$C_{31}H_{47}N_9O_6$	0.36	Arg-CO	Trp	This Study
19	3	AP 679	680.3763 $[M+H]^+$	2.95	$C_{35}H_{49}N_7O_7$	-0.5	D-Lys	Pro	Harms et al. 2016
20	3	AP 706	707.4490 $[M+H]^+$	3.02	$C_{39}H_{58}N_6O_6$	-0.03	Hty	HPhe	This Study
21	3	AP I	760.4600 $[M+H]^+$	3.44	$C_{38}H_{61}N_7O_9$	-0.44	Lxx-CO	Hty	Murakami et al. 2000
22	3	AP 765	766.4509 $[M+H]^+$	3.28	$C_{40}H_{59}N_7O_8$	-1.41	Hty-CO	Lxx	This Study
23	3	AP J	794.4447 $[M+H]^+$	3.65	$C_{41}H_{59}N_7O_9$	0	Lxx-CO	Phe	Murakami et al. 1997
24	3	AP 823	824.4548 $[M+H]^+$	3.20	$C_{42}H_{61}N_7O_{10}$	-0.6	Lxx-CO	Hty	This Study
25	3	AP D	828.4283 $[M+H]^+$	3.84	$C_{44}H_{57}N_7O_9$	-0.93	Hty	Phe	Mazur-Marzec et al. 2016
26	3	AP A	844.4229 $[M+H]^+$	3.22	$C_{44}H_{57}N_7O_{10}$	-1.25	Hty	Phe	Harada et al. 1995
27	3	AP 851	852.4843 $[M+H]^+$	3.59	$C_{40}H_{62}N_{13}O_8$	0.43	Hty	Lxx	This Study
28	3	AP T	866.5021 $[M+H]^+$	3.85	$C_{45}H_{66}N_7O_{10}$	-0.16	Hty	Lxx	Kodani et al. 1999
29	3	FA A	867.4400 $[M+H]^+$	3.87	$C_{46}H_{57}N_8O_9$	-1.7	Hty	Trp	Williams et al. 1996

30	3	AP 871	872.4548 $[M+H]^+$	3.58	$C_{46}H_{60}N_7O_{10}$	-0.5	Hty	Phe	Sanz et al. 2015
31	3	AP 885	886.4701 $[M+H]^+$	3.78	$C_{47}H_{63}N_7O_{10}$	-0.93	MeHty	Phe	This Study
32	3	AP KB899	900.4865 $[M+H]^+$	4.01	$C_{48}H_{64}N_7O_{10}$	-0.06	Hty	Lxx	Elkobi-Peer & Carmeli 2015
33	3	AP 901	902.4655 $[M+H]^+$	3.15	$C_{39}H_{67}N_9O_{13}S$	0.3	Hty	Val	This Study
34	3	AP MM913	914.5015 $[M+H]^+$	4.30	$C_{49}H_{67}N_7O_{10}$	-0.75	Hty	HPhe	Zafrir-Ilan et al. 2010
35	3	AP 915	916.4813 $[M+H]^+$	3.32	$C_{48}H_{65}N_7O_{11}$	-0.24	Hty	Lxx	Okumura et al. 2009
36	3	AP 929	930.4034 $[M+H]^+$	3.80	$C_{50}H_{56}N_7O_{11}$	0.22	Hty-CO	Lxx	This Study
37	3	AP G-Itou	930.4962 $[M+H]^+$	3.52	$C_{49}H_{67}N_7O_{11}$	-1.06	Hty-CO	Lxx	Itou et al. 2009
38	3	AP 931	932.5140 $[M+H]^+$	3.10	$C_{42}H_{69}N_{13}O_9S$	0.55	Hty-CO	Lxx	This Study
39	3	AP 947	948.5081 $[M+H]^+$	2.70	$C_{49}H_{69}N_7O_{12}$	0.38	Hty-CO	Lxx	This Study
40	4	AP 971	487.2673 $[M+2H]^{2+}$	2.33	$C_{50}H_{69}O_{10}N_8S$	1.54	Arg-CO	Phe	This Study
41	4	AP B	837.4613 $[M+H]^+$	2.56	$C_{41}H_{60}N_{10}O_9$	-0.54	Arg-CO	Tyr	Harada et al. 1995
42	4	AP 844	845.5231 $[M+H]^+$	2.68	$C_{40}H_{73}N_6O_{13}$	0.07	Arg-CO	Hty	This Study
43	4	AP E	851.4774 $[M+H]^+$	2.61	$C_{42}H_{63}N_{10}O_9$	-0.05	Arg-CO	Phe	Shin et al. 1998
44	4	AP 858	859.5400 $[M+H]^+$	2.75	$C_{42}H_{71}N_{10}O_9$	0.05	Arg-CO	Hty	This Study
45	4	AP HU892	893.5237 $[M+H]^+$	2.82	$C_{45}H_{67}N_{10}O_9$	-0.68	Arg-CO	Hty	Gesner-Apter & Carmeli 2009

46	4	AP 906b	907.5401 $[M+H]^+$	2.89	$C_{46}H_{70}N_{10}O_9$	0.11	Arg-CO	Hty	Puddick & Prinsep 2008
47	4	AP G-Erhard	909.5194 $[M+H]^+$	2.58	$C_{45}H_{67}N_{10}O_{10}$	0.09	Arg-CO	Hty	Erhard et al. 1999
48	4	AP H	923.5349 $[M+H]^+$	2.65	$C_{46}H_{70}N_{10}O_{10}$	-0.07	Arg-CO	Hty	Itou et al. 2009
49	4	AP 923	924.5198 $[M+H]^+$	2.99	$C_{46}H_{69}N_9O_{11}$	0.98	Hty	Trp	This Study
50	5	AC B	533.1998 $[M+H]^+$	5.83	$C_{24}H_{32}O_4N_6S_2$	-0.28	Tzl	Tzl	Peña et al. 2013
51	5	AC A	535.2155 $[M+H]^+$	5.62	$C_{24}H_{34}N_6O_4S_2$	-0.23	Tzln	Tzl	Portmann et al. 2008
52	6	Md A	583.1906 $[M+H]^+$	2.83	$C_{26}H_{30}N_8O_4S_2$	0.33	Tzl	Tzl	Ishida & Murakami 2000
53	6	Md 584	585.2059 $[M+H]^+$	2.76	$C_{26}H_{32}N_8O_4S_2$	-0.24	Tzln	Tzl	McDonald 2021
54	6	Md 586	587.2217 $[M+H]^+$	2.64	$C_{26}H_{34}N_8O_4S_2$	0.04	Tzln	Tzln	McDonald 2021
55	7	CB 420	421.1543 $[M+H]^+$	2.27	$C_{19}H_{25}O_5N_4S$	0.74	???	???	This Study
56	7	CB 614 (1)	615.2054 $[M+H]^+$	3.23	$C_{28}H_{34}O_6N_6S_2$	0	Tzl	Phe	This Study
57	7	CB 614 (2)	615.2384 $[M+H]^+$	5.93	$C_{32}H_{34}O_5N_6S$	-0.02	Tzl	Phe	This Study
58	7	CB 616	617.2211 $[M+H]^+$	3.44	$C_{28}H_{37}O_6N_6S_2$	0.03	Tzl	Phe	This Study
59	7	Md GL616	617.2747 $[M+H]^+$	5.49	$C_{29}H_{40}N_6O_7S$	0.19	Tzl	Phe	
60	7	CB 632A	633.2162 $[M+H]^+$	2.59	$C_{36}H_{32}O_5N_4S$	-0.66	Unk	Phe	This Study
61	7	CB 632B	633.2483 $[M+H]^+$	4.65	$C_{32}H_{44}O_5N_2S_3$	0.02	Unk	Phe	This Study

62	7	CB 650A	651.2270 $[M+H]^+$	2.39	$C_{36}H_{34}O_6N_4S$	-0.28	Unk	Phe	This Study
63	7	CB 650B	651.2590 $[M+H]^+$	3.32	$C_{32}H_{46}O_6N_2S_3$	-0.16	Unk	Phe	This Study
64	7	CB 684	685.3377 $[M+H]^+$	5.49	$C_{34}H_{49}O_7N_6S$	-0.12	Unk	Phe	This Study
65	8	MP MZ845	828.4609 $[M+H-H_2O]^+$	3.35	$C_{40}H_{63}N_9O_{11}$	-0.57	Lxx-Ahp	*NMe*Phe	Zafrir E. & Carmeli S 2010
66	8	CP 866	849.4395 $[M+H-H_2O]^+$	4.70	$C_{44}H_{60}N_6O_{11}$	0.27	Lxx-Ahp	*NMe*Phe	This Study
67	8	CP 908 (880)	863.4528 $[M+H-H_2O]^+$	5.03	$C_{39}H_{64}N_{10}O_{13}$	-0.04	Lxx-Ahp	*NMe*Phe	Okumura et al. 2009
68	8	CP 882	865.4335 $[M+H-H_2O]^+$	4.45	$C_{44}H_{62}N_6O_{13}$	-0.77	Lxx-Ahp	*NMe*Hty	This Study
69	8	CP 875	876.4817 $[M+H]^+$	3.25	$C_{41}H_{65}N_9O_{12}$	-0.98	Lxx-Ahp	*NMe*Tyr	This Study
70	8	MP LH911B	894.4019 $[M+H-H_2O]^+$	3.06	$C_{39}H_{61}N_9O_{14}S$	-0.81	Val-Ahp	*NMe*Phe	Vegman & Carmeli 2013
71	8	CP S	926.4286 $[M+H]^+$	3.57	$C_{40}H_{63}N_9O_{14}S$	-0.22	Lxx-Ahp	*NMe*Phe	Jakobi et al. 1995
72	8	CP 955	956.4388 $[M+H-H_2O]^+$	3.23	$C_{45}H_{57}N_{13}O_{11}$	1.52	Lxx-Ahp	*NMe*Hty	This Study
73	9	CP 1017 (2)	1000.5132 $[M+H-H_2O]^+$	3.57	$C_{52}H_{76}N_9O_8S_2$	-1.47	Phe-Ahp	*NMe*Tyr	McDonald 2021
74	9	CP 1017 (1)	1000.5136 $[M+H-H_2O]^+$	3.40	$C_{52}H_{67}N_{13}O_9$	-1.63	Phe-Ahp	*NMe*Tyr	McDonald 2021
75	9	CP 1000	1001.5671 $[M+H]^+$	3.47	$C_{48}H_{76}N_{10}O_{13}$	0.53	Lxx-Ahp	*NMe*Tyr	This Study
76	9	CP 1051 (A)	1034.4634 $[M+H-H2O]^+$	4.49	$C_{52}H_{70}N_7O_{14}Cl$	-0.25	Phe-Ahp	*NMe*Tyr-Cl	McDonald 2021
77	9	CP 1051 (C)	1034.4751 $[M+H-H2O]^+$	3.70	$C_{50}H_{74}N_5O_{17}Cl$	1.27	Phe-Ahp	*NMe*Tyr-Cl	This Study

78	9	CP 1051 (B)	1034.4969 $[M+H-H2O]^+$	3.48	$C_{49}H_{70}N_{13}O_{11}Cl$	-0.43	Phe-Ahp	*NMe*Tyr-Cl	This Study
79	9	CP 1085	1068.4618 $[M+H-H2O]^+$	3.81	$C_{53}H_{73}N_7O_{13}Cl_2$	0.69	Phe-Ahp	*NMe*Tyr-Cl	This Study
80	9	CP 1109	1110.5739 $[M+H]^+$	3.22	$C_{54}H_{79}N_9O_{16}$	1.89	Lxx-Ahp	*NMe*Tyr	This Study
81	9	CP 1143	1144.5564 $[M+H]^+$	3.32	$C_{57}H_{77}N_9O_{16}$	0.26	Phe-Ahp	*NMe*Tyr	McDonald 2021
82	9	CP 750	733.3916 $[M+H-H2O]^+$	3.44	$C_{34}H_{54}N_8O_{11}$	-1.96	Phe-Ahp	*NMe*Tyr	McDonald 2021
83	9	CP 784	767.3521 $[M+H-H2O]^+$	3.82	$C_{38}H_{53}N_8O_8Cl$	-2.55	Phe-Ahp	*NMe*Tyr-Cl	This Study
84	9	CP 805	787.4025 $[M+H-H2O]^+$	2.81	$C_{35}H_{56}N_{12}O_{10}$	-1.36	Phe-Ahp	*NMe*Tyr	This Study
85	9	CP 838	821.3637 $[M+H-H2O]^+$	2.66	$C_{35}H_{59}N_6O_{15}Cl$	0.78	Phe-Ahp	*NMe*Tyr-Cl	This Study
86	9	CP 886	869.4761 $[M+H-H2O]^+$	3.02	$C_{43}H_{65}N_8O_{12}$	-0.76	Lxx-Ahp	*NMe*Tyr	This Study
87	9	CP 920	903.4368 $[M+H-H2O]^+$	3.28	$C_{39}H_{61}N_{14}O_{10}Cl$	1.27	Lxx-Ahp	*NMe*Tyr-Cl	McDonald 2021
88	9	NP BN920	903.4618 $[M+H-H2O]^+$	3.13	$C_{46}H_{64}N_8O_{12}$	0.78	Phe-Ahp	*NMe*Tyr	Von Elert et al. 2005
89	9	CP 954	937.4211 $[M+H-H2O]^+$	3.41	$C_{46}H_{63}N_8O_{12}Cl$	-1.03	Phe-Ahp	*NMe*Tyr-Cl	Von Elert et al. 2005
90	9	CP 981	962.4854 $[M+H-H2O]^+$	4.20	$C_{49}H_{70}N_7O_{14}$	-1.42	Lxx-Ahp	*NMe*Tyr	This Study
91	9	CP 990	973.5609 $[M+H]^+$	3.42	$C_{48}H_{76}N_8O_{13}$	0.46	Lxx-Ahp	*NMe*Tyr	This Study
92	9	CP 1003	986.4974 $[M+H-H2O]^+$	3.17	$C_{50}H_{69}N_9O_{13}$	-0.77	Phe-Ahp	*NMe*Tyr	This Study
93	9	CP 1007	990.5165 $[M+H-H2O]^+$	5.18	$C_{52}H_{69}N_{11}O_{10}$	0.86	Lxx-Ahp	*NMe*Tyr	This Study

94	9	CP 1013	996.4691 $[M+H-H2O]^+$	4.36	$C_{47}H_{67}N_9O_{16}$	1.84	Phe-Ahp	*NMe*Tyr	McDonald 2021
95	9	CP 800	783.3711 $[M+H-H2O]^+$	3.01	$C_{42}H_{52}N_6O_{10}$	-1.55	Phe-Ahp	*NMe*Tyr	Welker et al. 2006
96	None	CP 847	848.4496 $[M+H]^+$	2.65	$C_{43}H_{61}N_9O_7S$	0.99	Val-Ahp	*NMe*Tyr	This Study
97	11	MG 507	508.3383 $[M+H]^+$	3.11	$C_{27}H_{45}N_3O_6$	0.35	*Nme*-Ahda	*NMe*Lxx	McDonald 2021
98	11	MG 557	558.3185 $[M+H]^+$	2.74	$C_{30}H_{43}N_3O_7$	1.98	*Nme*-Ahda	*NMe*Tyr	McDonald 2021
99	13	MG 514	515.3442 $[M+H]^+$	2.73	$C_{25}H_{47}N_4O_7$	0.12	Ahoa	*NMe*Lxx	McDonald 2021
100	13	MG 516	517.3150 $[M+H]^+$	2.81	$C_{25}H_{45}N_4O_5Cl$	-0.24	Ahda-Cl	*NMe*Lxx	McDonald 2021
101	13	MG 528	529.3242 $[M+H]^+$	2.32	$C_{25}H_{45}O_8N_4$	1.94	Ahoa	*NMe*Lxx	McDonald 2021
102	13	MG 532	533.3098 $[M+H]^+$	2.75	$C_{25}H_{45}N_4O_6Cl$	-0.43	Ahda-Cl	NMeLxx	McDonald 2021
103	13	MG 556	557.3541 $[M+H]^+$	2.99	$C_{27}H_{48}N_4O_8$	-0.65	Ahoa	Lxx	McDonald 2021
104	13	MG 590	591.3154 $[M+H]^+$	2.95	$C_{27}H_{48}N_4O_8Cl$	-0.15	Ahoa	Lxx	This Study
105	13	MG 661	662.4123 $[M+H]^+$	2.85	$C_{34}H_{55}N_5O_8$	-0.08	Ahda	NMeLxx	McDonald 2021
106	12	MG 712	713.3500 $[M+H]^+$	2.40	$C_{35}H_{48}N_6O_{10}$	-0.7	Unk	*NMe*Tyr-Cl	This Study
107	12	MG 755	756.3821 $[M+H]^+$	2.91	$C_{38}H_{54}N_5O_{11}$	0.85	Ahoa	*NMe*Tyr	McDonald 2021
108	12	MG 922	923.4741 [M+H]+	2.84	$C_{52}H_{72}N_2O_8Cl_2$	0.23	Unk	*NMe*Tyr	This Study
109	12	MG 956	957.4368 [M+H]+	3.00	$C_{42}H_{66}N_{10}O_{11}Cl_2$	0.56	Unk	*NMe*Tyr-Cl	This Study

110	14	MG 551	551.3436 $[M+H]^+$	2.74	$C_{28}H_{46}N_4O_7$	-0.65	Ahda	Val NMeAla NMeTyr	McDonald 2021
111	14	MG KR604	605.3912 $[M+H]^+$	2.87	$C_{32}H_{52}N_4O_7$	0.49	*NMe*-Ahda	NMeLxx	Lodin-Friedman & Carmeli 2018
112	14	MG 606A	607.4059 $[M+H]^+$	3.04	$C_{32}H_{54}N_4O_7$	-0.97	*NMe*-Ahda	NMeLxx Val	Zervou et al. 2020
113	14	MG 620A/B	621.4222 $[M+H]^+$	3.13	$C_{33}H_{56}N_4O_7$	-0.28	*NMe*-Ahda	NMeLxx Val	McDonald 2021
114	14	MG 654	655.3701 $[M+H]^+$	2.64	$C_{35}H_{51}N_4O_8$	-0.03	*NMe*-Ahda	*NMe*Tyr Tyr Pro	This Study
115	14	MG 684	685.4166 $[M+H]^+$	3.20	$C_{37}H_{56}N_4O_8$	-0.72	*NMe*-Ahda		McDonald 2021
116	14	MG FR13	770.4697 $[M+H]^+$	3.17	$C_{41}H_{63}N_5O_9$	-0.24	*NMe*-Ahda		Ishida et al. 1998
117	14	MG FR12	784.4850 $[M+H]^+$	3.24	$C_{42}H_{65}N_5O_9$	-0.66	*NMe*-Ahda		Fastner J. et al. unpub.

Table S3. cAB sample collection and extraction information. Coordinates are given as general locations for each watercourse.

No.	Watercourse	Date	Coordinates (UTMs)	Filtered Volume (mL)	Biomass (mg)
1	Constance Lake	25-05-2017	18N 423749 5028281	500.0	0.9
2	Constance Lake	24-08-2017	18N 423749 5028281	500.0	18.6
3	Constance Lake	07-06-2017	18N 423749 5028281	500.0	3.7
4	Constance Lake	07-06-2017	18N 423749 5028281	500.0	1.4
5	Bagot Long Lake	16-11-2016	18N 370503 5010538	17.5	5.4
6	Bagot Long Lake	16-11-2016	18N 370503 5010538	20.0	2.6
7	Bagot Long Lake	16-11-2016	18N 370503 5010538	25.0	7.2
8	Bagot Long Lake	16-11-2016	18N 370503 5010538	25.0	6.5
9	Bagot Long Lake	16-11-2016	18N 370503 5010538	25.0	2.7
10	Lac Breton	21-05-2017	18N 559865 5080298	500.0	1.6
11	Lac Breton	21-05-2017	18N 559865 5080298	250.0	1.06
12	Lac Breton	02-09-2019	18N 559865 5080298	15.0	9.1
13	Lac Philippe	19-06-2019	18N 421615 5050587	4.0	6.0
14	Lac Meech	17-06-2019	18N 429703 5043516	30.0	8.62
15	Bay of Quinte	23-08-2016	18N 317877 4892117	450.0	2.22
16	Mill Pond	12-07-2016	18N 528512 5017325	400.0	2.0
17	South Nation River	12-07-2016	18N 492468 5017900	400.0	9.4
18	Lake Nipissing	07-08-2016	17N 614192 5117154	400.0	1.4
19	Lake Nipissing	07-08-2016	17N 614192 5117154	10.0	12.1
20	Lac Meech	29-11-2016	18N 429703 5043516	1000.0	0.5
21	Lac Meech	29-11-2016	18N 429703 5043516	4.0	11.5

22	Lac Meech	29-11-2016	18N 429703 5043516	4.0	8.5
23	Lake X	09-07-2013	18N 454295 5009941	200.0	3.9
24	Lac Breton	04-07-2011	18N 559865 5080298	500.0	1.6
25	Lac Meech	29-11-2016	18N 429703 5043516	1000.0	0.8
26	Lac Meech	29-11-2016	18N 429703 5043516	3.0	7.2
27	Lake X	09-07-2013	18N 454295 5009941	20.0	9.5
28	Lake X	18-06-2013	18N 454295 5009941	10.0	41
29	Lake X	18-06-2013	18N 454295 5009941	15.0	16.1
30	Lake X	10-10-2013	18N 454295 5009941	20.0	7.3
31	Lake X	10-10-2013	18N 454295 5009941	10.0	12
32	Lake X	16-07-2013	18N 454295 5009941	10.0	11.9
33	Lake X	16-07-2013	18N 454295 5009941	10.0	34.8
34	Lake X	10-10-2013	18N 454295 5009941	15.0	13.5
35	Lake X	16-07-2013	18N 454295 5009941	100.0	6.4
36	Lake X	10-10-2013	18N 454295 5009941	10.0	12.4
37	Lake X	10-10-2013	18N 454295 5009941	20.0	7.4
38	McNally Bay	14-08-2014	18N 394252 4945862	30.0	8.54
39	Lake X	16-07-2013	18N 454295 5009941	10.0	7.3
40	C-shore	08-10-2009	N/A	1000.0	103.3
41	Lac Heney	11-06-2009	18N 428226 5097705	1000.0	0.8
42	Lake X	18-06-2013	18N 454295 5009941	20.0	8.2
43	Lac Heney	30-06-2009	18N 428226 5097705	1000.0	1.5
44	Lake X	16-07-2013	18N 454295 5009941	10.0	14.4

45	Lake X	10-06-2013	18N 454295 5009941	100.0	16.2
46	Stand Out	27-08-2009	N/A	1000.0	24.8
47	Rice Lake	22-08-2013	17N 725067 4895267	500.0	3.0
48	McNally Bay	14-08-2014	18N 394252 4945862	400.0	5.8
49	Lac Heney	11-06-2009	18N 428226 5097705	100.0	25.5
50	Lake X	16-07-2013	18N 454295 5009941	200.0	1.5
51	Lac Heney	11-06-2009	18N 428226 5097705	4.0	6.3
52	Lake X	18-06-2013	18N 454295 5009941	10.0	10.6
53	Lake X	10-10-2013	18N 454295 5009941	5.0	30.6
54	Lac Breton	13-08-2019	18N 559865 5080298	15.0	9.8
55	Lac Breton	13-08-2019	18N 559865 5080298	50.0	51.0

Table S4. Colonel By Lake sediment core results for CRS chronology, Itrax-XRF elemental data (cps), Chlorophyl-a (mg/g) and Total MC concentrations (ng/g)

Depth interval (cm)		Age (CE)	Itrax-XRF elemental data (normalized cps)													Elemental ratios (unitless)			[Chl-a] (mg/g)	Total [MC] (ng/g)
Top	Bottom		Ti	Zr	Sr	Fe	Rb	Mg	K	Ca	CIR	S	P	Si	Cl	Si/Ti	Ti/K	Zr/Rb		
0	0.25	2020	12588	62	752	25121	365	10	16016	18265	1.92	91	27	2359	122	0.187	0.786	0.171	0.026	–
0.25	0.5	2020	12772	169	743	25562	276	9	15610	18241	1.94	79	8	2149	616	0.168	0.818	0.614	0.026	–
0.5	1	2020	12176	158	766	26332	322	11	14498	17812	1.84	83	18	1931	117	0.159	0.840	0.491	0.030	–
1	1.5	2019	11901	129	755	25097	294	18	14919	17520	1.93	91	13	2221	48	0.187	0.798	0.437	0.031	<LOQ
1.5	2	2019	12636	121	736	25624	313	19	15782	18179	1.88	80	14	2415	43	0.191	0.801	0.387	0.032	–
2	2.5	2018	12189	147	725	24948	346	18	15148	17853	1.89	104	17	2322	49	0.190	0.805	0.427	0.032	–
2.5	3	2017	11862	132	725	24498	322	12	14353	16718	1.84	77	7	1935	134	0.163	0.826	0.410	0.033	–
3	3.5	2016	12637	133	720	24138	249	14	17054	19833	2.02	129	14	3345	50	0.265	0.741	0.533	0.033	–
3.5	4	2015	13000	141	717	25441	248	17	16823	19306	2.01	155	23	2988	38	0.230	0.773	0.571	0.034	–
4	4.5	2014	12878	142	758	24652	265	16	16865	19472	1.93	162	12	3061	79	0.238	0.764	0.537	0.031	–
4.5	5	2013	13231	205	730	25072	330	21	16887	19711	1.92	203	9	2890	28	0.218	0.783	0.621	0.031	–
5	5.5	2012	12036	165	717	22976	276	14	15939	17954	2.01	202	9	2968	26	0.247	0.755	0.596	0.032	–
5.5	6	2011	12911	145	709	24757	306	14	16558	18420	1.88	193	22	2864	37	0.222	0.780	0.475	0.033	–
6	6.5	2010	12535	108	691	24946	303	15	15731	17603	1.94	195	8	2479	31	0.198	0.797	0.356	0.032	–
6.5	7	2009	12116	154	668	24042	307	14	15604	17975	1.96	270	19	2634	35	0.217	0.776	0.500	0.032	–
7	7.5	2007	12034	149	724	24434	298	19	15161	17320	1.90	155	9	2282	27	0.190	0.794	0.501	0.030	–
7.5	8	2006	12833	147	756	25620	295	17	15982	17895	1.87	137	22	2455	35	0.191	0.803	0.499	0.028	–
8	8.5	2005	13162	177	748	24859	272	20	16879	18434	2.03	152	9	3109	26	0.236	0.780	0.653	0.027	–
8.5	9	2004	12749	152	750	25102	317	18	16498	18130	1.99	187	16	2821	27	0.221	0.773	0.480	0.028	–
9	9.5	2002	13195	194	743	25861	283	7	16604	18600	1.93	225	13	2684	31	0.203	0.795	0.685	0.027	–
9.5	10	2001	12080	165	735	24851	280	18	15376	17332	1.89	251	6	2460	31	0.204	0.786	0.587	0.027	–
10	10.5	2000	12472	175	750	24891	310	15	15997	17434	1.90	136	5	2435	25	0.195	0.780	0.564	0.027	–
10.5	11	1999	12672	137	756	25762	324	19	16164	17118	1.93	143	10	2609	25	0.206	0.784	0.424	0.027	–
11	11.5	1997	13274	184	787	26374	283	24	17022	18392	1.89	168	10	2759	29	0.208	0.780	0.648	0.026	–
11.5	12	1996	12918	143	733	25120	326	16	16395	18630	1.94	221	23	2787	36	0.216	0.788	0.439	0.027	–
12	12.5	1995	12541	163	721	24856	281	14	16144	17567	1.94	174	10	2823	24	0.225	0.777	0.581	0.026	–
12.5	13	1994	11809	151	678	22142	273	12	15371	16932	2.05	186	16	2741	26	0.232	0.768	0.553	0.027	–
13	13.5	1992	13055	124	674	25599	300	26	17197	18877	2.06	238	19	3309	23	0.253	0.759	0.413	0.026	–

13.5	14	1991	12932	123	712	25909	283	21	16701	18251	1.93	228	15	2874	24	0.222	0.774	0.433	0.027	–
14	14.5	1990	12654	118	693	24740	323	20	15955	17675	1.91	154	11	2694	27	0.213	0.793	0.366	0.026	–
14.5	15	1988	13069	169	723	26354	277	18	16397	18537	1.96	206	11	2627	31	0.201	0.797	0.612	0.027	–
15	15.5	1987	12706	156	788	25482	329	19	15843	18309	1.91	163	12	2391	21	0.188	0.802	0.474	0.026	–
15.5	16	1986	11986	156	695	24543	322	10	14831	16843	1.88	165	8	2251	31	0.188	0.808	0.486	0.027	–
16	16.5	1984	11903	150	703	24398	271	10	14489	16840	1.95	135	3	1877	22	0.158	0.822	0.554	0.026	–
16.5	17	1983	12530	154	760	26732	325	9	14525	17023	1.87	149	0	1769	26	0.141	0.863	0.474	0.027	–
17	17.5	1981	11521	152	744	25002	298	25	13838	16279	1.80	138	4	1786	33	0.155	0.833	0.510	0.026	–
17.5	18	1979	12403	156	728	26297	276	13	14555	16724	1.84	136	5	1821	23	0.147	0.852	0.563	0.028	–
18	18.5	1978	11715	169	778	25508	303	16	13737	16518	1.84	159	6	1644	30	0.140	0.853	0.556	0.026	–
18.5	19	1976	11317	126	771	25441	331	22	12532	15357	1.74	123	6	1234	23	0.109	0.903	0.379	0.027	–
19	19.5	1974	11164	201	789	25096	355	10	12872	15440	1.77	121	3	1311	30	0.117	0.867	0.566	0.027	–
19.5	20	1972	11876	176	753	25745	280	21	13811	17177	1.80	214	7	1643	27	0.138	0.860	0.630	0.028	–
20	20.5	1969	11643	190	764	25005	292	15	13710	16241	1.80	127	16	1518	41	0.130	0.849	0.651	0.026	–
20.5	21	1967	11811	141	787	25743	307	12	14148	16018	1.81	137	12	1660	51	0.141	0.835	0.460	0.026	–
21	21.5	1964	11649	163	753	25376	365	19	13302	15340	1.85	136	2	1349	27	0.116	0.876	0.446	0.026	–
21.5	22	1961	11859	142	756	26208	326	22	13095	15411	1.79	139	6	1291	23	0.109	0.906	0.435	0.026	–
22	22.5	1958	12340	117	762	26202	302	22	14315	16243	1.81	148	11	1591	36	0.129	0.862	0.388	0.023	–
22.5	23	1954	12692	161	773	27093	300	15	14335	16341	1.74	138	7	1533	78	0.121	0.885	0.535	0.023	–
23	23.5	1951	12158	177	701	25488	316	5	14052	15881	1.80	135	2	1559	26	0.128	0.865	0.561	0.021	–
23.5	24	1946	13158	135	782	26758	292	20	15603	16795	1.88	123	4	1922	20	0.146	0.843	0.463	0.021	–
24	24.5	1942	13106	177	778	27851	316	11	15733	16724	1.85	122	4	1913	29	0.146	0.833	0.560	0.021	–
24.5	25	1937	13146	157	760	26476	303	13	16434	17399	1.91	121	11	2270	30	0.173	0.800	0.518	0.020	–
25	25.5	1932	13082	176	727	26488	286	14	15948	17105	1.91	138	13	2095	29	0.160	0.820	0.615	0.019	–
25.5	26	1927	12764	158	754	25966	320	7	15379	16329	1.86	108	8	1940	25	0.152	0.830	0.493	0.019	–
26	26.5	1921	13329	188	741	27929	307	15	15900	16807	1.88	133	15	1921	32	0.144	0.838	0.613	0.019	–
26.5	27	1914	13607	121	754	27298	313	16	16549	17170	1.87	119	18	2213	33	0.163	0.822	0.387	0.019	–
27	27.5	1907	14079	174	772	28501	284	23	17095	17979	1.89	181	8	2298	21	0.163	0.824	0.613	0.018	–
27.5	28	1900	13491	138	729	27327	300	8	16943	19274	1.91	326	6	2471	26	0.183	0.796	0.459	0.018	–
28	28.5	1892	13614	172	756	26965	310	19	16475	17552	1.91	201	8	2210	24	0.162	0.826	0.556	0.017	–
28.5	29	1883	13376	193	777	25799	301	15	16440	18667	1.91	284	8	2415	17	0.181	0.814	0.640	0.017	–

Table S5. Dog Lake sediment core results for the CRS chronology, Itrax-XRF elemental data (cps), Chlorophyll-a (mg/g) and Total MC concentrations (ng/g)

Depth interval (cm)		Age (CE)	Itrax-XRF elemental data (normalized cps)													Elemental ratios (unitless)				[Chl-a] (mg/g)	Total [MC] (ng/g)
Top	Bottom		Ti	Zr	Sr	Fe	Rb	Mg	K	Ca	CIR	S	P	Si	Cl	Si/Ti	Ca/Sr	Ca/Ti	Ti/K		
0	0.5	2020	9674	86	770	21255	284	24	10855	16310	1.74	270	14	1273	44	0.08	21.19	1.69	0.89	0.032	29.0
0.5	1	2019	8441	31	739	20226	329	18	8701	13659	1.64	222	10	720	35	0.05	18.49	1.62	0.97	0.032	55.0
1	1.5	2018	8257	15	763	20482	337	16	8541	13338	1.64	223	6	742	29	0.06	17.49	1.62	0.97	0.033	86.8
1.5	2	2018	8390	26	722	19790	369	30	8806	13624	1.62	239	18	799	32	0.06	18.86	1.62	0.95	0.031	237.5
2	2.5	2018	8418	34	708	20126	286	17	8894	13838	1.65	238	7	832	30	0.06	19.56	1.64	0.95	0.033	13.1
2.5	3	2017	7982	22	704	19086	388	20	8526	13037	1.60	236	15	836	36	0.06	18.52	1.63	0.94	0.034	161.4
3	3.5	2017	7726	47	697	19148	343	21	7908	12586	1.62	217	8	656	20	0.05	18.06	1.63	0.98	0.035	237.0
3.5	4	2016	8243	43	682	20232	337	37	8461	13633	1.62	251	6	680	24	0.05	19.99	1.65	0.97	0.034	36.2
4	4.5	2015	7846	41	714	19439	356	33	8045	13249	1.65	256	7	674	23	0.05	18.55	1.69	0.98	0.035	93.9
4.5	5	2014	8573	42	699	18983	325	18	9643	14757	1.60	313	12	1084	45	0.07	21.12	1.72	0.89	0.035	80.7
5	5.5	2013	8335	49	747	20526	340	31	8727	13730	1.58	237	6	739	139	0.05	18.38	1.65	0.96	0.035	171.2
5.5	6	2012	8173	55	708	19850	346	33	8573	13326	1.60	236	8	758	30	0.06	18.82	1.63	0.95	0.035	23.8
6	6.5	2010	8395	80	725	20151	314	39	8666	13493	1.63	234	9	730	22	0.05	18.62	1.61	0.97	0.033	41.7
6.5	7	2009	8165	29	708	20487	313	31	8035	12955	1.58	214	3	547	23	0.04	18.30	1.59	1.02	0.033	229.8
7	7.5	2007	7526	12	789	20049	334	22	7158	12199	1.58	204	0	435	20	0.04	15.45	1.62	1.05	0.032	138.5
7.5	8	2005	8297	39	793	20571	352	17	7972	12909	1.63	231	3	529	32	0.04	16.27	1.56	1.04	0.030	<LOQ
8	8.5	2003	7761	38	757	20533	345	24	7615	12401	1.60	227	7	534	25	0.04	16.38	1.60	1.02	0.030	6.7
8.5	9	2001	7617	63	757	20513	348	33	7347	12294	1.60	217	2	475	24	0.04	16.24	1.61	1.04	0.029	5.1
9	9.5	1998	7998	72	761	20684	319	21	7774	13008	1.58	240	6	523	28	0.04	17.09	1.63	1.03	0.032	6.7
9.5	10	1996	7802	53	767	20429	342	32	7396	12463	1.59	238	4	473	17	0.04	16.25	1.60	1.05	0.031	<LOQ
10	10.5	1993	7794	46	721	20594	385	15	7601	12362	1.58	225	4	546	20	0.04	17.15	1.59	1.03	0.032	13.7
10.5	11	1990	7475	33	792	20059	351	30	7347	12565	1.59	274	7	547	24	0.04	15.87	1.68	1.02	0.031	6.1
11	11.5	1987	8248	59	735	21076	370	24	8561	13527	1.58	289	11	729	23	0.05	18.40	1.64	0.96	0.032	<LOQ
11.5	12	1984	9151	61	697	21241	312	33	10189	15106	1.67	342	18	1173	34	0.08	21.66	1.65	0.90	0.032	<LOQ
12	12.5	1980	8247	73	761	20873	232	29	8150	13214	1.59	279	3	600	24	0.05	17.36	1.60	1.01	0.033	<LOQ
12.5	13	1977	8774	57	747	21216	397	22	9969	14935	1.59	349	16	1086	34	0.07	20.00	1.70	0.88	0.032	20.8
13	13.5	1973	8381	63	744	20759	294	19	8656	14292	1.64	341	4	766	23	0.05	19.21	1.71	0.97	0.033	<LOQ
13.5	14	1969	9193	25	726	21197	255	22	10631	15751	1.63	369	12	1199	20	0.08	21.69	1.71	0.86	0.031	<LOD

14	14.5	1965	9230	78	759	21740	336	26	10223	15582	1.59	348	3	1057	24	0.07	20.52	1.69	0.90	0.032	<LOQ
14.5	15	1961	8609	53	778	20814	357	41	9284	14316	1.62	331	12	911	28	0.06	18.40	1.66	0.93	0.032	<LOQ
15	15.5	1957	8712	54	731	21374	293	26	9310	14750	1.61	337	5	860	29	0.06	20.19	1.69	0.94	0.032	<LOQ
15.5	16	1952	8657	73	786	21733	348	28	9483	15043	1.62	364	11	924	25	0.06	19.13	1.74	0.91	0.032	<LOQ
16	16.5	1947	8535	36	764	21403	303	18	9033	14443	1.61	358	8	765	35	0.05	18.91	1.69	0.94	0.031	<LOQ
16.5	17	1943	8344	66	783	21041	355	26	8553	14639	1.58	386	6	615	40	0.04	18.70	1.75	0.98	0.030	<LOQ
17	17.5	1938	7697	36	763	19911	340	15	7613	13480	1.60	355	8	522	24	0.04	17.66	1.75	1.01	0.030	<LOQ
17.5	18	1932	8081	37	756	20778	349	28	7897	13725	1.61	333	3	504	20	0.04	18.15	1.70	1.02	0.031	–
18	18.5	1927	8602	42	779	22177	371	17	8806	14694	1.58	403	2	695	28	0.05	18.87	1.71	0.98	0.029	–
18.5	19	1922	7376	39	694	20317	344	21	7186	12477	1.58	401	3	519	18	0.04	17.97	1.69	1.03	0.028	–
19	19.5	1916	6783	22	759	19755	360	22	6722	12493	1.57	489	3	504	25	0.04	16.46	1.84	1.01	0.028	–
19.5	20	1910	5445	53	654	16747	347	29	5471	12553	1.55	674	0	492	16	0.04	19.19	2.31	1.00	0.030	–
20	20.5	1904	4438	13	666	15346	308	28	3961	10879	1.54	606	5	334	16	0.03	16.32	2.45	1.12	0.032	–
20.5	21	1898	3548	28	613	13229	336	24	3251	9782	1.53	568	1	238	9	0.02	15.97	2.76	1.09	0.033	–
21	21.5	1892	3644	41	598	13904	278	30	3331	8568	1.52	538	0	261	18	0.03	14.33	2.35	1.09	0.034	–
21.5	22	1886	4354	2	595	15295	255	29	4170	9729	1.57	632	2	359	17	0.04	16.36	2.23	1.04	0.034	–
22	22.5	1879	4855	19	604	15828	284	22	4497	10632	1.59	627	0	375	5	0.04	17.62	2.19	1.08	0.035	–
22.5	23	1872	3873	50	572	14095	305	18	3686	9335	1.54	619	0	321	13	0.03	16.31	2.41	1.05	0.034	–
23	23.5	1865	4273	51	765	16759	304	38	4308	10539	1.48	711	28	441	13	0.04	13.78	2.47	0.99	0.034	–
23.5	24	1858	4093	65	638	15495	254	31	3694	9133	1.56	651	4	311	11	0.03	14.32	2.23	1.11	0.034	–
24	24.5	1851	2220	4	597	11508	337	25	1965	7393	1.48	610	3	158	3	0.02	12.38	3.33	1.13	0.039	–
24.5	25	1844	3921	23	669	14166	275	19	3515	9707	1.55	618	1	322	16	0.03	14.50	2.48	1.12	0.036	–
25	25.5	1836	3387	5	557	14674	285	29	3078	9262	1.57	805	0	322	7	0.03	16.63	2.74	1.10	0.039	–
25.5	26	1828	2899	8	553	13956	312	33	2702	7400	1.53	814	3	344	6	0.05	13.39	2.55	1.07	0.045	–
26	26.5	1821	2223	14	568	14135	311	28	2194	8637	1.57	1171	0	339	9	0.04	15.21	3.88	1.01	0.051	–
26.5	27	1813	3080	8	533	16229	293	38	2905	6822	1.57	918	3	442	9	0.06	12.79	2.22	1.06	0.053	–
27	27.5	1804	2555	15	505	14753	187	33	2651	7991	1.58	1052	0	541	9	0.07	15.83	3.13	0.96	0.058	–