Supporting Information

Spatial and Temporal Variation of Suspect Screening derived Cyanobacterial Secondary Metabolite Mixtures during Harmful Algal Bloom Events in Shallow Agroecosystem Lakes

Stanley W. Kohls,[ab] Lyndy Holdt,[cd] Nervana Metwali,[c] Corey D. Markfort,[ab] Peter S. Thorne,[c] Gregory H. LeFevre*[ab]

[a]Department of Civil & Environmental Engineering, University of Iowa, 4105 Seamans Center, Iowa City, IA 52242, United States; [b]IIHR-Hydroscience &Engineering, 100 C. Maxwell Stanley Hydraulics Laboratory, Iowa City, IA 52242, United States; [c]Department of Occupational and Environmental Health, The University of Iowa, Iowa City, Iowa 52242, United States; [d]Friends of Lakeside Lab, 1838 Highway 86, Milford, IA 51351, United States

* Corresponding Author: gregory-lefevre@uiowa.edu; Phone: +319 335 5655, Department of Civil and Environmental Engineering, 4105 Seamans Center for Engineering, University of Iowa, Iowa City IA, 52242, United States

Total Pages (Including this page): 51
Total number of Figures: 5
Total number of Tables: 19

Supporting Information Table of Contents:

Section S1: Site and Sample Descriptions

The two lakes sampled in this study are in Iowa, USA, with the sampling sites' coordinates shown in Table S1. Land use in the watersheds of both lakes is primarily cultivated agricultural and hay/pasture, with detailed land use information shown in Figure S2 and Figure S1, derived from the National Land Cover Database (NLCD). This land-use data was used to inform the classification of these lakes as strongly influenced by agroecosystems. Examples of the conditions of the sampling sites during sampling are shown in Figure S3.

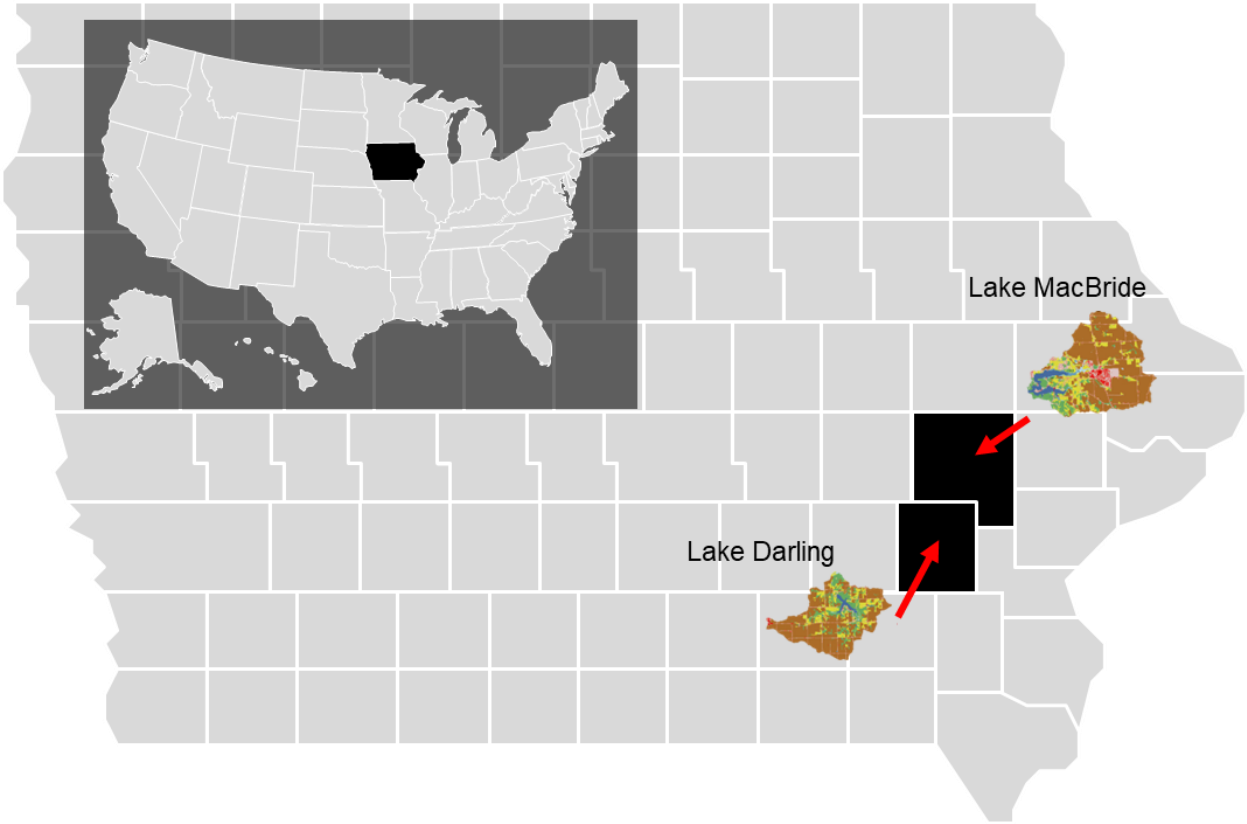
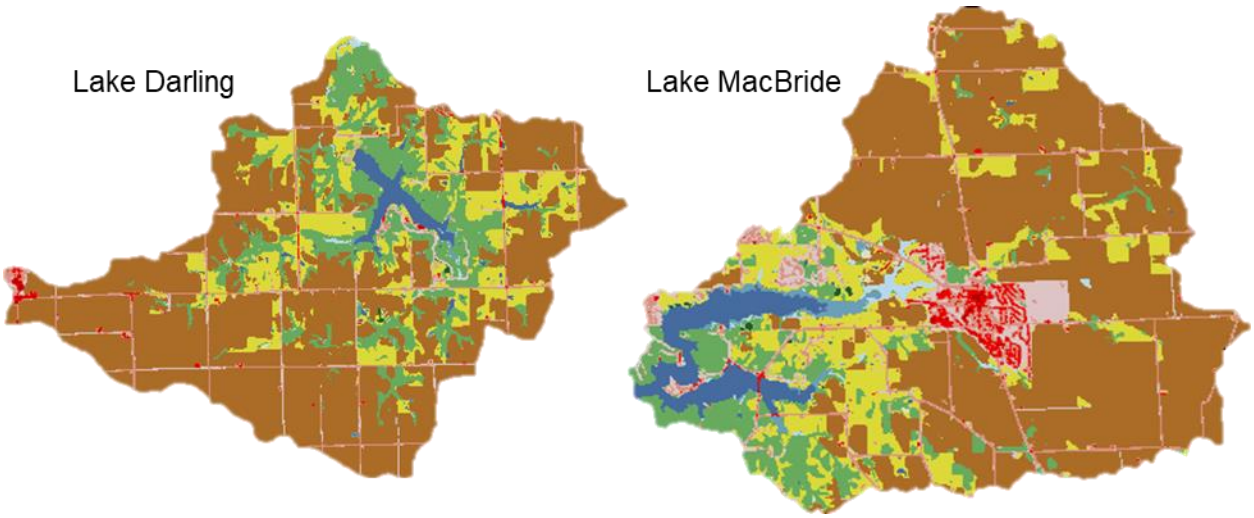


Figure S1: Visual Representation of Site Locations

Table S1: Site descriptions/sample collection

	Location	
Lake MacBride	**Latitude**	**Longitude**
North 1	41.797796	-91.563066
North 2	41.797742	-91.564402
South 1	41.7916511	-91.5500512
South 2	41.7915541	-91.5512981
Lake Darling	41.190278	-91.8925



Watershed	Open Water	Developed	Baren Land	Forest	Shrub/ Scrub	Herbaceuous	Hay/ Pasture	Cultivated Crops	Wetland
Lake MacBride (HUC 12-070802071008)	4.66%	12.86%	0.01%	9.74%	0.05%	0.18%	14.78%	55.77%	1.96%
Lake Darling (HUC 12-070801070304	2.51%	6.26%	0.06%	15.71%	0.24%	0.15%	15.44%	59.08%	0.56%

Figure S2: Land use (from the NLCD 2023) for watersheds encompassing Lake Darling and Lake MacBride displayed visually and numerically.

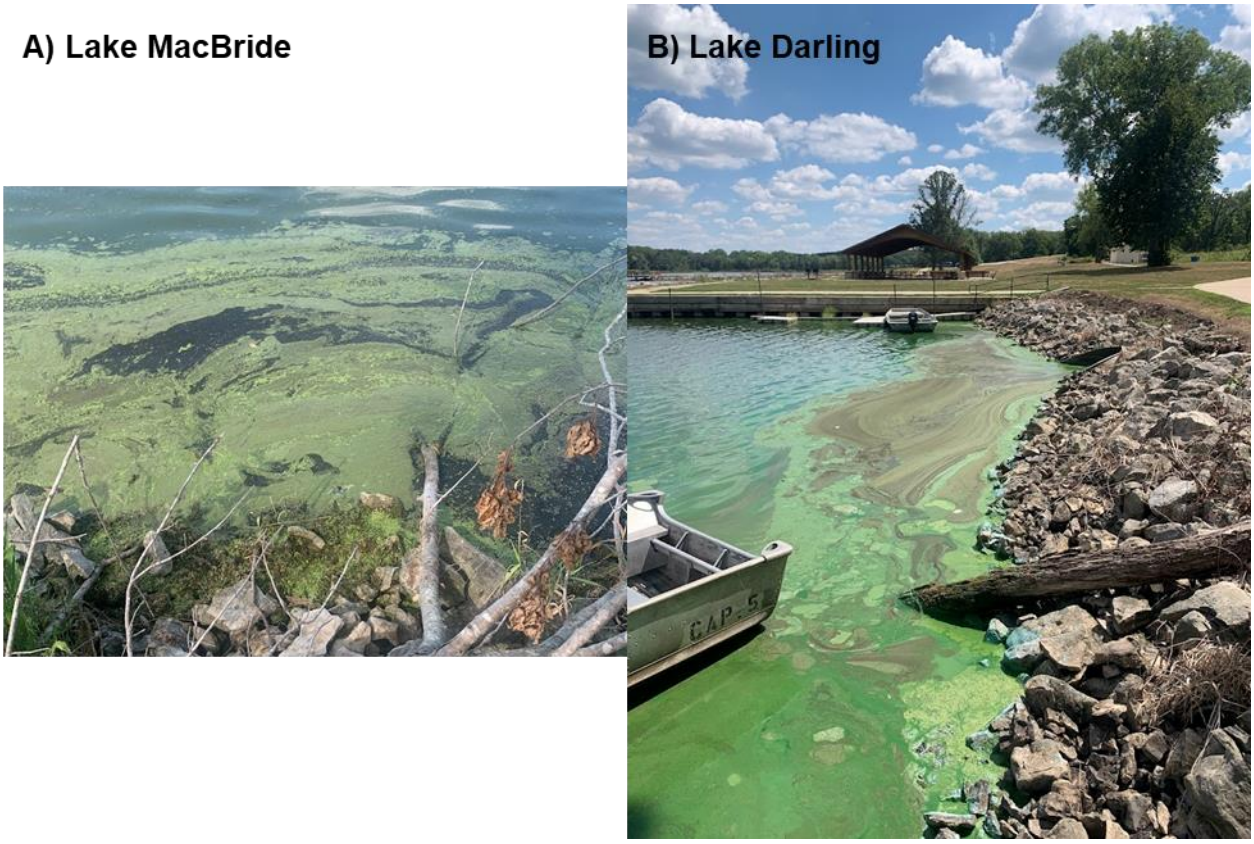


Figure S3: Example of conditions at (A) Lake MacBride and (B) Lake Darling.

Section S2: Analytical Methods and Chemicals

S2.1 Sample Collection:

The sampling protocol was adapted from the Ohio EPA Method 546, specifically the determination of total MCs in ambient water, and the sample collection methods from the Eurofins Abraxis Microcystins-ADDA Enzyme-Linked Immunosorbent Assay (ELISA) kit.[1] At a beach site, surface water samples were taken at both ends of the designated swimming area and at the center. At each area (left, center, and right), the polyethylene sample dipper was washed in the water three times, to remove surface scum, at approximately 1 meter off the shore before taking about 200 mL of water to pour into a pre-labeled and sterilized 250 mL amber glass jar. Samples were immediately put on ice in a cooler to be transported back to the laboratory, where they were stored in -20°C freezers.

S2.2 ELISA

Analysis was performed following instructions given by the kit manufacturer, eurofins Abraxis, which can be found online. Kits used were the Microcystins-ADDA ELISA (product no. 520011, Microcystins-ADDA ELISA) the Anatoxin-a Receptor Binding Assay (product no. 520050. Anatoxin-a ELISA), and the Saxitoxin (PSP) ELISA, Microtiter Plate (product no. 52255B, Saxitoxin (PSP) ELISA). The Microcystins-ADDA kit measured total MCs and nodularins (NOD) in a sample, based on detection of a characteristic feature common to microcystin and nodularin congeners (structural variants), specifically, the ADDA (3-amino-9- methoxy-2,6,8-trimethyl-10-phenyldeca-4,6-dienoic acid) amino acid side chain. The kit was calibrated against the specific congener, MC-LR for standardization. The ANTX assay is a direct 60 competitive ELISA based on the recognition of anatoxin-a by a monoclonal antibody, as well as the SXT assay but using more than one specific antibodies. Five standards in duplicate were used, ranging from zero to five parts per billion (ppb, equivalent to µg/L). As well as a Quality Control Standard (QCS) within the range of 0.75 ± 0.185 ppb and a Low Calibration Range Check (LCRC) control within the range of 0.40 ± 0.16 ppb. There were no issues with the standards or controls in any assay performed. The resulting microtiter plate was read using a SpectraMax plate reader at a wavelength of 450 nm (Molecular Devices), analyzed using the associated Softmax Pro 5.4 software. Samples below 0.15 µg/L or above 5 µg/L were out of the range of detection of the assay for both the MC and ANTX kit. The lower LOD for the SXT kit was 0.02 µg/L. MC samples above the limit of detection (LOD) were diluted 5- and 10-fold, while ANTX samples were diluted 100-, 1000-, and 10,000-fold to accurately quantify the toxin concentration.

S2.3 Suspect Screening

Chromatography parameters used for non-target analysis. A Thermo Vanquish Flex UHPLC system was fitted with an Agilent C18 column for peak separation prior to mass spectrometry. Guard column: Agilent Eclipse Plus C18, 2.1×5 mm, 1.8 µm. Analytical column: Agilent Poroshell 120, EC-C18 2.7 µm, 2.1×100 (PN 695775-902(T)). Mass spectrometry was performed in succession with the UHPLC. The HRMS setup was a Thermo Scientific Q Exactive hybrid quadrupole-Orbitrap mass spectrometer. Each sample was injected in duplicate using this polarity-switching full-scan method.

A summary of where information relevant to the BP4NTA (Best Practices for Non-Targeted Analysis; https://nontargetedanalysis.org/) SRT (Study Reporting Tool) "scoresheet" is found in Table S2.[22] The full SRT scoring document was shared at the time of review, and the original document and related paper can be found in the references.

Table S2: Locations of information relevant to the BP4NTA SRT

Section	Category	Sub-Category	Rationale for score
Methods	Study Design	Objectives & Scope	Described in the last paragraph of the introduction in the main text for objectives, hypotheses, and scope
		Sample Information & Preparation	Described in the main text under the Methods subheadings "Experimental Design" and "Analytical Methods."
		QC Spikes & Samples	NA to this paper
	Data Acquisition	Analytical Sequence	Describe in the Analytical Methods U*HPLC-HRMS* section and Table S6
		Chromatography	Describe in the Analytical Methods UHPLC-HRMS section and Table S3
		Mass Spectrometry	Describe in the Analytical Methods UHPLC-HRMS section and Tables S4 and S5
	Data Processing & Analysis	Data Processing	Described in the Data processing and analysis section, Table S7, and Figures S5 and S6
		Statistical & Chemometric Analysis	Described in the Data processing and analysis section
		Annotation & Identification	Described in the Data processing and analysis section

Results	Data Outputs	Statistical & Chemometric Outputs	Described in Data processing and analysis. Shown in Figure 1 (HCA), use of cos(θ) metric
		Identification & Confidence Levels	Described in Data processing and analysis (note that the Schamaski framework was not used), Details of compounds in Tables S9-S14
	QA/QC Metrics	Data Acquisition QA/QC	Described in data processing and analysis, Analytical Methods (UHPLC-HRMS), and reported in Tables S10 and S14
		Data Processing & Analysis QA/QC	Referenced throughout the Methods and Results of the main text and in the SI, not in one specific location.

Table S3: Chromatography gradient

A= Water with 0.1% Formic Acid

B= Acetonitrile with 0.1% Formic Acid

Time Interval	Time (min)	Flow (mL/min)	%B
1	0.00	0.40	10.0
2	4.00	0.40	20.0
3	8.00	0.40	40.0
4	10.00	0.40	60.0
5	15.00	0.40	90.0
6	15.20	0.40	10.0
7	20.00	0.40	10.0

Table S4: Full Scan MS Acquisition Parameters.
Mass spectrometry was performed in succession with the UPLC listed above. The HRMS setup was a Thermo Scientific Q Exactive hybrid quadrupole-Orbitrap mass spectrometer. Each sample was injected in duplicate using this polarity-switching full-scan method.

Polarity Switching - Full Scan					
Method of Q Exactive		***Setup***		***Experiments***	
Overall method settings		**Tunefiles**		**Full MS – SIM**	
Global Settings		**General**		**General**	
Use lock masses	off	Switch Count	0	Runtime	0 to 15 min
Lock mass injection	–	Base Tunefile C:\Xcalibur\methods\300ul-min_020122.mstune		Polarity	positive
Chrom. peak width (FWHM)	3 s	**Contact Closure**		In-source CID	0.0 eV
Time		**General**		**Full MS – SIM**	
Method duration	20.00 min	Used	FALSE	Microscans	1
Customized Tolerances (+/-)		Start in Closed	TRUE	Resolution	70,000
Lock Masses	–	Switch Count	0	AGC target	1e6
Inclusion	10.00 ppm	**Syringe**		Maximum IT	200 ms
Exclusion	–	**General**		Number of scan ranges	1
Neutral Loss	–	Used	FALSE	Scan range	100 to 1500 m/z
Mass Tags	–	Start in OFF	TRUE	Spectrum data type	Profile
Dynamic Exclusion	–	Stop at end of run	FALSE	**Full MS – SIM**	
		Switch Count	0	**General**	
		Pump setup		Runtime	0 to 15 min
		Syringe type	Hamilton	Polarity	negative
		Flow rate	3.000 µL/min	In-source CID	0.0 eV
		Inner diameter	2.303 mm	**Full MS – SIM**	
		Volume	250 µL	Microscans	1
		Divert Valve A		Resolution	70,000
		General		AGC target	1e6
		Used	TRUE	Maximum IT	200 ms
		Start in 1-2	TRUE	Number of scan ranges	1
		Switch Count	1	Scan range	100 to 1500 m/z
		Element 1		Spectrum data type	Profile
		At	15.00 min		
		Switches to	1-6		

Table S5: ddMS2 Acquisition Parameters

Positive or Negative ddMS2					
Method of Q Exactive		***Experiment***		***Setup***	
Overall method settings		**Full MS / dd-MS² (TopN)**		**Tunefiles**	
Global Settings		**General**		**General**	
Use lock masses	off	Runtime	0 to 15 min	Switch Count	0
Lock mass injection	–	Polarity	positive or negative	Base Tunefile	C:\Xcalibur\methods\300ul-min_020122.mstune
Chrom. peak width (FWHM)	3 s	In-source CID	0.0 eV	**Contact Closure**	
Time		Default charge state	1	**General**	
Method duration	20.00 min	Inclusion	–	Used	FALSE
Customized Tolerances (+/-)		Exclusion	–	Start in Closed	TRUE
Lock Masses	–	Tags	–	Switch Count	0
Inclusion	10.00 ppm	**Full MS**		**Syringe**	
Exclusion	–	Microscans	1	**General**	
Neutral Loss	–	Resolution	70,000	Used	FALSE
Mass Tags	–	AGC target	1.00E+06	Start in OFF	TRUE
Dynamic Exclusion	–	Maximum IT	200 ms	Stop at end of run	FALSE
		Number of scan ranges	1	Switch Count	0
		Scan range	100 to 1500 m/z	**Pump setup**	
		Spectrum data type	Profile	Syringe type	Hamilton
		dd-MS² / dd-SIM		Flow rate	3.000 µL/min
		Microscans	1	Inner diameter	2.303 mm
		Resolution	17,500	Volume	250 µL
		AGC target	1.00E+05	**Divert Valve A**	
		Maximum IT	50 ms	**General**	
		Loop count	3	Used	TRUE
		MSX count	1	Start in 1-2	TRUE
		TopN	3	Switch Count	1
		Isolation window	1.0 m/z	**Element 1**	
		Isolation offset	0.0 m/z	At	15.00 min
		Scan range	200 to 2000	Switches	1-6

	m/z	to	
Fixed first mass	–		
(N)CE / stepped (N)CE	nce: 20, 40, 60		
Spectrum data type	Centroid		
dd Settings			
Minimum AGC target	4.00E+03		
Intensity threshold	8.00E+04		
Apex trigger	–		
Charge exclusion	–		
Multiple charge states	all		
Peptide match	–		
Exclude isotopes	on		
Dynamic exclusion	3.0s		
If idle ..	pick others		

Table S6: Data Acquisition sequences

Lake Darling Acquisition Sequence		Lake MacBride Acquisition Sequence	
1	Blank	1	Blank
2	LD1122_full_1	2	HAB_MAC_081722_1_PM_full_1
3	LD1122_full_2	3	HAB_MAC_081722_1_PM_full_2
4	LD1122_ddms2_neg	4	HAB_MAC_081722_1_PM_ddms2_neg
5	LD1122_ddms2_pos	5	HAB_MAC_081722_1_PM_ddms2_pos
6	LD1220_full_1	6	HAB_MAC_081722_2_PM_full_1
7	LD1220_full_2	7	HAB_MAC_081722_2_PM_full_2
8	LD1220_ddms2_neg	8	HAB_MAC_081722_2_PM_ddms2_neg
9	LD1220_ddms2_pos	9	HAB_MAC_081722_2_PM_ddms2_pos
10	LD1400_full_1	10	HAB_MAC_081722_AP4_PM_full_1
11	LD1400_full_2	11	HAB_MAC_081722_AP4_PM_full_2
12	LD1400_ddms2_neg	12	HAB_MAC_081722_AP4_PM_ddms2_neg
13	LD1400_ddms2_pos	13	HAB_MAC_081722_AP4_PM_ddms2_pos
14	Blank2	14	HAB_MAC_081722_AP5_PM_full_1
15	LD1415_full_1	15	HAB_MAC_081722_AP5_PM_full_2
16	LD1415_full_2	16	HAB_MAC_081722_AP5_PM_ddms2_neg
17	LD1415_ddms2_neg	17	HAB_MAC_081722_AP5_PM_ddms2_pos
18	LD1415_ddms2_pos	18	ACN2
19	LD1430_full_1		
20	LD1430_full_2		
21	LD1430_ddms2_neg		
22	LD1430_ddms2_pos		
23	LD1445_full_1		
24	LD1445_full_2		
25	LD1445_ddms2_neg		
26	LD1445_ddms2_pos		
27	LD1500_full_1		
28	LD1500_full_2		
29	LD1500_ddms2_neg		
30	LD1500_ddms2_pos		
31	Pooled_full_1		
32	Pooled_full_2		
33	Pooled_ddms2_neg		
34	Pooled_ddms2_pos		
35	Blank3		

Section S3: Data Processing

S3.1 Compound Discoverer

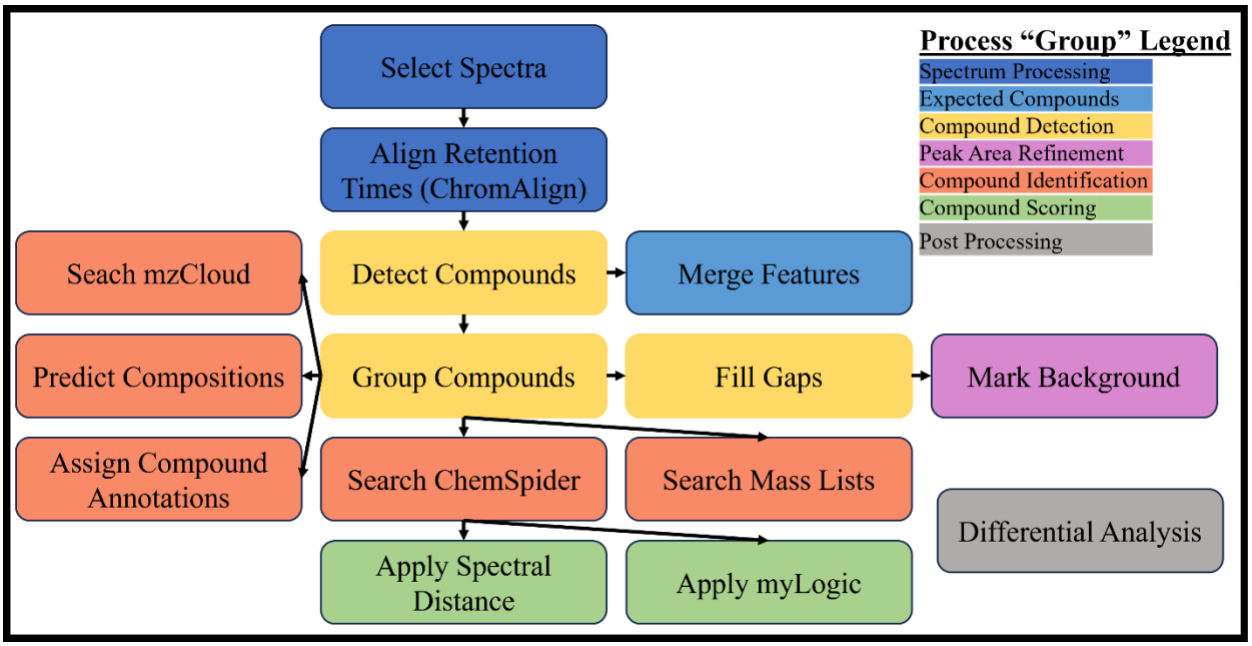


Figure S4: Workflow Tree in Compound Discover used for the analysis of all datasets. The workflow tree is adapted from the preset workflow "*Environmental w Stats Unknown ID w Online and Local Database Searches"*. The two datasets were processed individually, with one for Lake Darling and Lake MacBride, respectively.

Table S7: Parameters used for Compound Discover Analysis, applied using the workflow shown in **Figure S4.**

Select Spectra	
1. Spectrum Properties Filter	
Lower RT Limit	0
Upper RT Limit	15.0
2. Scan Event Filters	
Polarity Mode	Any
Align Retention Times (ChromAlign)	
1. General Settings	
Reference File	First sample in list
Detect Compounds	
1. General Settings	
Mass Tolerance [ppm]	5 ppm
Min. Peak Intensity	100000
Use Most Intense Isotope Only	True
3. Peak Detection	
Chromatographic S/N Threshold	1.5
Remove Baseline	False
4. Isotope Pattern Detection	
Group Isotopes for	Br; Cl
Ions	33/33 Checked
Merge Features	
1. Peak Consolidation	
Mass Tolerance	5 ppm
RT Tolerance [min]	0.1
Group Compounds	
1. General Settings	
Mass Tolerance	5 ppm
RT Tolerance [min]	0.1
Minimum Valley %	10
Align Peaks	False
Preferred Ions	[M+H]+1; [M-H]-1
Area Integration	Most Common Ion
2. Peak Rating Contributions	
Area Contribution	3
CV Contribution	10
FWHM to Base Contribution	5
Jaggedness Contribution	5
Modality Contribution	5
Zig-Zag Index Contribution	5
3. Peak Rating Filter	

Peak Rating Threshold	4.5
Number of Files	3

Fill Gaps	
1. General Settings	
Mass Tolerance	5 ppm
S/N Threshold	1.5

Mark Background Compounds	
1. General Settings	
Max. Sample/Blank	5
Max. Blank/Sample	0
Hide Background	True

Search mzCloud	
1. General Settings	
Compound Classes	All
Library	Autoprocessesd; Reference
Search MSn Tree	False
2. DDA Search	
Identity Search	HighChem HighRes
Match Activation Type	True
Match Activation Energy	Match with Tolerance
Activation Energy Tolerance	20
Apply Intensity Threshold	True
Similarity Search	None
Match Factor Threshold	50
3. DIA Search	
Use DIA Scans for Search	True
Max. Isolation Width [Da]	500
Match Activation Type	False
Match Activation Energy	Any
Activation Energy Tolerance	100
Apply Intensity Threshold	True
Match Factor Threshold	20

Predict Compositions	
1. Prediction Settings	
Mass Tolerance	5 ppm
Min. Element Counts	C H
Max. Element Counts	C90 H190 Br3 Cl8 F18 N10 O18 P3 S5
Min. RDBE	0
Max. RDBE	40
Min. H/C	0.1
Max. H/C	3.5
Max. # Candidates	10
2. Pattern Matching	

Intensity Tolerance [%]	30
Intensity Threshold [%]	0.1
S/N Threshold	3
Use Dynamic Recalibration	True
3. Fragments Matching	
Use Fragments Matching	True
Mass Tolerance	5 ppm
S/N Threshold	3

Assign Compound Annotations	
1. General Settings	
Mass Tolerance	5 ppm
2. Data Sources	
Data Source #1	MassList Search
Data Source #2	Predicted Compositions
Data Source #3	mzCloud Search
Data Source #4	mzVault Search
Data Source #5	ChemSpider Search
3. Scoring Rules	
Use mzLogic	True
Use Spectral Distance	True
SFit Threshold	20
SFit Range	20
4. Reprocessing	
Clear Names	False

Search ChemSpider	
1. Search Settings	
Database(s)	ACToR: Aggregated Computational Toxicology Resource; DrugBank; EAWAG Biocatalysis/Biodegradation Database; EPA DSSTox; EPA Toxcast; FDA UNII – NLM: Toxin, Toxin-Target Database
Search Mode	By Formula or Mass
Mass Tolerance	5 ppm
Max. # of results per compound	20
Max. # of Predicted Compositions to be searched per Compound	3

Search Mass Lists	
1. Search Settings	
Mass Lists	CyanoMetDD_V3
Use Retention Time	False
RT Tolerance [min]	0.5
Mass Tolerance	5 ppm

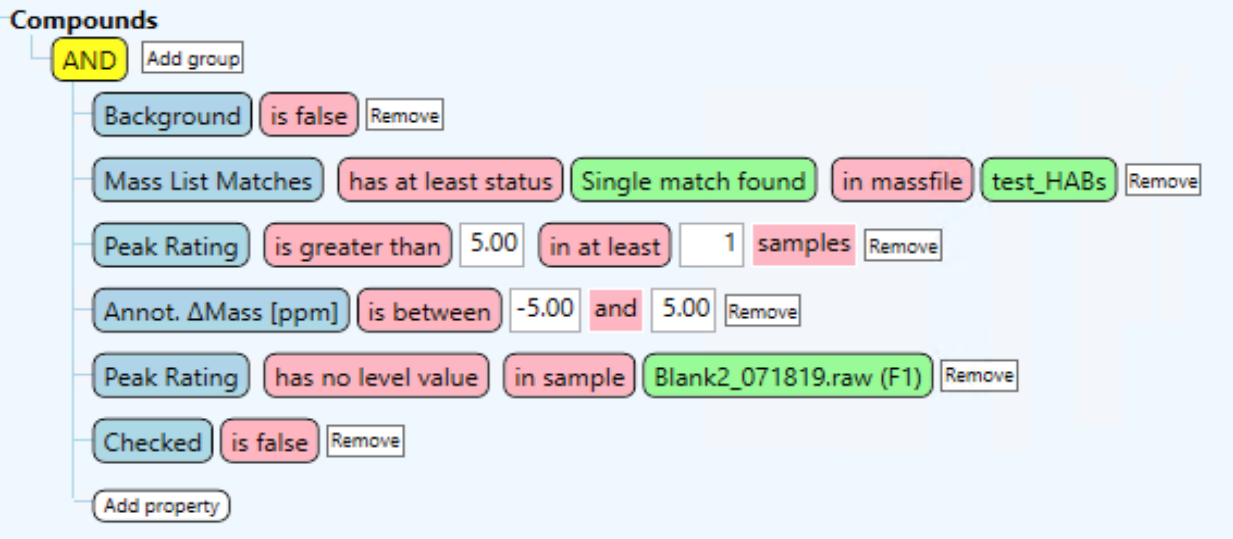
Apply mzLogic

1. Search Settings	
Max. # Compounds	0
Max. # mzCloud Similarity Results to consider per Compound	10
Match Factor Threshold	30

Apply Spectral Distance	
1. Pattern Matching	
Mass Tolerance	5 ppm
Intensity Tolerance [%]	30
Intensity Threshold [%]	0.1
S/N Threshold	3
Use Dynamic Recalibration	True

Differential Analysis	
1. General Settings	
Log10 Transform Values	True
2. Peak Rating Contributions	
Update Peak Rating	False
Area Contribution	3
CV Contribution	10
FWHM to Base Contribution	5
Jaggedness Contribution	5
Modality Contribution	5
Zig-Zag Index	5

Figure S5: Compound Discover filter set for creation of working feature lists. Features were "checked" when a split peak resulted in duplicate feature entries at different retention times, and the highest-quality peak was not removed.



S3.2 MS2Query

As a secondary library-search method, the MS2Query[3] tool was used to cross-check our MS2 spectra against a spectral library outside mzCloud. From the final filtered data lists, raw spectral data (m/z, intensity pairs) were exported, along with the relevant metadata (RT, Precursor m/z, charge, and ion mode). Using the code from the GitHub repository of the initial MS2query publication, the data were processed with the original GNPS spectral database[4] as the basis for spectral matching. In interpreting the output from MS2query, a prediction score (0-1) > 0.9 and a minimal precursor m/z difference indicate that the proposed ID is a likely candidate when compared to the provided MS2 spectra. While the GNPS database does not have spectra for several of the compounds we initially identified from our results, it does have spectra for some of the more "important" compounds, such as certain Microcystins and Anabaenopeptins. While this tool may assign different IDs to some of our proposed CSMs, this could be due to similar MS2 spectra, the absence of the expected compound in the database, or an incorrect initial tentative ID. However, because this tool was used as a secondary verification against spectral libraries, its results were not used to change IDs; instead, correlated results were treated as elevating the confidence of our initial identifications.

Section S4: Supplemental Results Figures, Tables, and Statistics

Table S8: Anatoxin-a and Microcystin concentrations at Lake MacBride site

Sample ID	Sample Name	Anatoxin-a Results (ppb)	Microcystin Results (ppb)
HAB-MAC-081722-1-PM	North 1	0.049	2.19
HAB-MAC-081722-2-PM	North 2	0.044	2.035
HAB-MAC-081722-AP4-PM	South 2	9.532	36.864
HAB-MAC-081722-AP5-PM	South 1	0.431	0.646

Table S9: Summary of Feature Identification Confidence for each site, based on the Schymanski scale

ID Confidence	ID Confidence Meaning (as applied to this study)	Lake MacBride	Lake Darling
Level 1	Matched with a reference standard (Not applicable to this study)	NA	NA
Level 2(a)	Spectral Library Match	0	0
Level 2(b)	Best fit compound match with MS2Query	2	1
Level 3(a)	Theoretical MS2 Match with good *in silico* match	8	2
Level 3(b)	MS2 present, but poor *in silico* match or low quality MS2	11	8
Level 4	Molecular Formula and Mass List Match with CyanoMetDB only	9	8
Level 5	Features Unmatched by the CyanoMetDB	990	1141

Table S10: MS2Query results for positive mode spectra

query spectrum expected ID	ms2query model prediction score	precursor mz difference	precursor mz query spectrum	precursor mz analog	inchikey	analog compound name
MC-RR (LMB)	0.9479	0.0003	519.7903	519.79	JIGDOBKZMULDHS	Microcystin RR
MC-RR (LD)	0.942	0.0008	519.7892	519.79	JIGDOBKZMULDHS	Microcystin RR
Santacruzamate A (LD)	0.9305	0.0003	279.1588	279.1591	DOIRQSBPFJWKBE	Di-n-butyl phthalate
Koshikalide (LMB)	0.9239	0.0004	295.2264	295.226	LUZSWWYKKLTDHU	9-Oxo-10E,12Z-octadecadienoic acid
Palythene (LD)	0.915	0.0001	268.1039	268.104	OIRDTQYFTABQOQ	ADENOSINE
35-O-β-6 amino-6-deoxyglucopyranosyl-bacteriohopanetetrol (LMB)	0.9095	0.0004	708.5396	708.54	AGAUYSZNDYQXOM	Bacteriohopanetetrol cyclitol ether
4-Oxo-beta-apo-13-Carotenone (LMB)	0.6649	91.9746	273.1843	365.1589	JSKXQQKSUOVSKS	7,8-Dihydroxy calonectrin
Palmyrrolinone (LD)	0.5869	0.0005	238.1069	238.1074	GOVWOKSKFSBNGD	Ethopabate
Palythine (LMB)	0.586	0.0008	245.0992	245.1	YBJHBAHKTGYVGT	biotin
Palmyrrolinone (LMB)	0.4912	195.0137	238.1073	433.121	SUHJYAOFQCJYHR	(2S)-N-(2,4-difluorophenyl)-1-(4,5-dimethoxy-3-oxo-1H-2-benzofuran-1-yl)-5-oxopyrrolidine-2-carboxamide"
Radisoumin B (LMB)	0.4677	17.0157	445.2436	428.2279	AABILZKQMVKFHP	Lasiocarpine N-oxide
Tricholactone (LD)	0.3669	129.1157	255.1587	384.2745	XPDBYTRMUWQPTF	SCHEMBL21808143
cis-Carboxydihydroanatoxin-a (LMB)	0.3188	216.1226	212.1644	428.287	STIBDTVUVQTFGW	p(El1-co-MeOx9+CH3)H+2n

Table S11: MS2Query results for negative mode spectra

query spectrum expected ID	ms2query model prediction score	precursor mz difference	precursor mz query spectrum	precursor mz analog	inchikey	analog compound name
Anabaenopeptin A (LMB)	0.9585	0.0002	842.4096	842.4094	JWVKNHDSRADFFA	Anabaenopeptin A
[D-Asp3]MC-HilR (LMB)	0.9585	0.0006	993.5421	993.5415	ZYZCGGRZINLQBL	MCLR
Koshikalide (LD)	0.9569	0.0002	293.1758	293.176	NLDDIKRKFXEWBK	6-Gingerol CollisionEnergy:205060
Malyngic acid (LMB)	0.9569	0.0004	327.2176	327.218	MKYUCBXUUSZMQB	(10E,15E)-9,12,13-trihydroxyoctadeca-10,15-dienoic acid
15,16-dihydrosacrolide A (LMB)	0.6195	0	309.2071	309.2071	UYQGVDXDXBAABN	13-HpOTrE
Sacrolide A (LMB)	0.6181	0	307.1915	307.1915	BYWWNDLILWPPJP	albocycline
7-methoxy-9-methylhexadecadienoic acid (LD)	0.557	0.0004	295.2275	295.2279	FBUKMFOXMZRGRB	9,10-EODE
Gloeolactone (LD)	0.5371	510.2232	291.1965	801.4197	WCHBFWOEFOZHMK	Ziyuglycoside I
Malyngic acid (LD)	0.5113	0.0008	327.2176	327.2168	NIOKCFABUMZUDL	FA 18:2+3O
15,16-dihydrosacrolide A (LD)	0.4345	462.2595	309.2069	771.4664	YPZYGIQXBGHDBH	Salinomycin, Sodium
Deoxygadusol (LMB)	0.4345	70.1146	187.0612	257.1758	MBGYSHXGENGTBP	Mono(2-ethylhexyl) adipate
Gloeolactone (LMB)	0.4345	170.0061	291.1967	461.2028	FSTIKTPQGMHLFJ	Urceolide
Puna'auic acid (LMB)	0.4345	272.0697	311.1865	583.2562	BPYKTIZUTYGOLE	Bilirubin
Puna'auic acid (LD)	0.3854	276.3183	311.1862	587.5045	QETRZBRSOYIFDZ	FAHFA 38:3
12-hydroxy-2-oxo-11-epi-hinesol (LMB)	0.316	312.3392	251.1653	563.5045	PGKKGBQMNNEIHV	FAHFA 36:1
478-Da MAA (LD)	0.2872	43.1558	477.1722	520.328	MHOZNCAVRHAOKT	Glu-CDCA
7-methoxy-9-methylhexadecadienoic acid (LMB)	0.2872	307.138	295.2278	602.3658	CBEIHVWVTGRGDR	CHEBI:181256
Hapalindole A-formamide (LMB)	0.2872	402.266	355.1581	757.4241	PGRWPWIGJLNDHN	2-[[2-[2-[2-[2-[[2-[2-(2-hydroxypropanoylamino)-3-methylbutanoyl]oxy-3-methylbutanoyl]amino]-3-methylbutanoyl]oxypropanoylamino]-3-methylbutanoyl]oxy-3-methylbutanoyl]amino]-3-methylbutanoic acid
Sacrolide A (LD)	0.2872	108.074	306.9188	414.9928	YURRUMFCDGTGRW	2-(benzoylamino)ethyl 5-(aminosulfonyl)-2,4-dichlorobenzoate

Table S12: Tentatively Identified Cyanometabolites from Lake MacBride and their associated information as reported in the CyanMetDB.

Identified Features	CyanoMet_ID	Compound Class	Alternative Class name	Genus initially reported in	Species initially reported in	Original ID Location
12-hydroxy-2-oxo-11-epi-hinesol	CyanoMetDB_2209	other linear non-peptide	Spirovetivane	*Calothrix*		Hawaii (freshwater)
15,16-dihydrosacrolide A	CyanoMetDB_1632	other cyclic non-peptide		*Aphanothece*	*sacrum*	Japan (freshwater)
35-O-β-6 amino-6-deoxyglucopyranosyl-bacteriohopanetetrol	CyanoMetDB_0823	other linear non-peptide	Triterpenoid	Synechocystis	sp.	Europe (Lab Culture)
4-Oxo-beta-apo-13-Carotenone	CyanoMetDB_2222	other linear non-peptide		*Anabaena*	*cylindrica*	UK (freshwater)
7-methoxy-9-methylhexadecadienoic acid	CyanoMetDB_1629	other linear non-peptide		*Lyngbya*	*aestuarii*	Hawaii (marine)
[D-Asp3]MC-HilR	CyanoMetDB_1821	Microcystin		*Planktothrix; Dolichospermum/ Anabaena*	*[Planktothrix]sp.; [Dolichospermum]sp.*	Germany (freshwater))
Aeruginosin 205A	CyanoMetDB_1122	Aeruginosin		*Oscillatoria/ Planktothrix*	*agardhii*	Japan (lab culture)
Aeruginosin 850	CyanoMetDB_0827	Aeruginosin	Depsipeptide	*NA*		Japan (marine)
Anabaenopeptin 871	CyanoMetDB_0668	Anabaenopeptin		*Nostoc*	*n.a.*	Brazil (Freshwater)
Anabaenopeptin A	CyanoMetDB_0760	Anabaenopeptin		*Anabaena/ Dolichospermum*	*flos aquae*	Ohio, USA (Lab Culture)
Anabaenopeptin B	CyanoMetDB_0867	Anabaenopeptin		*Anabaena/ Dolichospermum*	*flos aquae*	Japan (lab culture)
Caldorin	CyanoMetDB_1600	other linear non-peptide		*Caldora*	*penicillata*	Japan (marine)
cis-Carboxydihydroanatoxin-a	CyanoMetDB_2550	Anatoxin-a		*Oscillatoria*	*n.a.*	France (Freshwater)
Dehydroradiosumin	CyanoMetDB_1509	other linear peptide		*Anabaena*	*cylindrica*	Japan (lab Culture)
Deoxygadusol	CyanoMetDB_0	other linear non-peptide	Mycosporine-	*n.a.*		Himalayan Lake

	296		like amino acids			(Freshwater)
Radiosumin B	CyanoMetDB_2866	other linear non-peptide		Dolichospermum¬†	planctonicum	Finland (freshwater)
Gloeolactone	CyanoMetDB_1638	other linear non-peptide		*Gloeotrichia*		Montana, USA (freshwater)
Hapalindole A-formamide	CyanoMetDB_2299	other cyclic peptide		Hapalosiphon	sp.	Commercial product, USA (freshwater)
Hermitamide A	CyanoMetDB_1487	other linear non-peptide	Lyngbya toxin	*Lyngbya/Moorea*	*majuscula/producens*	Papua New Guinea (marine)
Koshikalide	CyanoMetDB_1654	other linear non-peptide		*Lyngbya*		Japan (marine)
Malyngic acid	CyanoMetDB_1627	other linear non-peptide		*Lyngbya/Moorea*	*majuscula/producens*	Hawaii (marine)
MC-RR	CyanoMetDB_1802	Microcystin		*Microcystis*	*viridis*	Japan (freshwater)
Mycosporine-GABA	CyanoMetDB_1725	other linear non-peptide	Mycosporine-like amino acids	Nostoc	commune	Japan (terresital)
Nakienone A	CyanoMetDB_1746	other linear non-peptide		*Synechocystis*		Japan (marine)
Palmyrrolinone	CyanoMetDB_2408	other cyclic peptide	Molluscicide	*cf. Oscillatoria and Hormoscilla*		Palmyra Atoll (USA, Marine)
Palythine	CyanoMetDB_1769	other linear non-peptide	Mycosporine-like amino acids	Nostoc	sp.	India (freshwater)
Puna'auic acid	CyanoMetDB_1628	other linear non-peptide		*Pseudanabaena*		Pacific Ocean (marine)
Sacrolide A	CyanoMetDB_1635	other cyclic non-peptide		*Aphanothece*	*sacrum*	Japan (freshwater)
Santacruzamate A	CyanoMetDB_0233	other linear non-peptide		*Symploca*	*n.a.*	Coiba National Park, Panama (marine)
Tricholactone	CyanoMetDB_0040	other cyclic peptide	other cyclic depsipeptide			Texas, USA (freshwater)

Table S13: Tentatively Identified Cyanometabolites [with confidence levels shown next to name] from Lake MacBride with critical suspect screening results†

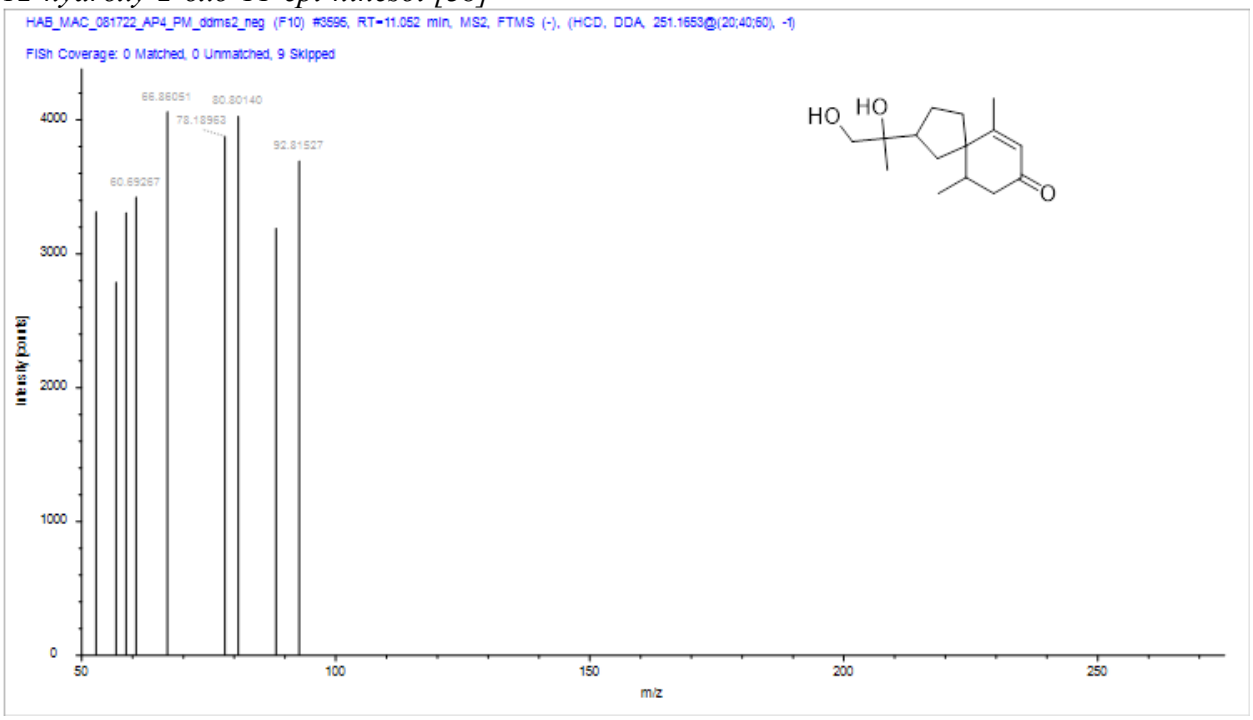
Identified Features [conf. level]	Calc. MW	m/z	Δppm	RT [min]	MS2	FISh Score (when Applicable)
12-hydroxy-2-oxo-11-epi-hinesol [3b]	252.17257	251.1653	0.12	11.049	DDA for preferred ion	NA (MS2 Not Abundant Enough)
15,16-dihydrosacrolide A [3b]	310.21412	309.20718	-0.7	10.769	DDA for preferred ion	0.00
35-O-β-6 amino-6-deoxyglucopyranosyl-bacteriohopanetetrol [3a]	707.53301	708.54031	-0.86	14.74	DDA for preferred ion	61.90
4-Oxo-beta-apo-13-Carotenone [3b]	272.17728	273.18456	-1.51	11.211	No MS2	68.25
7-methoxy-9-methylhexadecadienoic acid [3b]	296.2352	295.22792	0.18	12.366	DDA for preferred ion	NA (MS2 Not Abundant Enough)
[D-Asp3]MC-HilR [3b]	994.54672	995.5561	0.18	8.679	DDA for preferred ion	0.00
Aeruginosin 205A [4]	804.31238	805.31965	-0.86	6.139	No MS2	NA
Aeruginosin 850 [4]	850.46885	851.47596	0.07	6.916	No MS2	NA
Anabaenopeptin 871 [4]	871.44765	870.44147	-0.39	8.694	No MS2	NA
Anabaenopeptin A [2b]	843.41643	844.42319	-0.31	8.071	DDA for preferred ion	35.00
Anabaenopeptin B [4]	836.4532	837.46017	-1.53	6.555	No MS2	NA
Caldorin [4]	305.23517	306.24245	-1.01	11.731	No MS2	NA
cis-Carboxydihydroanatoxin-a [3a]	211.12085	212.12813	0.02	3.787	DDA for preferred ion	42.86
Dehydroradiosumin [4]	430.22132	431.22859	-0.71	0.982	No MS2	NA
Deoxygadusol [3b]	188.06846	187.06118	-0.08	3.389	DDA for preferred ion	0.00
Radiosumin B [3a]	444.23684	445.24412	0.31	1.088	DDA for preferred ion	72.00
Gloeolactone [3b]	292.2035	275.20012	0.03	10.759	DDA for preferred ion	0.00
Hapalindole A-formamide [3b]	356.16555	355.15828	-0.95	13.421	DDA for preferred ion	0.00
Hermitamide A [4]	359.28209	360.28936	-0.1	7.151	No MS2	NA
Koshikalide [3a]	294.18308	293.17606	-0.8	11.385	DDA for preferred ion	86.62
Malyngic acid [3a]	328.22491	327.21764	-0.14	8.314	DDA for preferred ion	100.00
MC-RR [2b]	1037.56429	519.79011	-0.95	7.513	DDA for other ion	60.00

Mycosporine-GABA [4]	273.12098	274.12825	-0.64	5.257	No MS2	NA
Nakienone A [4]	194.09417	193.08689	-0.36	10.027	No MS2	NA
Palmyrrolinone [3a]	237.10002	238.10729	-0.81	3.443	DDA for preferred ion	39.47
Palythine [3b]	244.10572	245.11299	-0.98	0.905	DDA for preferred ion	0.00
Puna'auic acid [3b]	312.22975	311.22287	-0.96	11.161	DDA for preferred ion	NA (MS2 Not Abundant Enough)
Sacrolide A [3b]	308.19845	307.19148	-1.06	9.858	DDA for preferred ion	0.00
Santacruzamate A [3b]	278.1629	279.17015	-0.18	3.821	DDA for preferred ion	15.38
Tricholactone [3a]	254.1516	253.14454	-0.83	9.891	DDA for preferred ion	93.33

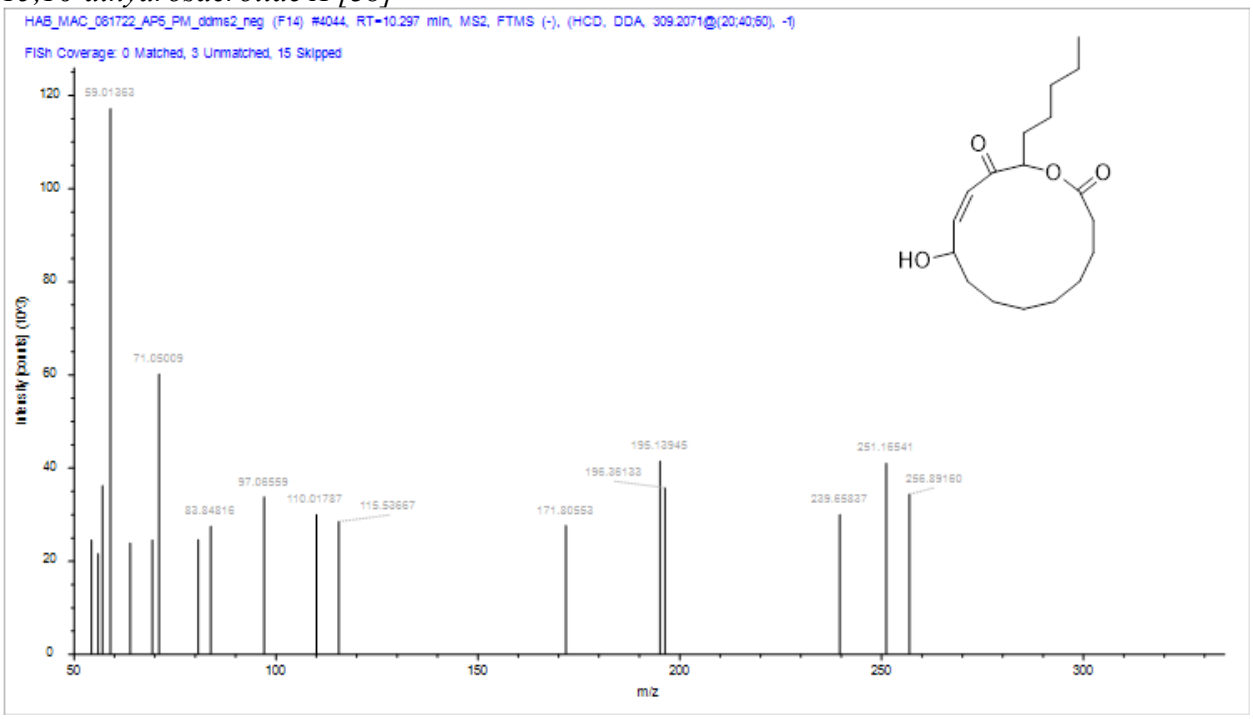
†NOTES: MW="molecular weight" (Da); m/z=mass to charge ratio; Δppm=mass error (parts per million); RT="retention time"; MS2 DDA= "Data Dependent Acquisition" for fragment ions; FISh="Fragment Ion Search" score

Table S14: Lake MacBride Tentatively Identified Compounds with annotated MS2 Spectra and structure for applicable features and identification confidence level in []

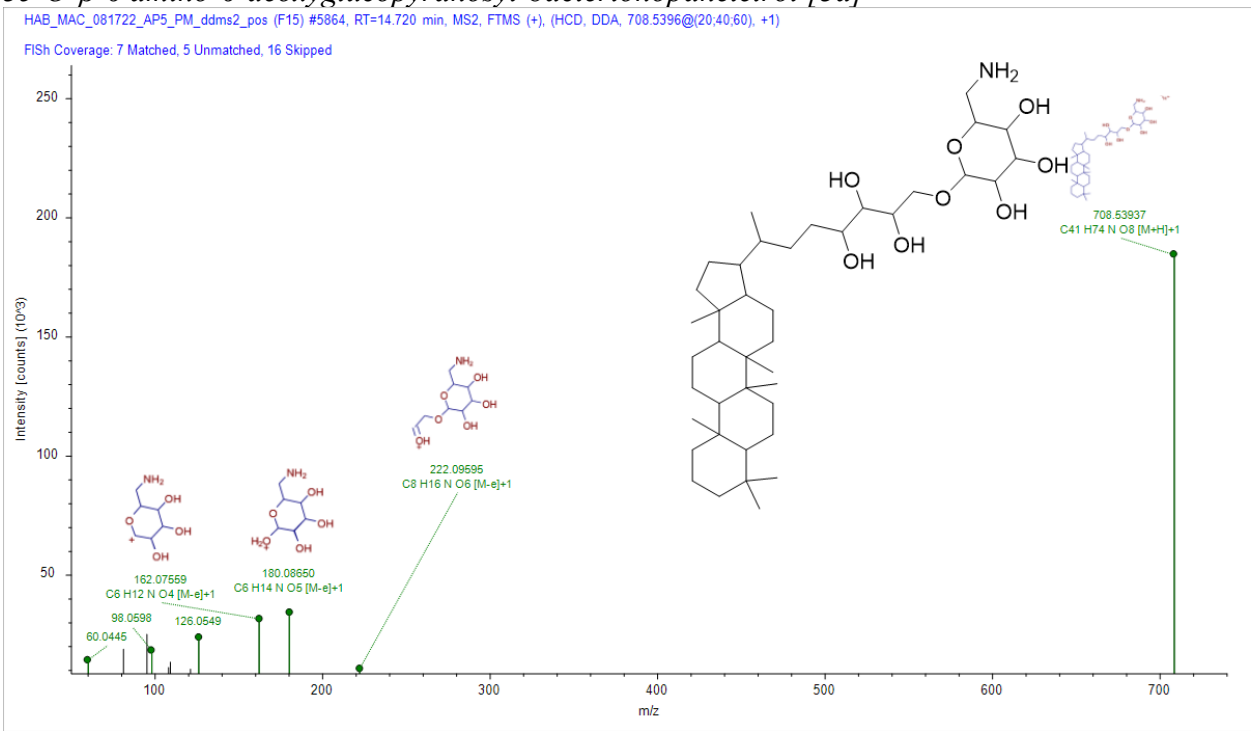
12-hydroxy-2-oxo-11-epi-hinesol [3b]



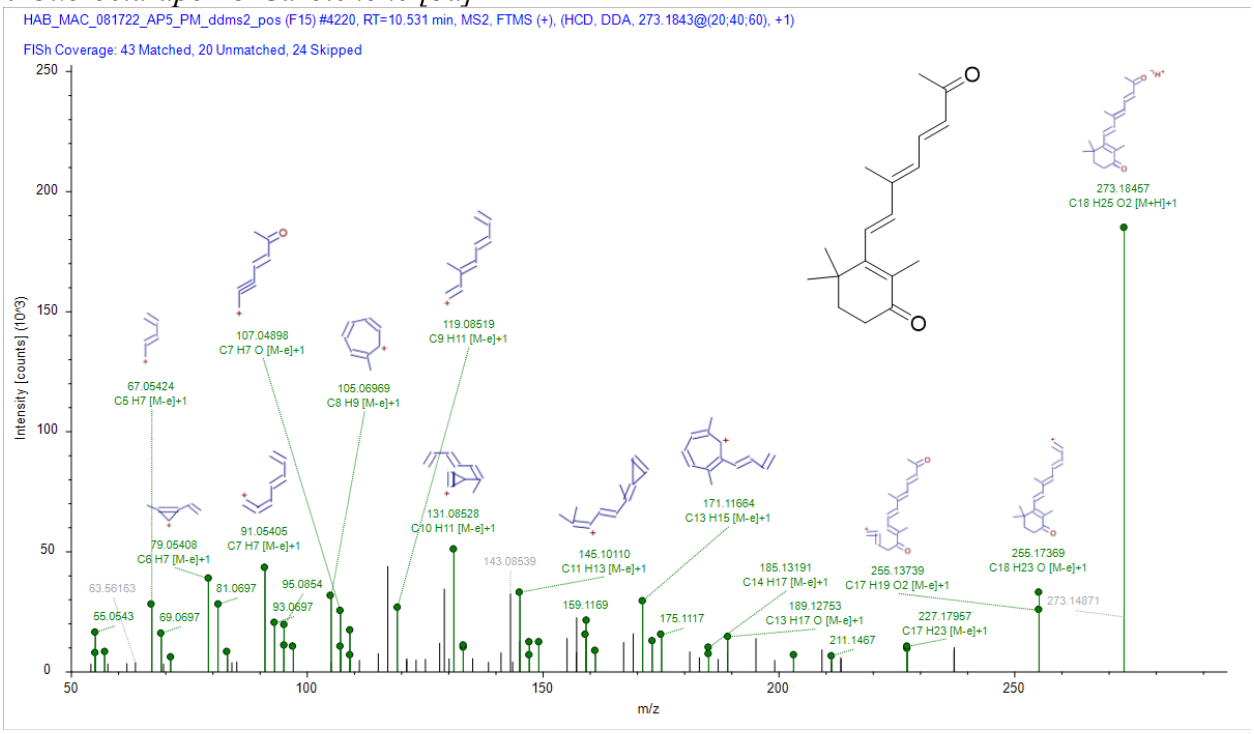
15,16-dihydrosacrolide A [3b]



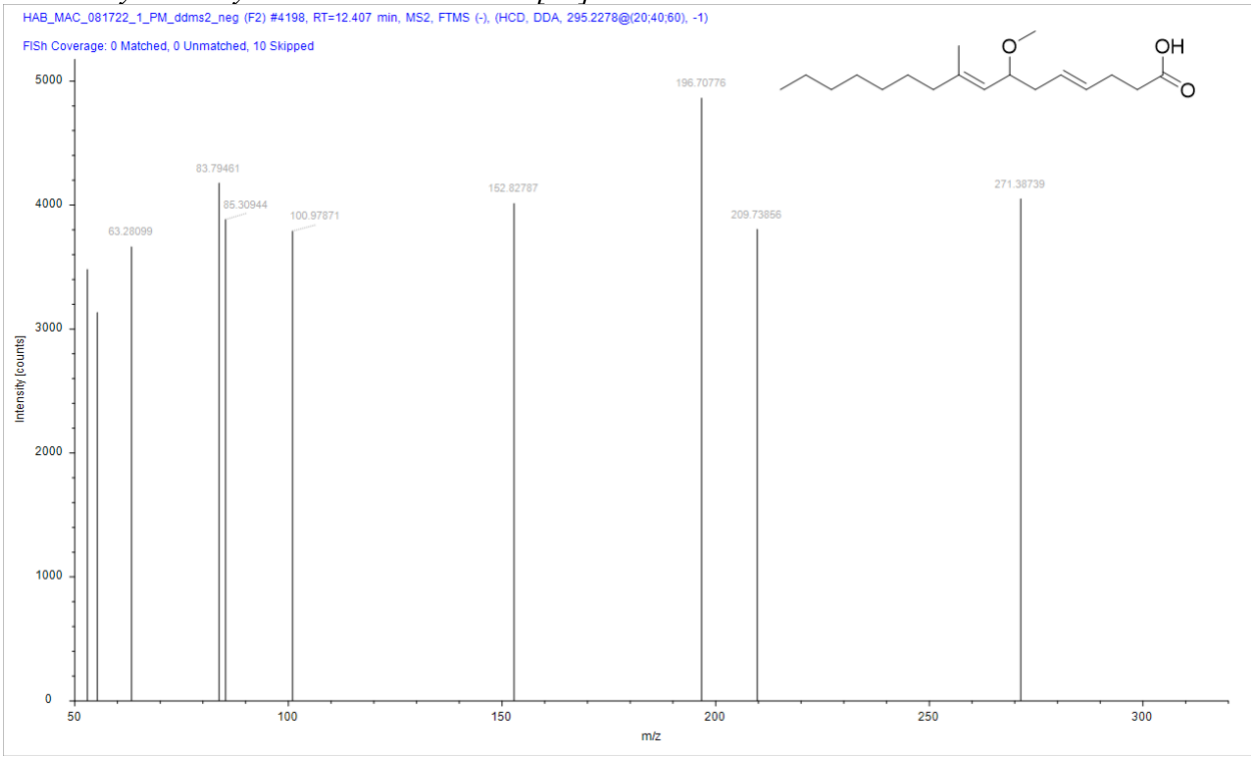
35-O-β-6 amino-6-deoxyglucopyranosyl-bacteriohopanetetrol [3a]



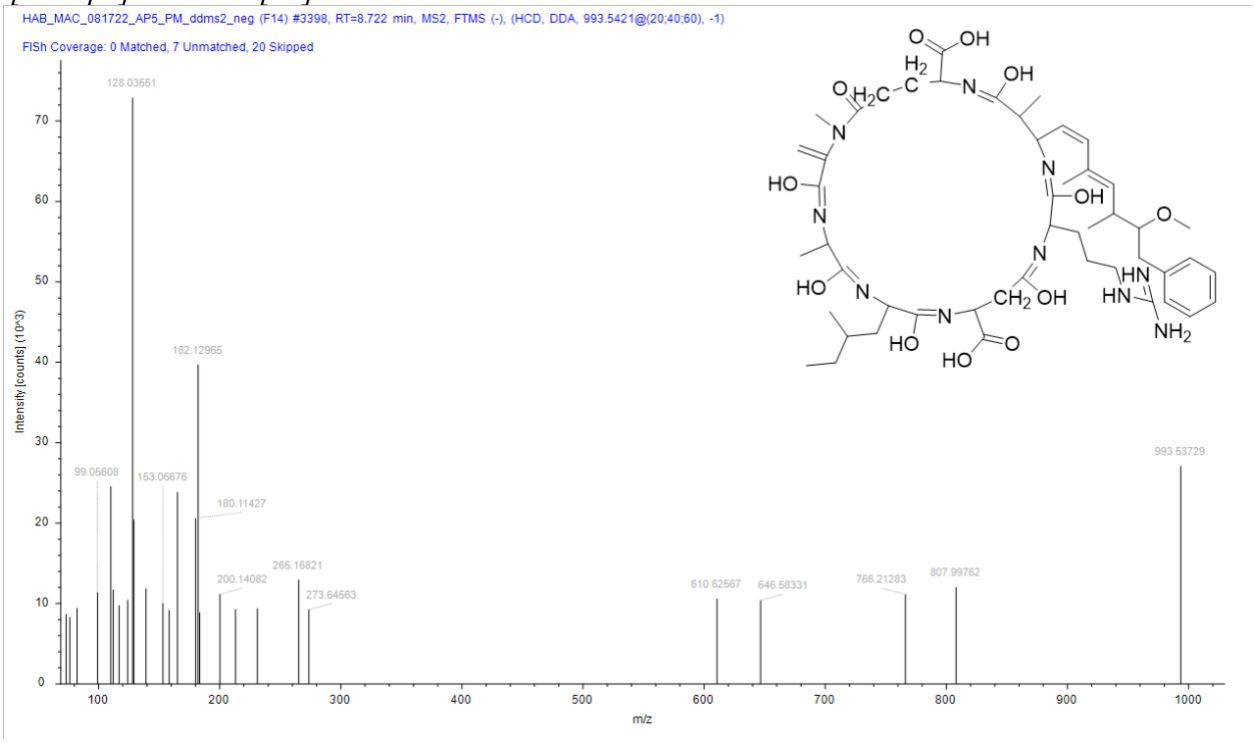
4-Oxo-beta-apo-13-Carotenone [3a]



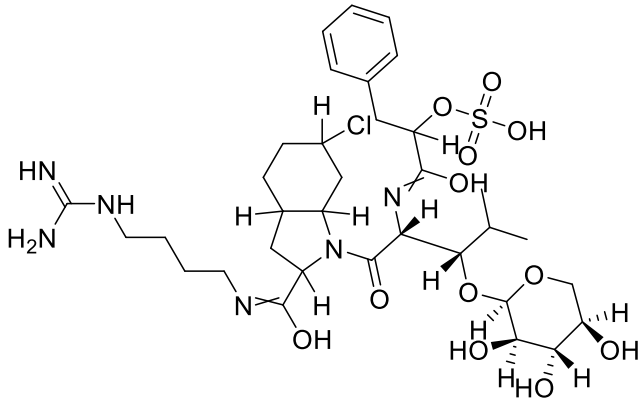
7-methoxy-9-methylhexadecadienoic acid[3b]

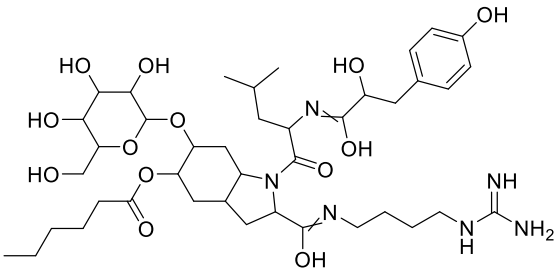


[D-Asp3]MC-HilR [3b]



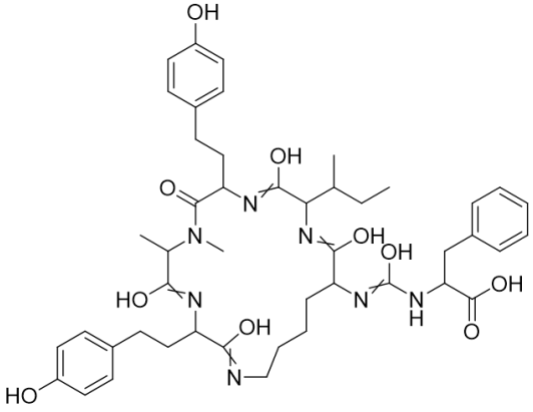
Aeruginosin 205A (No MS2) [4]



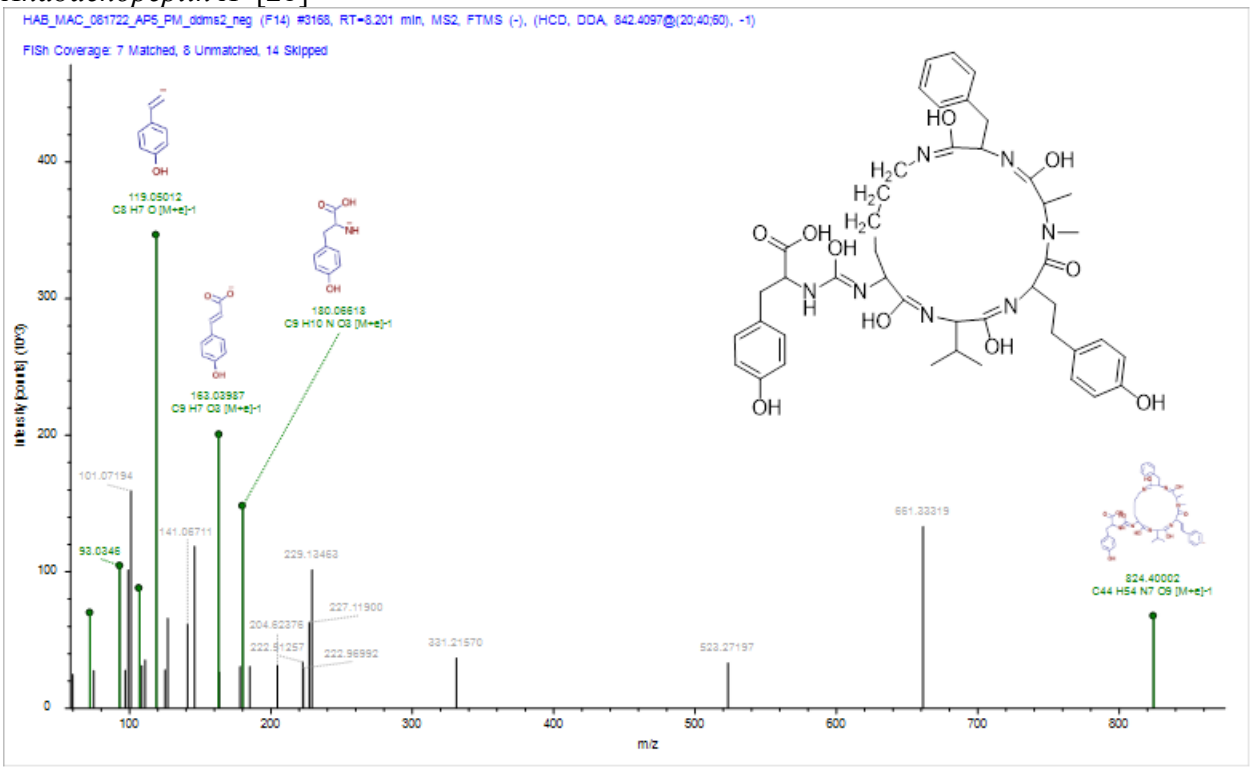


Aeruginosin 850 (No MS2) [4]

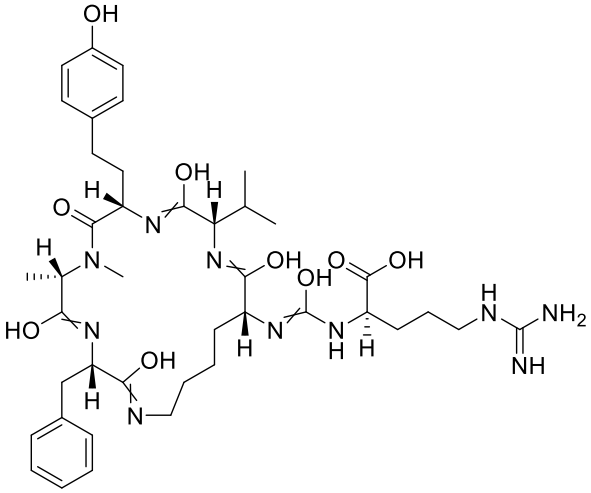
Anabaenopeptin 871 (No MS2) [4]



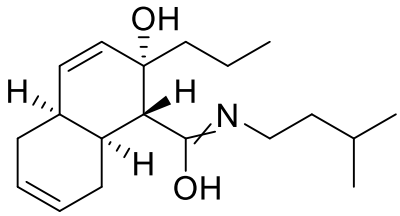
Anabaenopeptin A [2b]



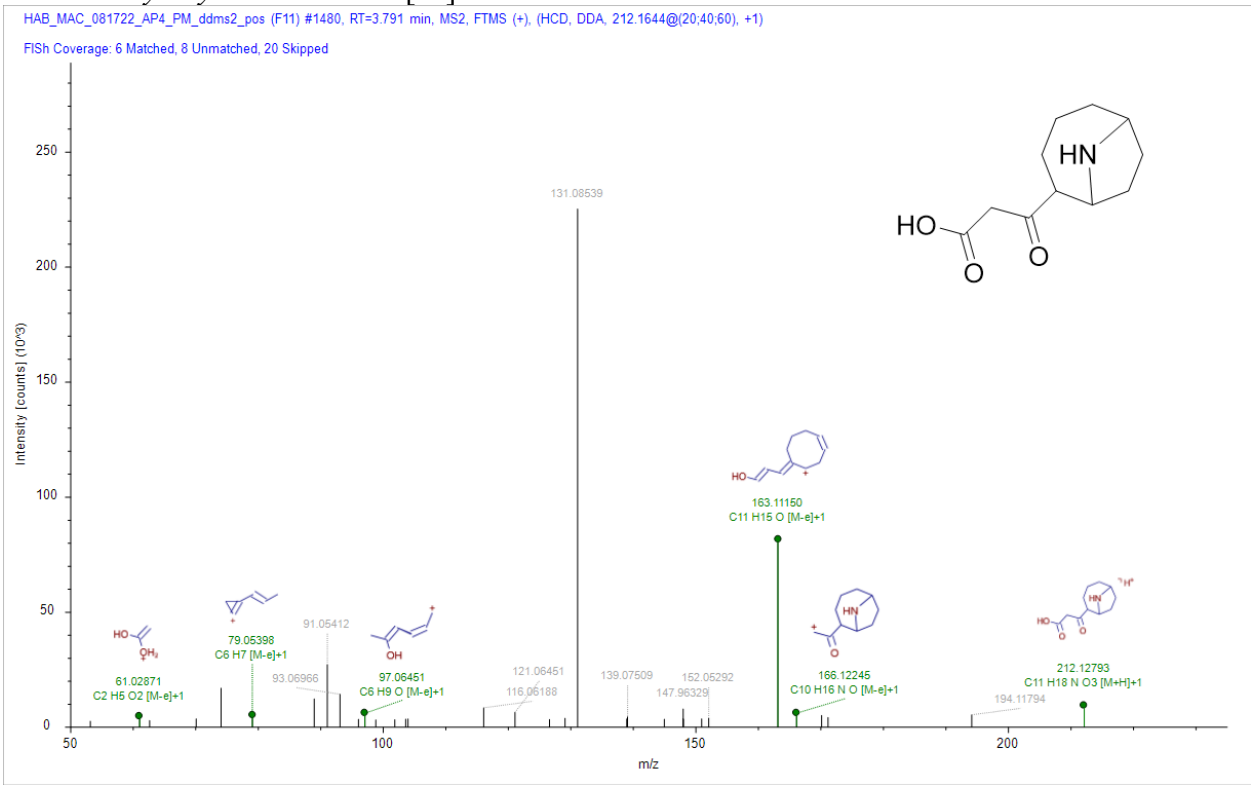
Anabaenopeptin B (No MS2) [4]



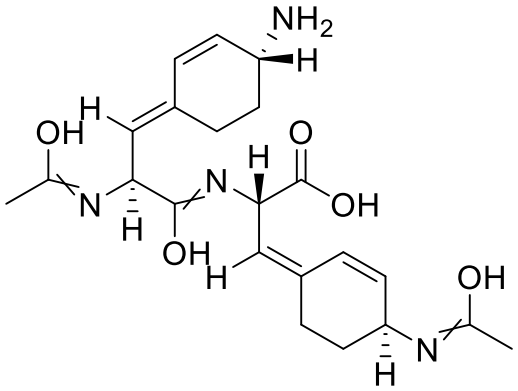
Caldorin (No MS2) [4]



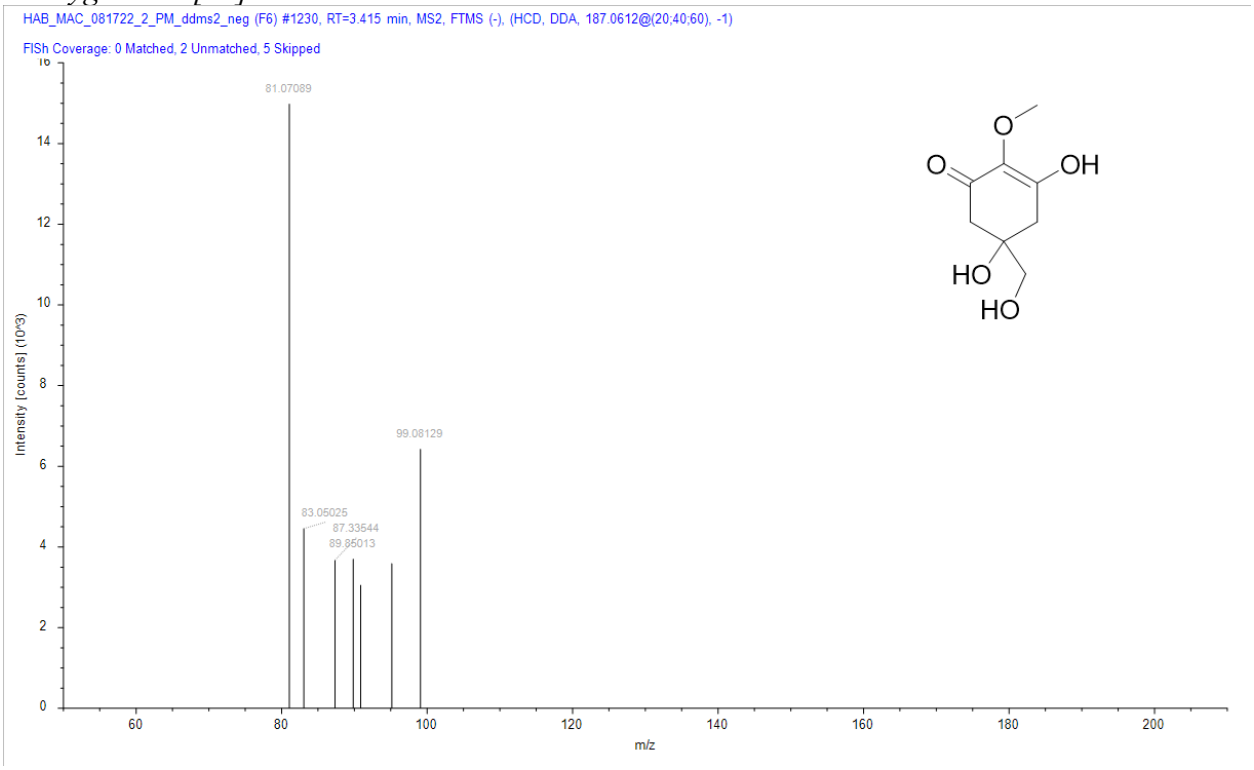
cis-Carboxydihydroanatoxin-a [3a]



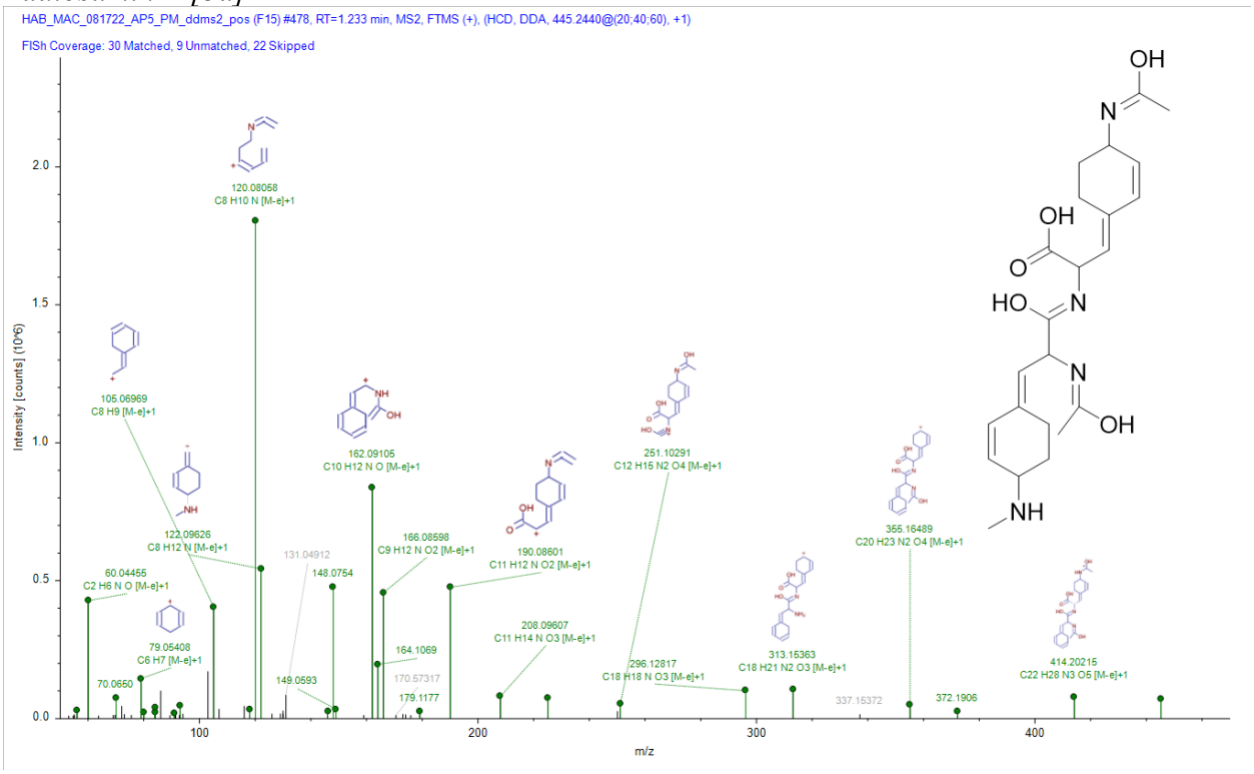
Dehydroradiosumin (No MS2) [4]



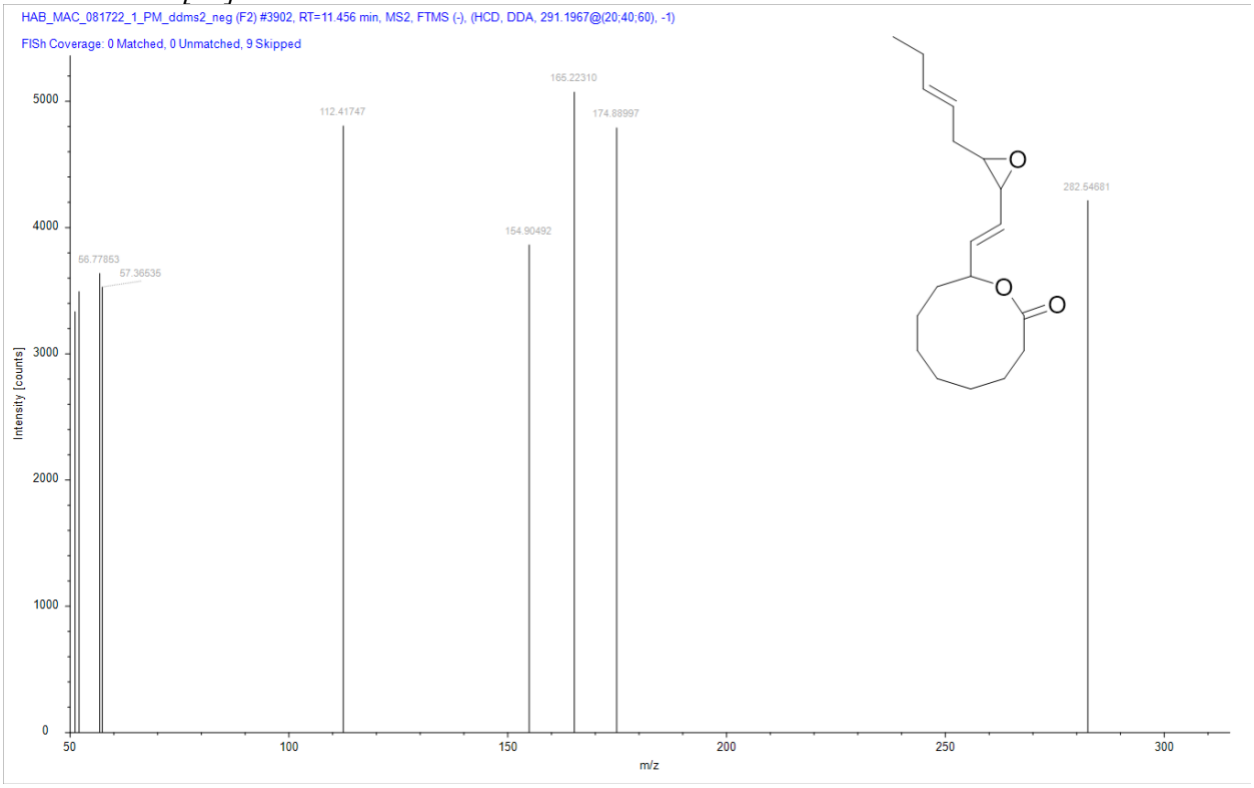
Deoxygadusol [3b]



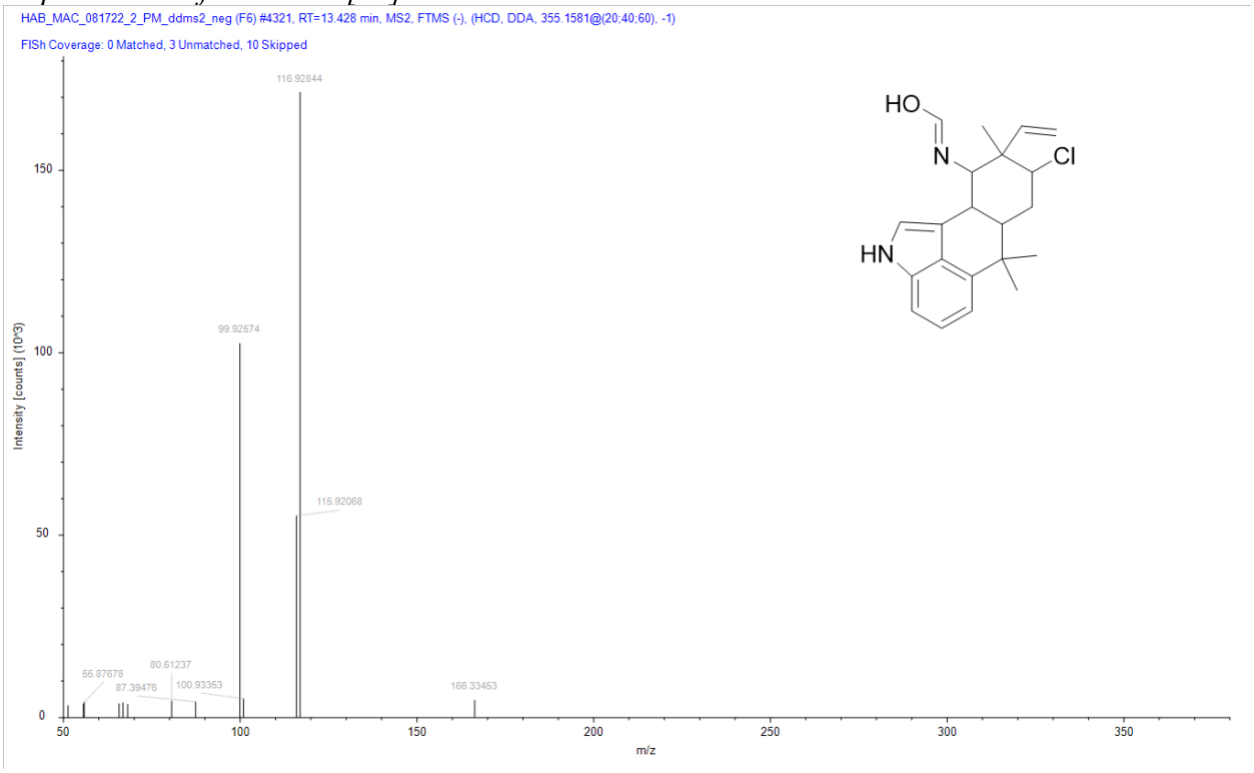
Radiosumin B [3a]



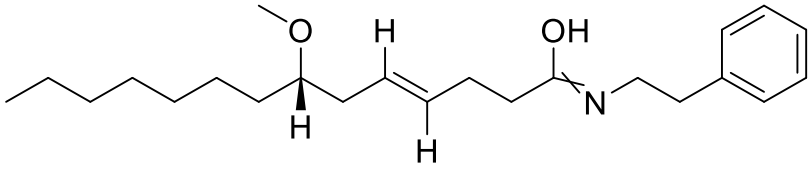
Gloeolactone [3b]



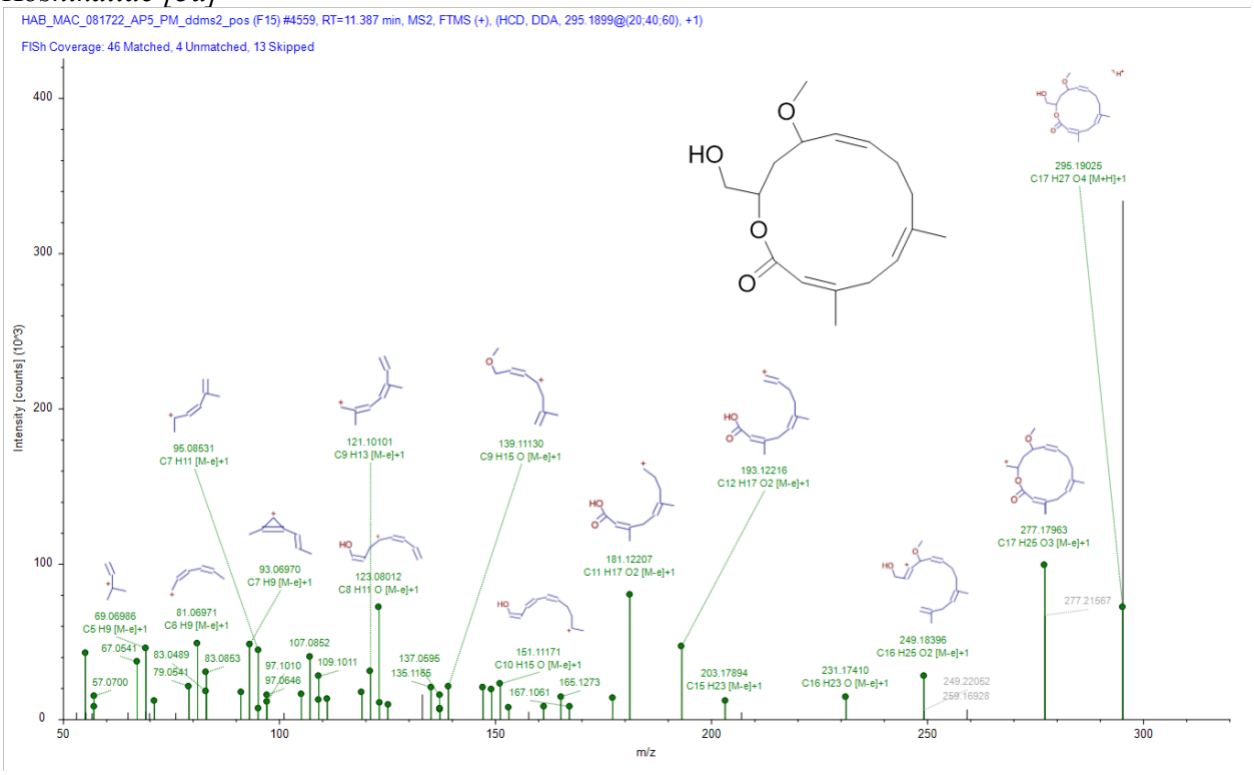
Hapalindole A-formamide [3b]



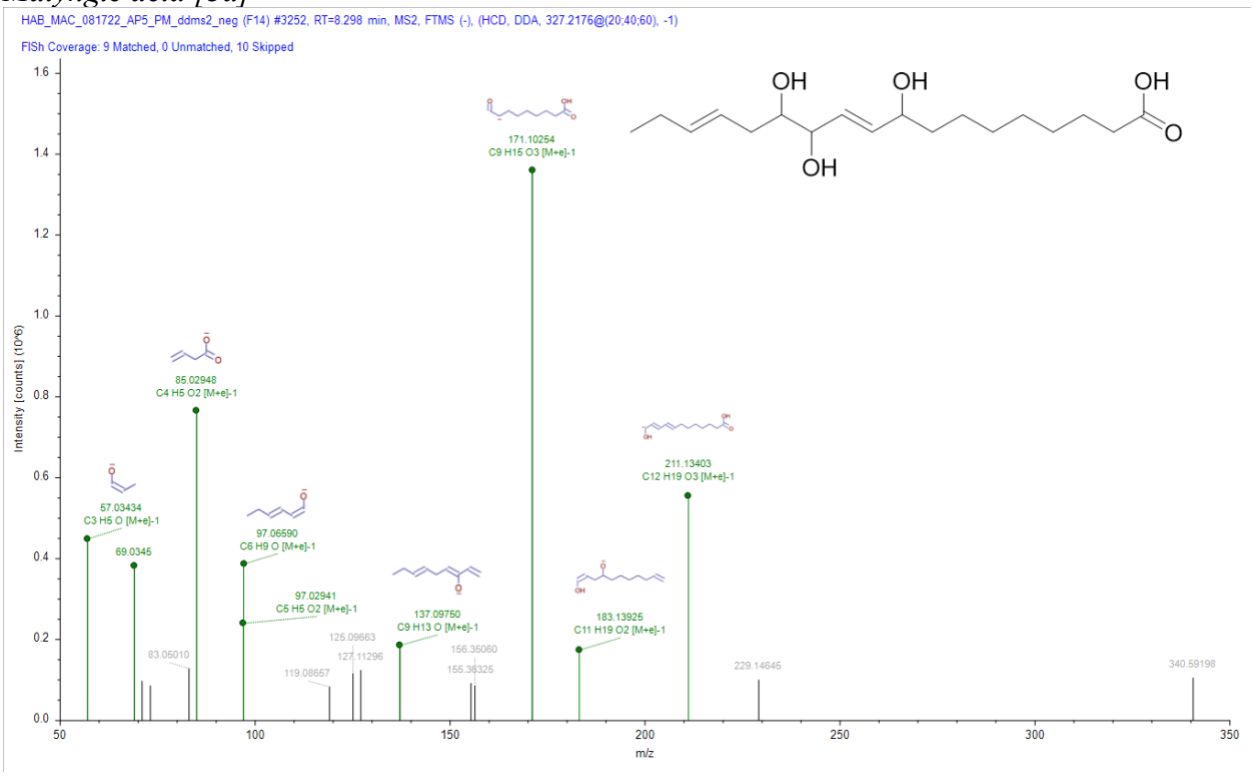
Hermitamide A (No MS2) [4]



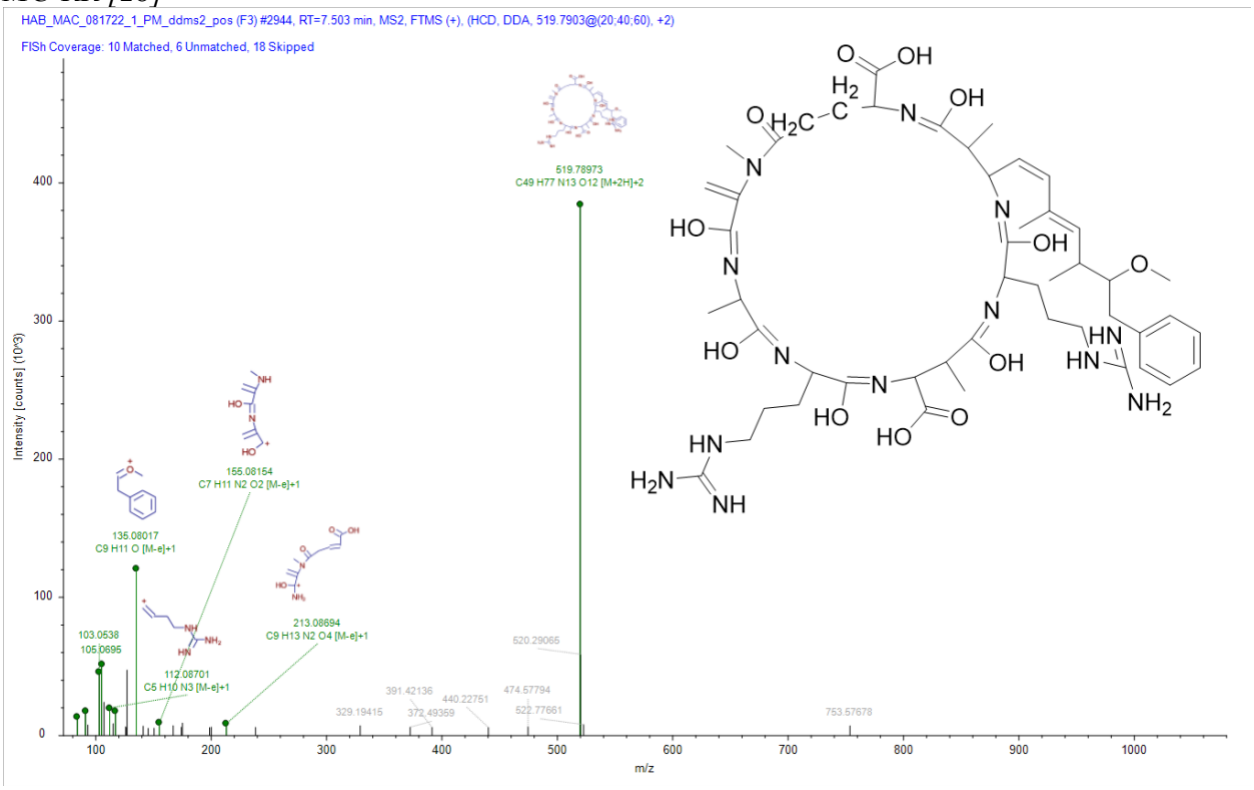
Koshikalide [3a]



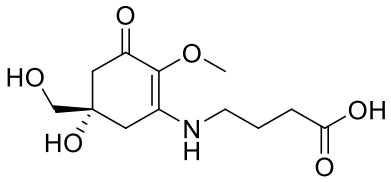
Malyngic acid [3a]



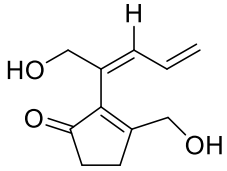
MC-RR [2b]



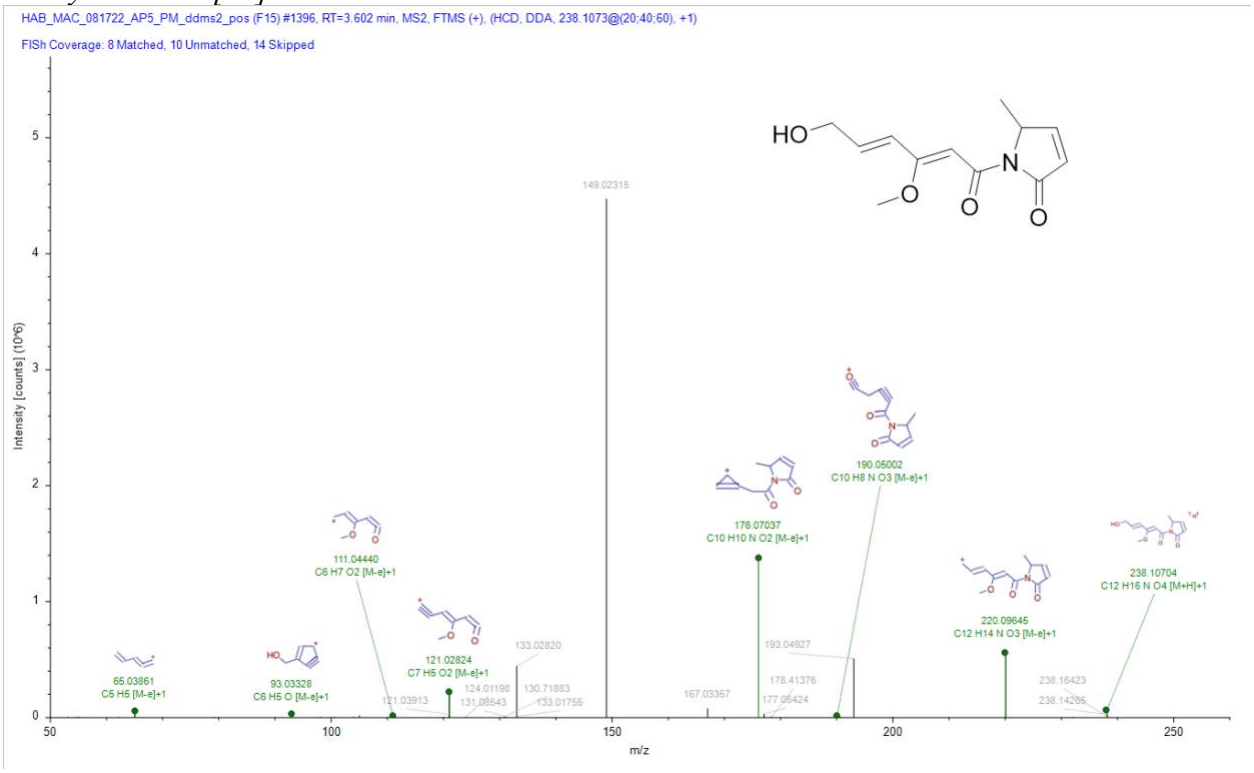
Mycosporine-GABA (No MS2) [4]



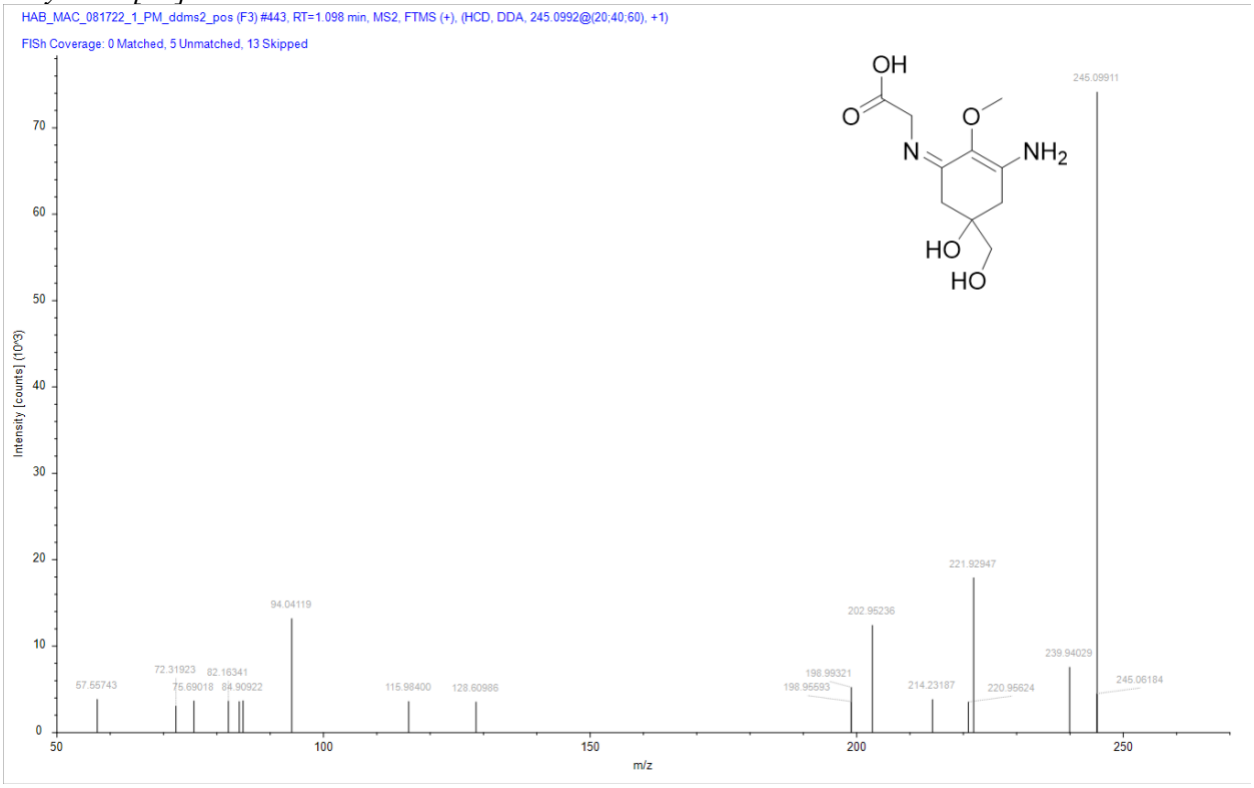
Nakienone A (No MS2) [4]



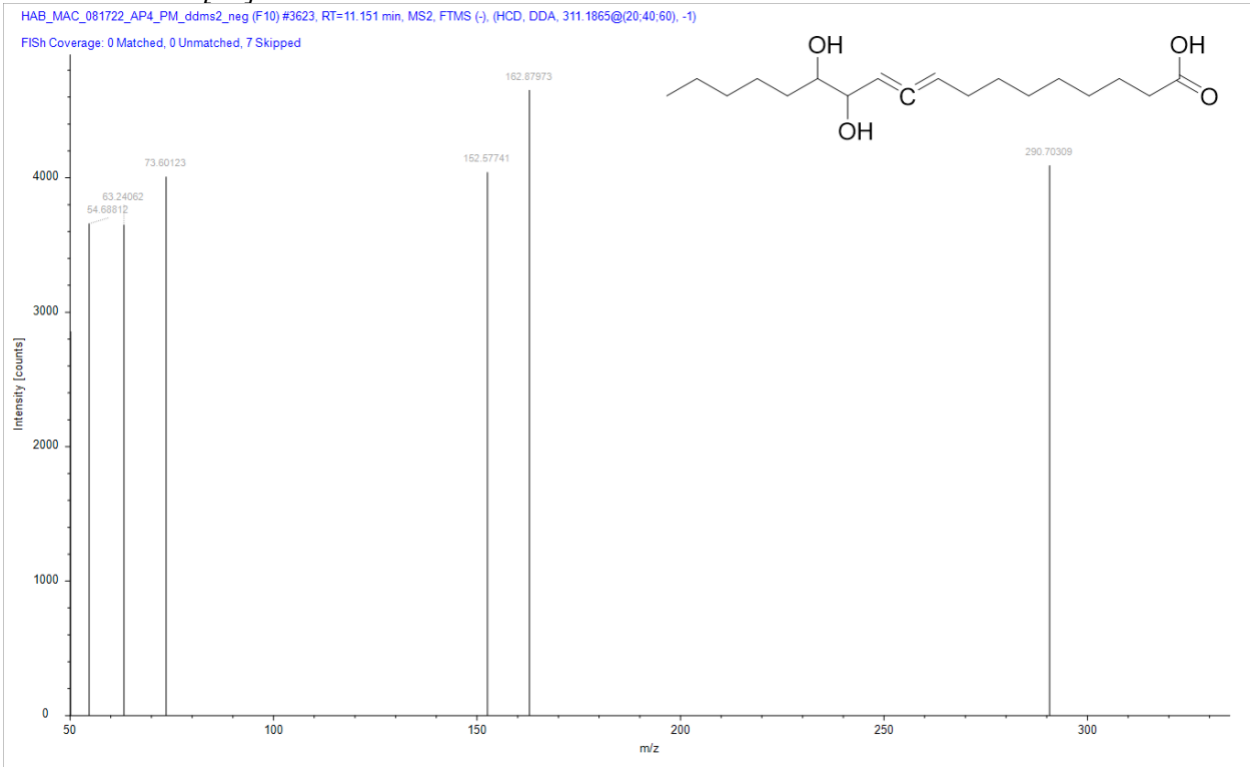
Palmyrrolinone [3a]



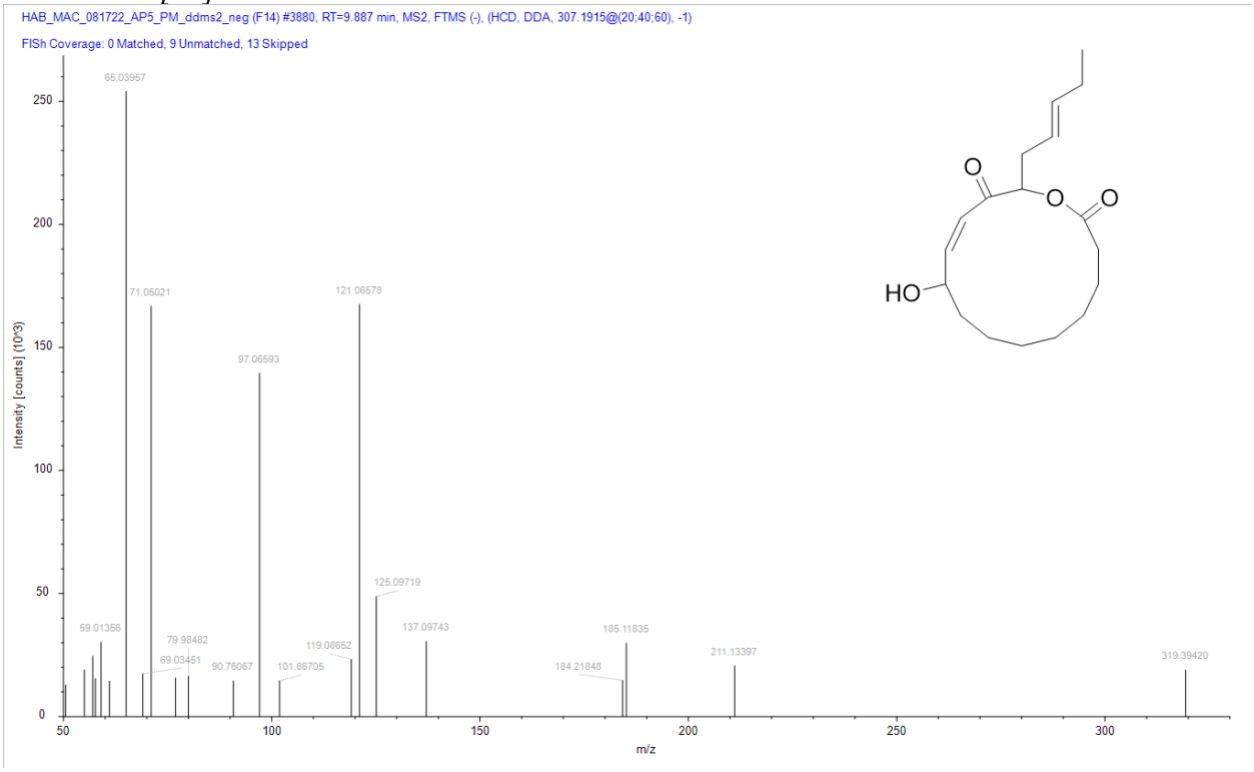
Palythine [3b]



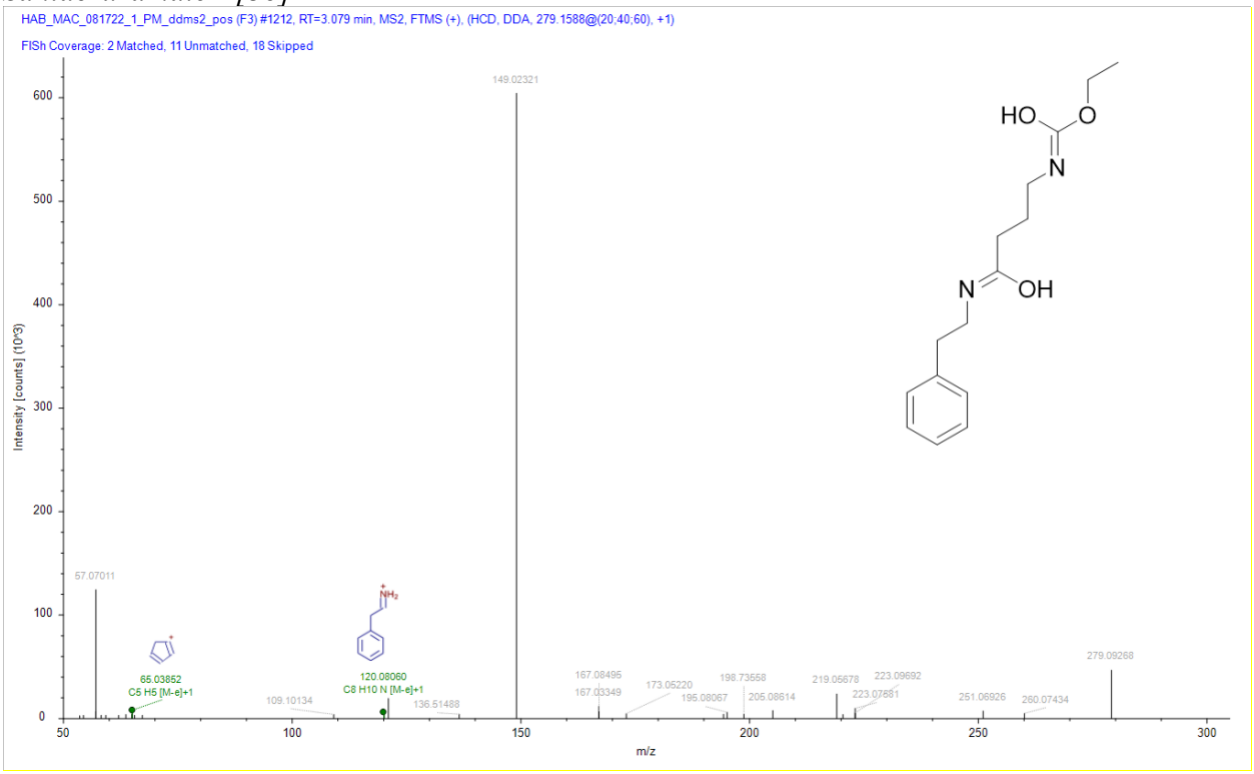
Puna'auic acid [3b]



Sacrolide A [3b]



Santacruzamate A[3b]



Tricholactone [3a]

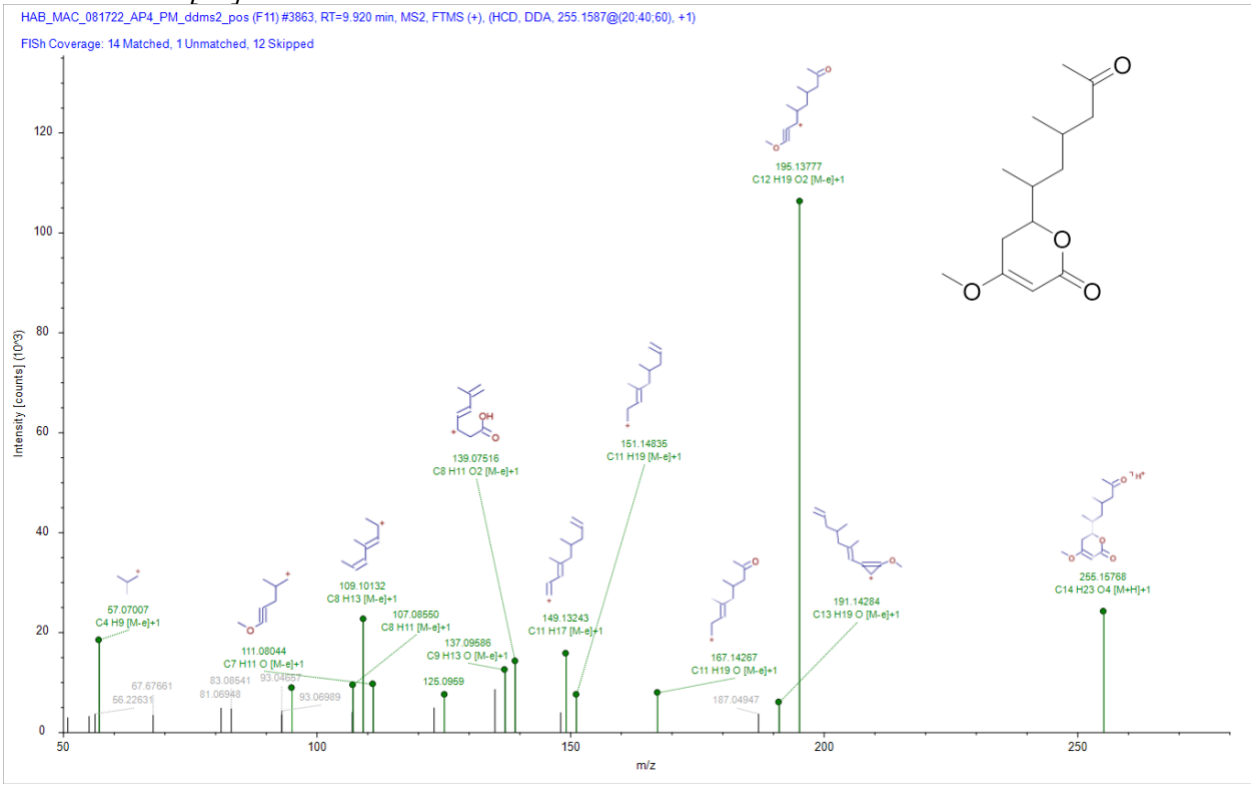


Table S15: Cosine theta values for the filtered (cyanometDB matched) results from Lake MacBride (based on normalized responses for each feature). The green highlight indicates cos(θ) > 0.85.

Site	North 1	North 2	South 2	South 1
North 1	1.000	0.562	0.534	0.452
North 2		1.000	0.984	0.375
South 2			1.000	0.378
South 1				1.000

Table S16: Tentatively Identified Cyanometabolites from Lake Darling and their associated information as reported in the CyanMetDB.

Identified Features	Cyanomet_ ID	Compound Class	Alternative Class name	Genus initial reported in	Species initial reported in	Original ID Location
[D-Asp3]MC-HilR	CyanoMetDB_1821	Microcystin		*Planktothrix; Dolichospermum/ Anabaena*	*[Planktothrix]sp.; [Dolichospermum]sp.*	Germany (freshwater)
11-Carboxyanatoxin-a	CyanoMetDB_1742	Anatoxin-a		*Aphanizomenon*	*issatschenkoi*	New Zealand (Freshwater)
15,16-dihydrosacrolide A	CyanoMetDB_1632	other cyclic non-peptide		*Aphanothece*	*sacrum*	Japan (freshwater)
478-Da MAA	CyanoMetDB_1603	other linear non-peptide	Mycosporine-like amino acids	Nostoc	commune	Japan (Freshwater)
6-Acetamidotridecane	CyanoMetDB_1685	other linear non-peptide	Zacetamide	*Microcoleus*	*lyngbyaceus*	Chuuk Island Atoll (marine)
7-methoxy-9-methylhexadecadienoic acid	CyanoMetDB_1629	other linear non-peptide		*Lyngbya*	*aestuarii*	Hawaii (marine)
Gloeolactone	CyanoMetDB_1638	other linear non-peptide		*Gloeotrichia*		Montana, USA (freshwater)
Koshikalide	CyanoMetDB_1654	other linear non-peptide		*Lyngbya*		Japan (marine)
Lyngbyacarbonate	CyanoMetDB_0286	other linear non-peptide		*Lyngbya/Moorea*	*majuscula/producens*	Oregon, USA (marine)
Malyngic acid	CyanoMetDB_1627	other linear non-peptide		*Lyngbya/Moorea*	*majuscula/producens*	Hawaii (marine)
MC-RR	CyanoMetDB_1802	Microcystin		*Microcystis*	*viridis*	Japan (freshwater)
Microcystbiopterin A	CyanoMetDB_1641	other linear non-peptide		*Microcystis*		Isreal, (freshwater)
N-butyryl-L-homoserine lactone	CyanoMetDB_0294	other linear non-peptide	AHLs	*Lyngbya/Moorea*	*majuscula/producens*	Hawaii (marine)
Palmyrrolinone	CyanoMetDB_2408	other cyclic peptide		*cf. Oscillatoria and Hormoscilla*		Palmyra Atoll (USA, Marine)
Palythene	CyanoMetDB_1711	other linear non-peptide	Mycosporine-like amino acids	*Aphanothece*	*halophytica*	India (freshwater)
Pukeleimide D	CyanoMetDB_1713	other linear non-peptide	Pukeleimide	*Lyngbya/Moorea*	*majuscula/producens*	Hawaii (marine)
Puna'auic acid	CyanoMetDB	other linear		*Pseudanabaena*		Pacific ocean (marine)

	_1628	non-peptide				
Sacrolide A	CyanoMetDB _1635	other cyclic non-peptide		*Aphanothece*	*sacrum*	Japan (freshwater)
Tricholactone	CyanoMetDB _0040	other cyclic peptide	other cyclic depsipeptide			Texas, USA (freshwater

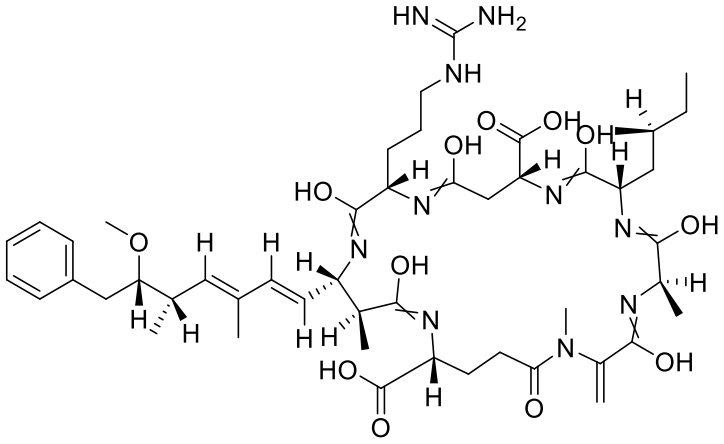
Table S17: Tentatively Identified Cyanometabolites from Lake Darling with critical suspect screening results†

Identified Features	Calc. MW	m/z	RT [min]	Δppm	MS2	FISh Score (when Applicable)
[D-Asp3]MC-HilR [4]	994.54672	995.55385	8.678	-2.06	No MS2	NA
11-Carboxyanatoxin-a [4]	209.10498	210.11225	0.794	-1.01	No MS2	NA
15,16-dihydrosacrolide A [3b]	310.21411	309.20703	10.757	-0.96	DDA for preferred ion	NA (MS2 not abundant enough)
478-Da MAA [3b]	478.17935	477.17246	0.883	-1.09	DDA for preferred ion	100.00
6-Acetamidotridecane [4]	241.24006	242.24734	10.03	-2.09	No MS2	NA
7-methoxy-9-methylhexadecadienoic acid [3b]	296.23486	295.22758	12.349	-0.96	DDA for preferred ion	NA (MS2 not abundant enough)
Gloeolactone [3b]	292.20378	291.1965	11.4	-0.23	DDA for preferred ion	NA (MS2 not abundant enough)
Koshikalide [3b]	294.18291	293.17563	10.375	-0.68	DDA for preferred ion	0.00
Lyngbyacarbonate [4]	232.05818	231.05088	0.793	-0.54	No MS2	NA
Malyngic acid [3b]	328.22468	327.21751	8.291	-0.91	DDA for preferred ion	NA (MS2 not abundant enough)
MC-RR [2b]	1037.56429	519.78942	7.498	-1.47	DDA for other ion	63.64
Microcystbiopterin A [4]	441.18527	442.19239	1.396	-1.56	No MS2	NA
N-butyryl-L-homoserine lactone [4]	171.08936	204.12284	0.626	-1.09	No MS2	NA
Palmyrrolinone [3a]	237.09966	238.1069	3.443	-1.9	DDA for preferred ion	30.00
Palythene [3b]	284.13668	268.11731	0.792	-1.89	DDA for other ion	50.00
Pukeleimide D [4]	296.10036	297.10764	0.875	-1.59	No MS2	NA
Puna'auic acid [4]	312.22975	311.22269	11.131	-1.32	DDA for preferred ion	NA (MS2 not abundant enough)
Sacrolide A [3b]	308.19845	307.19138	9.853	-0.34	DDA for preferred ion	0.00
Tricholactone [3a]	254.1516	253.14442	9.877	-1.58	DDA for preferred ion	81.82

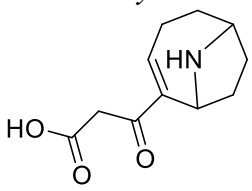
†NOTES: MW="molecular weight" (Da); m/z=mass to charge ratio; Δppm=mass error (parts per million); RT="retention time"; MS2 DDA= "Data Dependent Acquisition" for fragment ions; FISh="Fragment Ion Search" score

Table S18: Lake Darling Tentatively Identified Compounds with annotated MS2 Spectra and structure for applicable features and identification confidence level in []

[D-Asp3]MC-HilR (No MS2) [4]

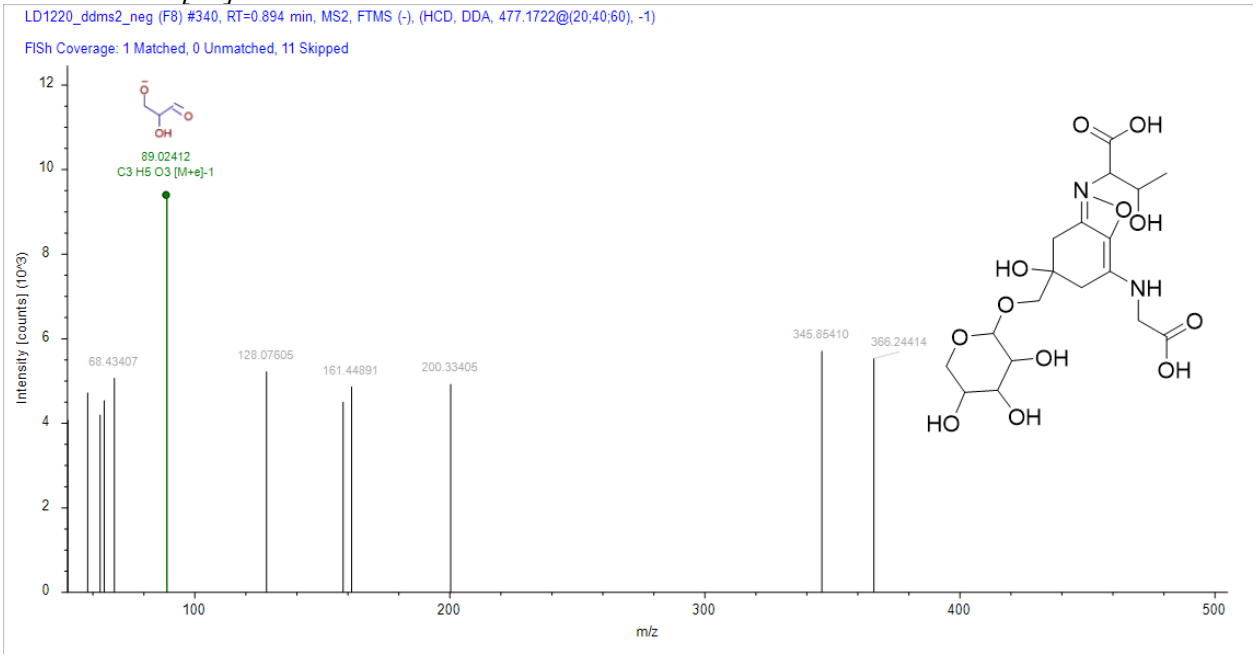


11-Carboxyanatoxin-a (No MS2) [4]

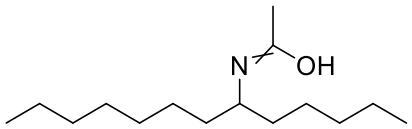


15,16-dihydrosacrolide A [3b]

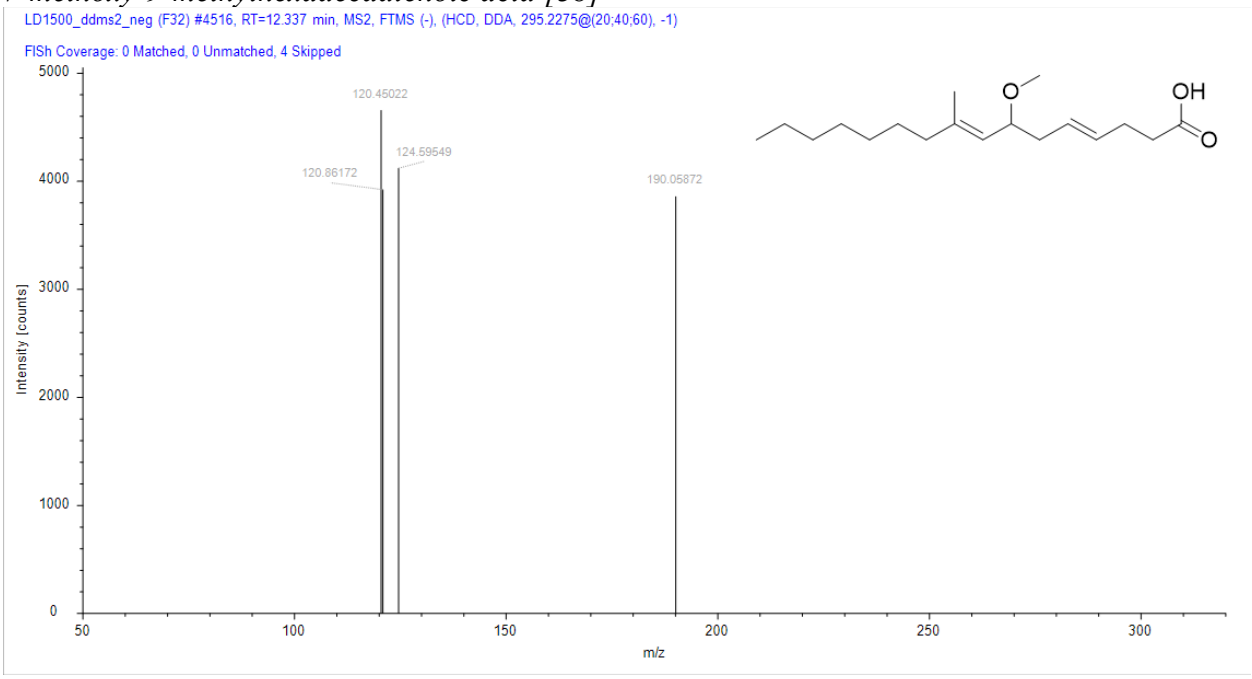
478-Da MAA [3b]



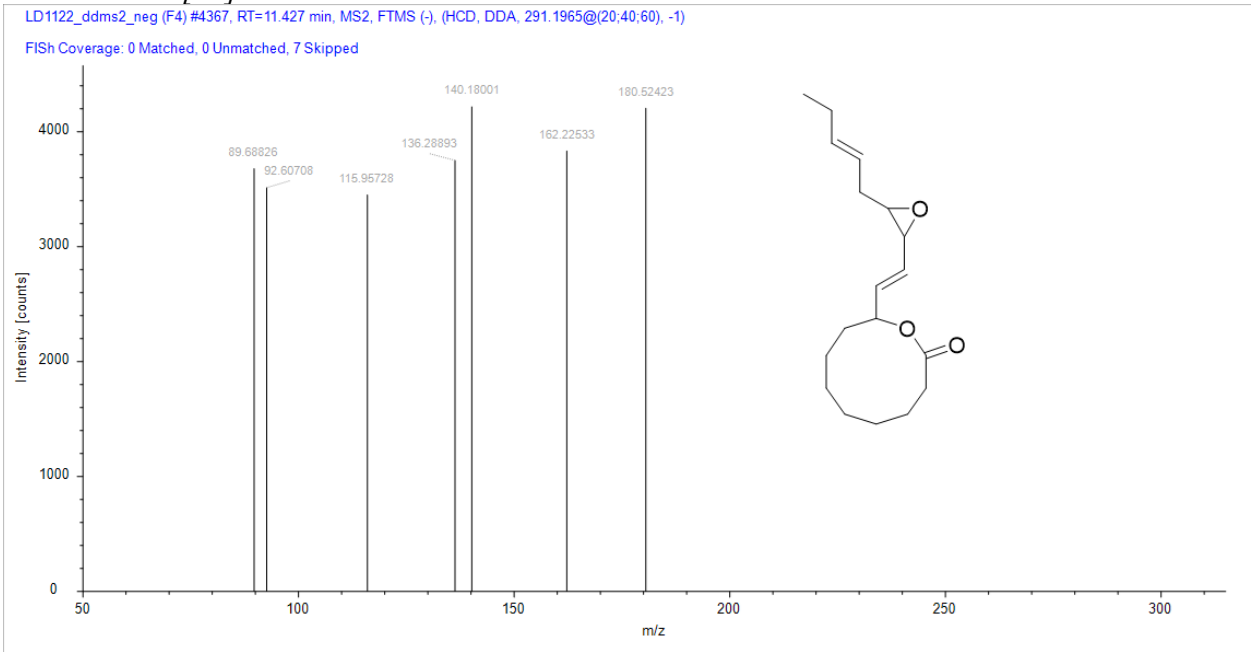
6-Acetamidotridecane, no MS2 [4]



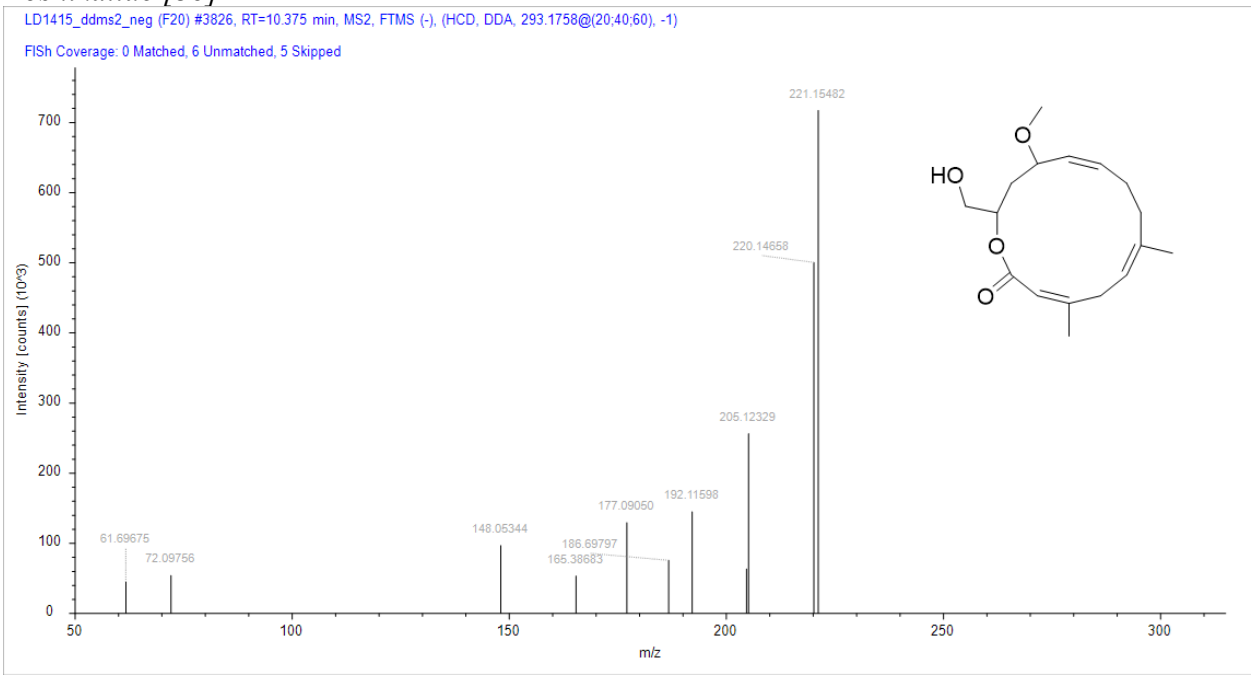
7-methoxy-9-methylhexadecadienoic acid [3b]



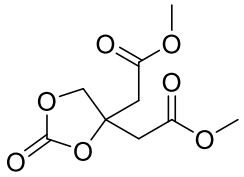
Gloeolactone [3b]



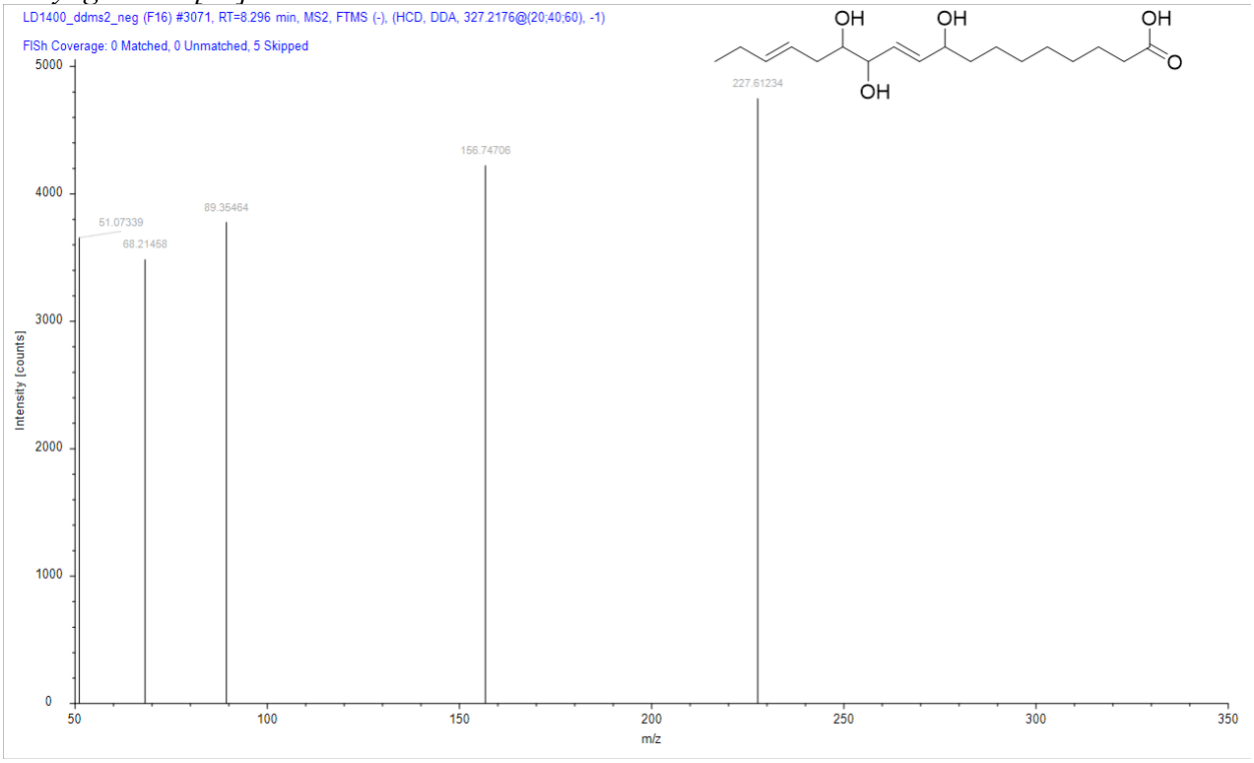
Koshikalide [3b]



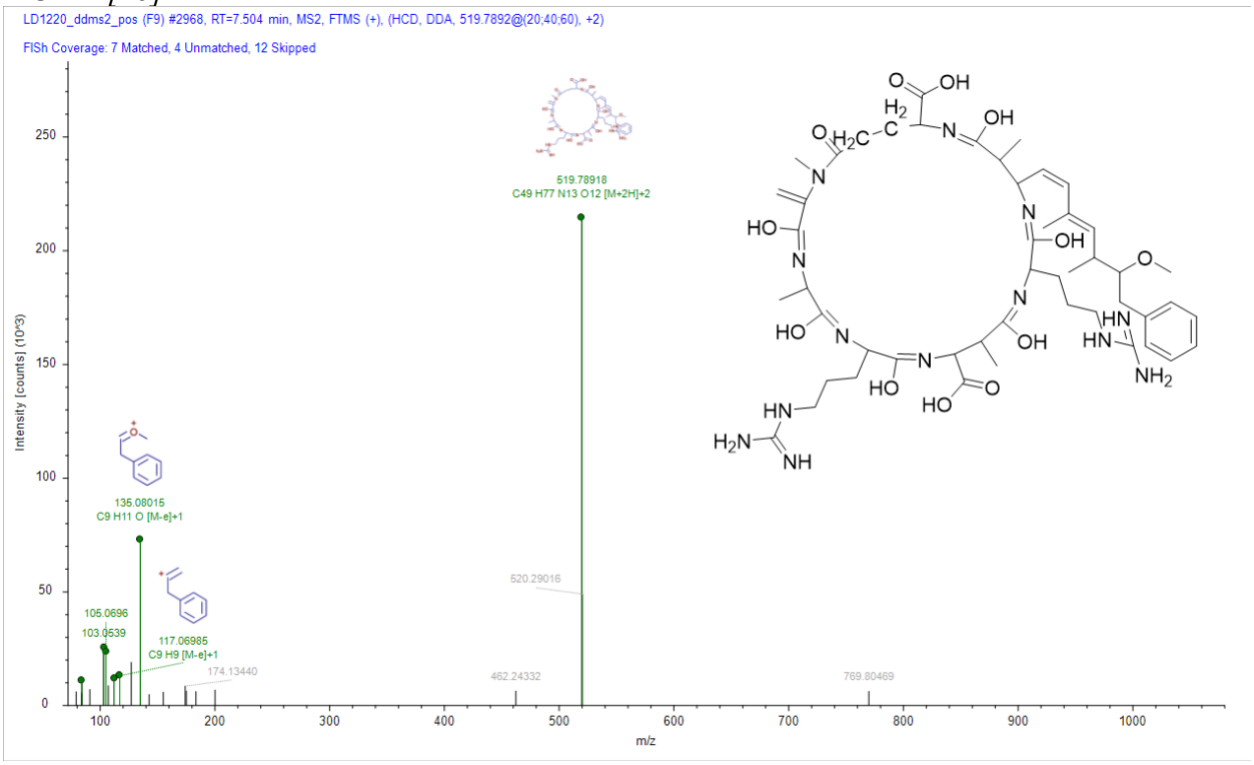
Lyngbyacarbonate, No MS2 [4]



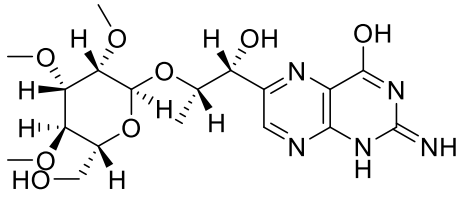
Malyngic acid [3b]



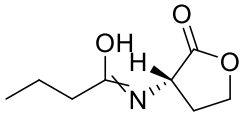
MC-RR [2b]



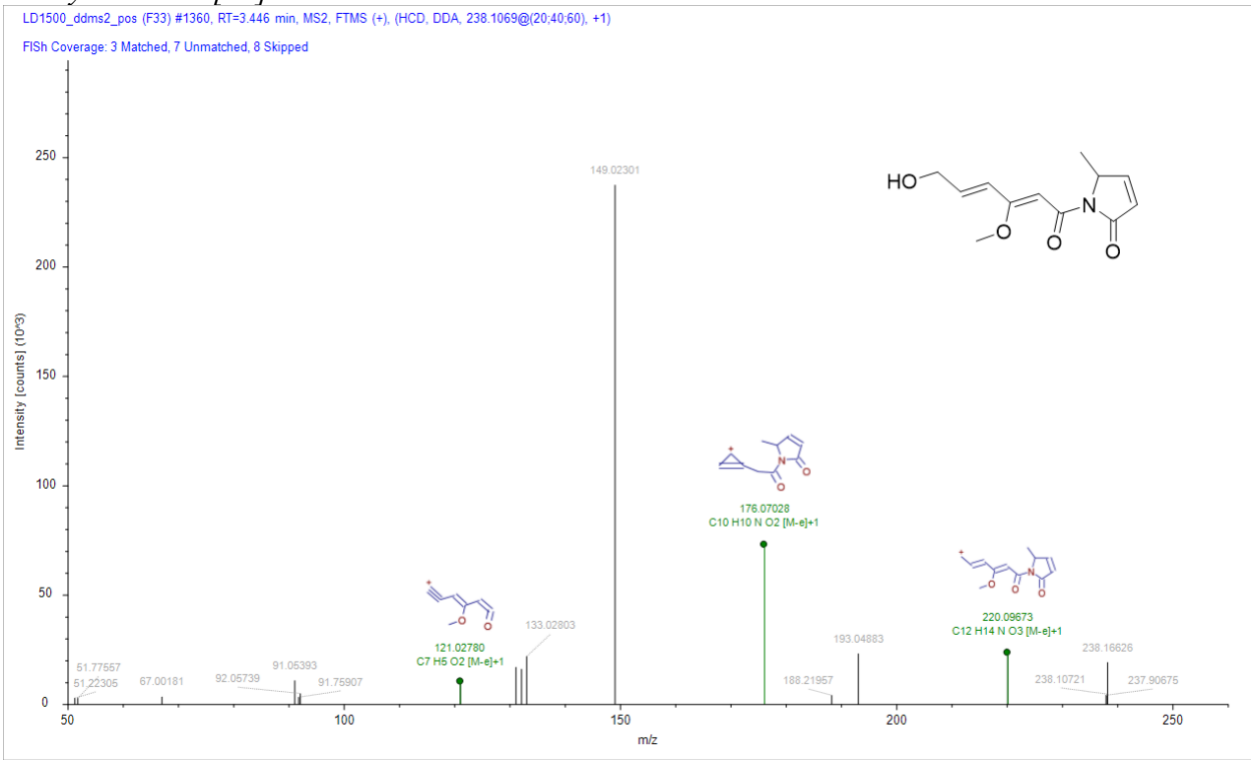
Microcystbiopterin A (No MS2) [4]



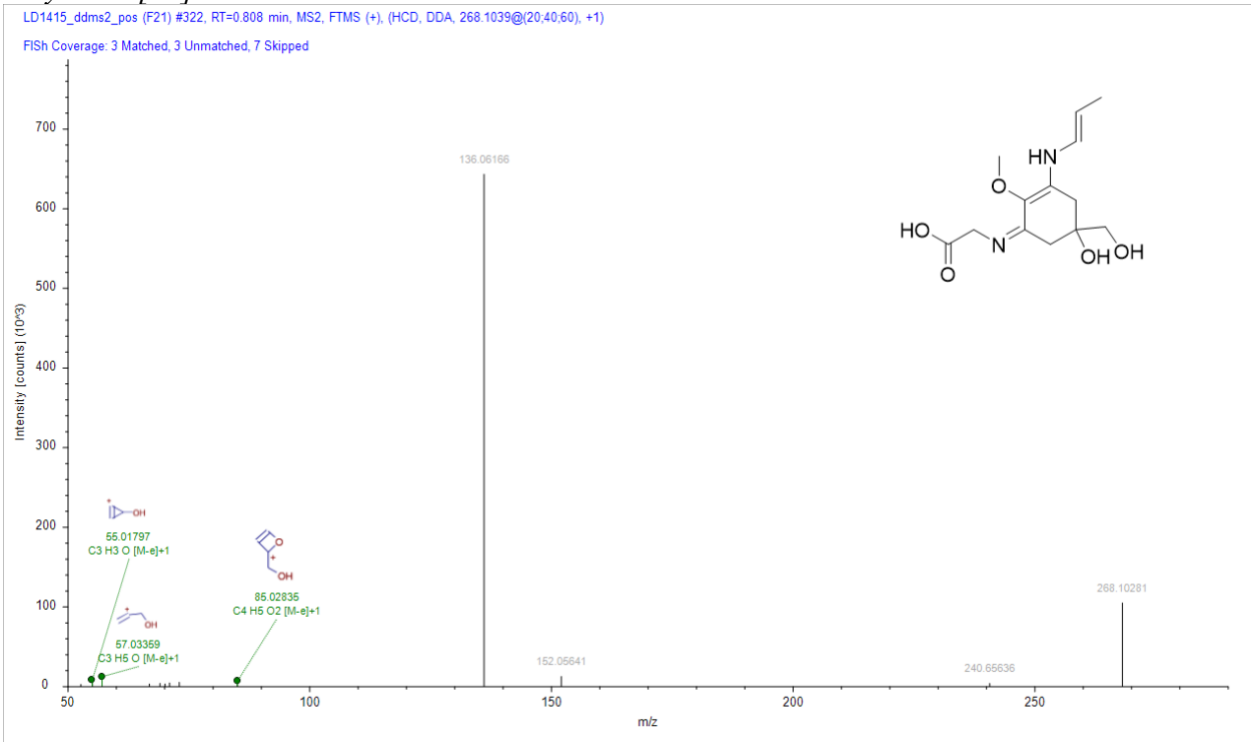
N-butyryl-L-homoserine lactone (No MS2) [4]



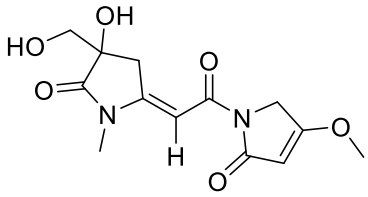
Palmyrrolinone [3a]



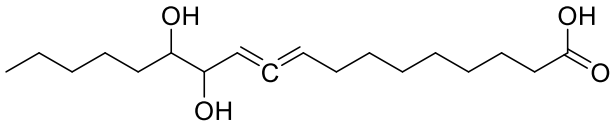
Palythene [3b]



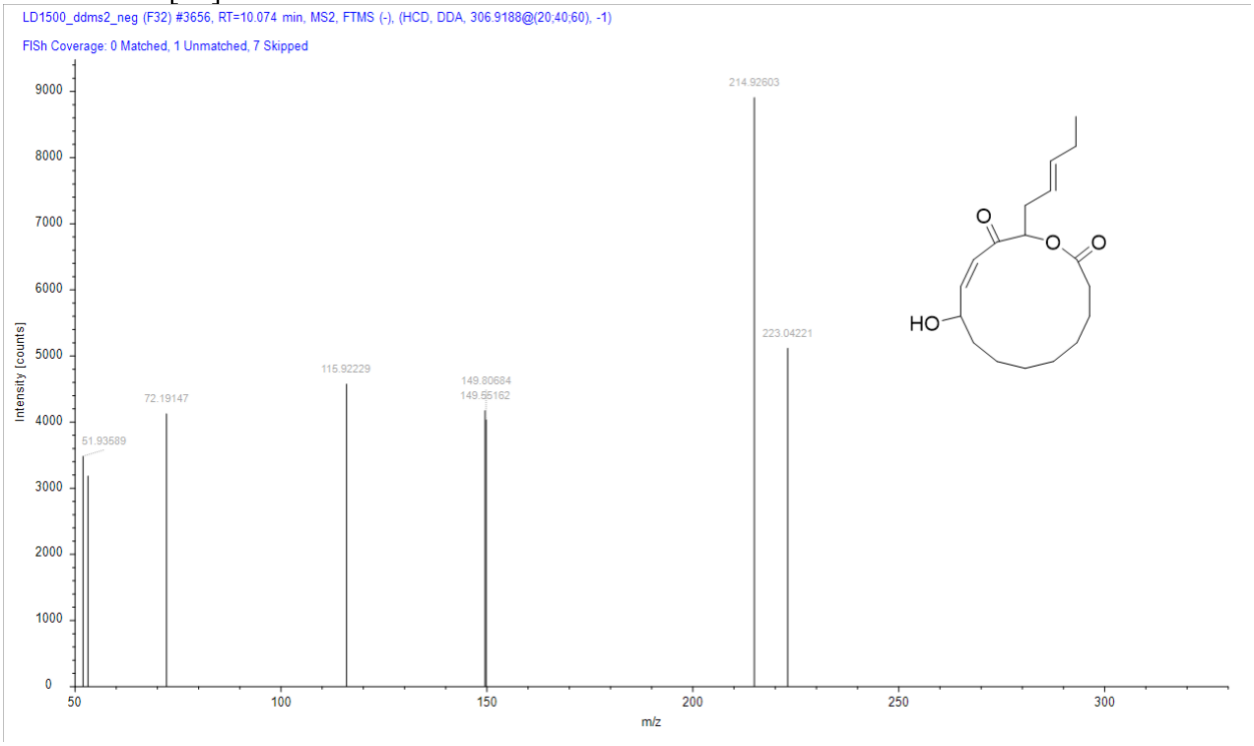
Pukeleimide D (No MS2) [4]



Puna'auic acid (No MS2) [4]



Sacrolide A [3b]



Tricholactone [3a]

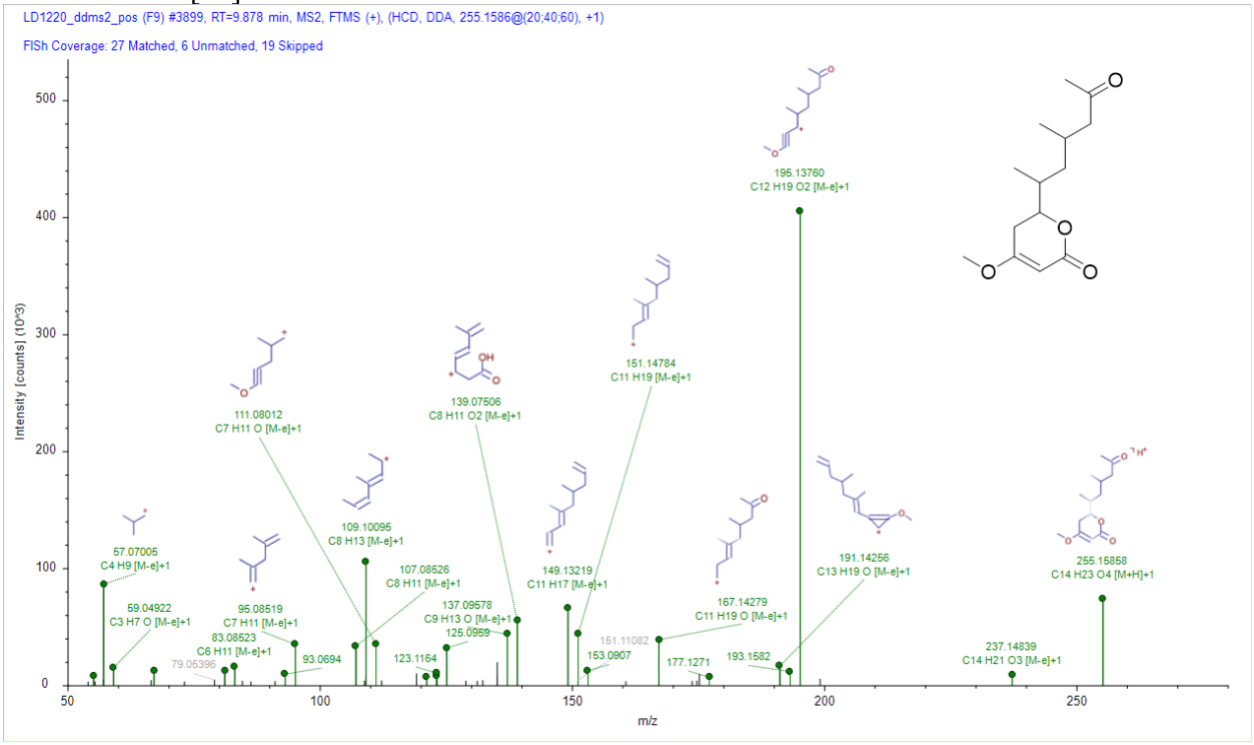


Table S19: Cosine theta values for the filtered (cyanometDB matched) results from Lake Darling (based on normalized responses for each feature). The green highlight indicates cos(θ) > 0.85.

Times	11:22	12:20	14:00	14:15	14:30	14:45	15:00
11:22	1.000	0.974	0.733	0.681	0.761	0.722	0.483
12:20		1.000	0.769	0.706	0.826	0.731	0.515
14:00			1.000	0.962	0.885	0.943	0.558
14:15				1.000	0.890	0.963	0.691
14:30					1.000	0.901	0.754
14:45						1.000	0.658
15:00							1.000

References

1. Zaffiro, A., Rosenblum, L. & Wendelken, S. C. Method 546: Determination of total microcystins and nodularins in drinking water and ambient water by Adda enzyme-linked immunosorbent assay. *United States Environ. Prot. Agency* 1–21 (2016).
2. Peter, K. T. *et al.* Nontargeted Analysis Study Reporting Tool: A Framework to Improve Research Transparency and Reproducibility. *Anal. Chem.* **93**, 13870–13879 (2021).
3. de Jonge, N. F. *et al.* MS2Query: reliable and scalable MS2 mass spectra-based analogue search. *Nat. Commun.* **14**, 1752 (2023).
4. 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).