Identification of toxic cyclopeptides based on mass spectral library matching

Dr. S.K. Mishra, Principal, S. Sinha College, Aurangabad, India.

Abstract : To gain perspective on the use of tandem mass spectral libraries for identification of toxic cyclic peptides, the new library was built from 263 mass spectra (mainly MS^2 spectra) of 59 compounds of that group, such as microcystins, amatoxins, and some related compounds. Mass spectra were extracted from the lit-erature or specially acquired on ESI-Q-ToF and MALDI-ToF/ToF tandem instruments. ESI-MS^2 product-ion mass spectra appeared to be rather close to MALDI-ToF/ToF fragment spectra which are uncommon for mass spectral libraries. Testing of the library was based on searches where reference spectra were in turn cross-compared. The percentage of 1st rank correct identifications (true positives) was 70% in a general case and 88–91% without including knowingly defective ('one-dimension') spectra as test ones. The percentage of 88–91% is the principal estimate for the overall performance of this library that can be used in a method of choice for identification of individual cyclopeptides and also for group recognition of individual classes of such peptides. The approach to identification of cyclopeptides based on mass spectral library matching proved to be the most effective for abundant toxins. That was confirmed by analysis of extracts from two cyanobacterial strains.

1. Introduction

Cyclic peptides are the groups of bioactive natural compounds of both non-ribosomal and ribosomal origin. The former have been detected in many bacterial and fungal strains and other natural sources for last decades. There were well-known cyanobacterial cyclopeptides such as microcystins, nodularins, anabaenopeptins, and related classes of toxins [1]. Other classes of non-ribosomal cyclopeptides are abundant amatoxins and phallotoxins that have been found in mushrooms [2]. In last years, many natural cyclopeptides that are synthesized via the ribosomal pathway also have been discovered [3].

Mass spectrometry has been playing an outstanding role in structure elucidation of new cyclic peptides and identification of known ones beginning from early reports, e.g. article [4]. In determination of known toxic peptides, mass of principal ions, predominantly $[M+H]^+/[M+2H]^{2+}$ in electrospray ionization mass spectrometry (ESI MS) and also their characteristic fragments in electrospray ionization tandem mass spectrometry (ESI-MS^2), were commonly measured (e.g. see [5]). Also, approaches to identification of cyclic peptides taken from proteomics are in progress. Here, researchers came up against the difficulty that there had been unusual amino acids in sequences and unusual types of fragment ions generated in mass spectrometer. In all circumstances, proteomics methods and approaches such as (a) searches of tandem mass spec-tra against genome databases [3,6] and (b) *de novo* interpretation/ annotation of tandem mass spectra [7–9], have been adopted for identification of ribosomal and non-ribosomal cyclopeptides, respectively. They can be considered as identification methods of 'unknown unknowns'.

Some other methods of proteomics can be also transformed for the use in mass spectrometry of analytes under the consideration. The analogy with linear peptides can be further continued. It is generally admitted that linear peptide fragment mass fingerprint-ing (see [10, p.217]) is the main method of protein identification. Peptides are firstly identified via database search of their tandem mass spectra against amino acid sequence databases. As proteomics advanced, it was suggested to build tandem mass spectral libraries for frequently detected peptides, in other words, 'known unknown' analytes [10, p.198].

At the same time, different libraries of the such type have been developed as new effective data systems for identification to cover other classes and groups of (bio)chemical compounds within metabolomics, toxicology, environmental analysis, and so on, e.g. see [10, p.188; 11]. Therefore, it was appropriate to expand this identification method to cover cyclic peptides relevant to these science fields. To develop the approach to recognition of cyclic peptides based on mass spectral libraries, we recently generated the short version of the new one containing 75 tandem mass spectra of 28 microcystins and some related toxic compounds [12]. Mass spectra only extracted from literature articles and Internet sites were entered into this library version. The library was

 DOI : 10.35291/2454-9150.2026.0007

cross-validated by test searches which resulted in the true positive rate of ≥73%. In the article, we report on the continuation of the work [12] and present the second/enlarged version of the tandem mass spectral library of microcystins, amatoxins, and other cyclic peptides and related compounds and discuss the library performance in identification of these compounds.

In the new version, the library was essentially updated. First, experimental product-ion mass spectra obtained by both ESI-MS^2 and matrix-assisted laser desorption ionization tandem time-of-flight (MALDI-ToF/ToF) technique and, second, new literature spectra, including 'one-dimension' (see below) data were entered in the second library. Library acquisition of all those spectra was necessitated to study a dependence of library search rates on the different kinds of spectra and take into account possibly all the literature spectral data on the cyclic peptides, including, among others, mass spectra of very rare/non-available compounds. Then, we continued to test the library for the efficiency of library searches leading to true results. The analysis of extracts from two cyanobacterial strains as real-world samples was the extra test performed. In the article, we outline the results of such validation tests and compare potentialities of the two general approaches to identification of cyclic peptides based on (a) library searches and (b) precise measurements of $[M+H]^+$ (or $[M+2H]^{2+}$) ion mass by means of high resolution mass spectrometry (HRMS).

1. Building mass spectral library

1.1 Library acquisition

Tandem spectra under the study were extracted from the literature/ Internet sources or recorded especially for this research.

To find spectra in the literature and WWW, more than 150 articles (or Internet sites/pages) on mass spectrometry of microcystins and related compounds were looked through. Of this amount, 47 articles ([9,13–33] and references [9–33] in the previous article [12]) with tandem mass spectra of the target compounds were selected by the following rules. These spectra had to be obtained for known compounds, e.g. reference materials/analytical standards, or analytes identified by several independent techniques. The spectra were rejected which (a) had been acquired in a very short *m/z* range and (b) contained very intense peaks not annotated or improperly annotated with *m/z* values. However, spectra with one or a few minor unannotated peaks were incorporated in the library. The maximum relative intensity of such signals was on the average of 15–20%. With that, most minor annotated peaks were taken into account.

Library searches with such or some other incomplete spectra available in the library may lead to false identification (see below). Other clear/assumed factors that may induce identification errors are (a) poor irreproducibility of tandem mass spectra, (b) reporting unrepresentative spectra in literature sources, (c) isobaric impurities in analytical standards, or (d) a distortion of mass spectra in printed/Internet copies. It should be noted we measured/reproduced peak intensities of spectral graphs with the uncertainty of not larger than about 1/20 of their values.

'One-dimension' literature mass spectra as the special ones were entered in this library for the first time. These are simply lists of *m/z* values without any linked peak intensities. This kind of defective mass spectral data was the only one available in some earlier publications, e.g. see [13,14]. In the library, these spectra were represented as common/'two-dimensional' ones with the same provisional relative intensity of 100%. If the last value was set as 30% or 10%, library searches (see below) tended to be a bit less efficient.

In generating experimental spectra, ESI and MALDI mass spectrometry was used (see below). Tens of experimental spectra were also incorporated in the library. ESI tandem spectra were recorded for seven pure substances: microcystins-LR, -RR, -LA, and -YR, alpha- and beta-amanitin, phalloidin. Those are the most cited microcystins (the first four compounds, see below) and the principal mushroom toxins. Corresponding MALDI fragment spectra were obtained for six compounds: microcystins-LR, -RR, and -YR, alpha- and beta-amanitin, phalloidin.

The *m/z* values of experimental and literature spectra rounded off to nearest integer and corresponding relative peak intensities were transferred to the library created as the low resolution version using the NIST MS Search 2.0 d software (NIST, Gaithersburg, MD, USA) [34]. Eventually, the electronic collection contained 263 mass spectra of 59 unique compounds. The spectra acquired by ESI with the use of mass analyzers of ion trap (IT), triple quadrupole (TQ), and quadrupole time-of-flight (Q-ToF) predominated. The most numerous were mass spectra of microcystin-LR, the most toxic microcystin of the group, and also microcystins-YR, -RR, -LA, and b-

 DOI : 10.35291/2454-9150.2026.0007

amanitin. These and other performances of the library are given in Tables 1 and 2. Mass spectra are exemplified in Fig. 1 for microcystin-LR.

In library searches performed by means of the MS Search pro-gram, settings of 'Similarity' and 'MS/MS' were selected for the 'Spectrum Search Type' option, with the precursor *m/z* value set in the 'Precursor Ion' box. For other details of searches, see [12]. In order to get realistic rates of different search results for miscellaneous data which may be relevant to much larger spectral libraries than built one, the 'bsa_2011_04_01_it' library [35] of 725 tandem mass spectra of 'common' (non-cyclic) peptides produced from bovine serum albumin, was added to the our data collection of cyclic ones. It was supposed that some amino acid fragments of cyclic and non-cyclic compounds were the same that might result in false non-identification of cyclic peptides. That library was provisionally taken and other options of linear-peptide libraries [35] led to about the same results.

1.1 E*xperimental*

Standard solutions of 10 lg/ml microcystins-LR, -RR, -LA, and –YR (purity >95% by HPLC) were from Abraxis (USA). The methanol solutions of the same concentration of alpha- and beta-amanitin, phalloidin were prepared from their solid-state standards (purity ≥90% by HPLC/high performance capillary electrophoresis, Sigma-Aldrich, USA). To obtain ESI-MS (MS^2) spectra, all the solutions were diluted twice with the 1:1 mixture of 0.05% trifluoric acid and acetonitrile and injected into the Maxis 4G ESI-Q-ToF mass spectrometer (Bruker, Germany). The two chemicals were of pure grade. The injection rate was 3 ll/min. The standard values of the needle and cone voltages, collision nitrogen pressure, and other experimental parameters were set. The high, intermediate, and low collision energy were set in the range of 10–75 eV. For each compound, three reference spectra obtained at different energy and then incorporated into the library were consisted from 30 ToF scans. The precursor ions were $[M+H]^+$ (all the analytes excluding microcystin-RR) and also $[M+2H]^{2+}$ species for four com-pounds (microcystins-LR, -RR, -YR, and beta-amanitin).

To acquire MALDI mass spectra, 1–2 ll aliquots of 10 lg/ml solutions of each of those compounds were combined with

Table 1

Principal features of the mass spectral library.

Feature	Data, value
Articles	47
Compounds	59
Spectra	263
Spectra origin (the number of spectra)	Literature (131), literature 'one-dimension' (78), experimental (54)
Ionization techniques (spectra)	ESI (199), FAB (33), MALDI (24), APCI (4), n.a. (3)
Mass analyzers (spectra)	IT (66), TQ (63), Q-ToF (39), electrostatic and magnetic sector (35), ToF/ToF (24), linear IT (13), IT-ToF (8), Orbitrap (5), other/n.a. (10)
Ion mode (spectra)	Positive (246), negative (17)
MS^n (spectra)	MS^2 (240), MS^3 (23)
Precursor ions[a] (spectra)	MS^2: $[M+H]^+$ (160), $[M+2H]^{2+}$ (39), in-source fragmentation (11), $[M+H—H_2O]^+$(10), $[M+Na]^+$(7), $[M-H]^-$ (9), other (4), MS^3: $[fragment]^+$ (15), $[fragment]^-$ (8)

[a] $[fragment]^{\pm}$ is a fragment ion formed from the first precursor ion by the elimination of neutral species with the mass of some tens or hundreds Da.

Table 2

Group	The number of spectra	Individual compound (the number of spectra)
Microcystins	176	Microcystins-LR (47), -YR (21), -RR (19), -LA (15), -LF (12), -LW (6), [ADMAdda^5]microcystin-LR (6), [D-Asp^3,ADMAdda^5]microcystin-LR (5), [ADMAdda^5]microcystin-Lhar (5), [ADMAdda^5,Mser^7]microcystin-LR (3), [D-Asp^3]microcystin-LR (3), [DMadda^5]microcystin-LR (3), [D-Asp^3,ADMAdda^5]microcystin-Lhar (3), [Dha^7]microcystin-RR (3), [D-Ser^1,ADMAdda^5]microcystin-LR (3), microcystin-WR (2), [Dha^7]microcystin-LR (2), microcystin-LR, glutathione conjugate (2), [D-Asp^3,E-Dhb^7]microcystin-RR (2), 14 other compounds (1)
Amatoxins	29	Beta-Amanitin (15), alpha-amanitin (10), amanexitide (3), gamma-amanitin (1)
Cyanopeptolins	16	Microcystilide A (6), microcystilide C (5), microcystilide B (4), aeruginopeptin 228-A (1)

 DOI : 10.35291/2454-9150.2026.0007

Nodularins	10	Nodularin (5), [D-Asp[1]]nodularin (3), [DMAdda[3]]nodularin (1), [(6Z)-Adda[3]]nodularin (1)
Phallotoxins	9	Phalloidin (8), phallacidin (1)
Anabaenopeptins	7	Keramamide F (3), anabaenopeptin A (2), anabaenopeptin B (2),
Other	16	9 Compounds (1 or 2)

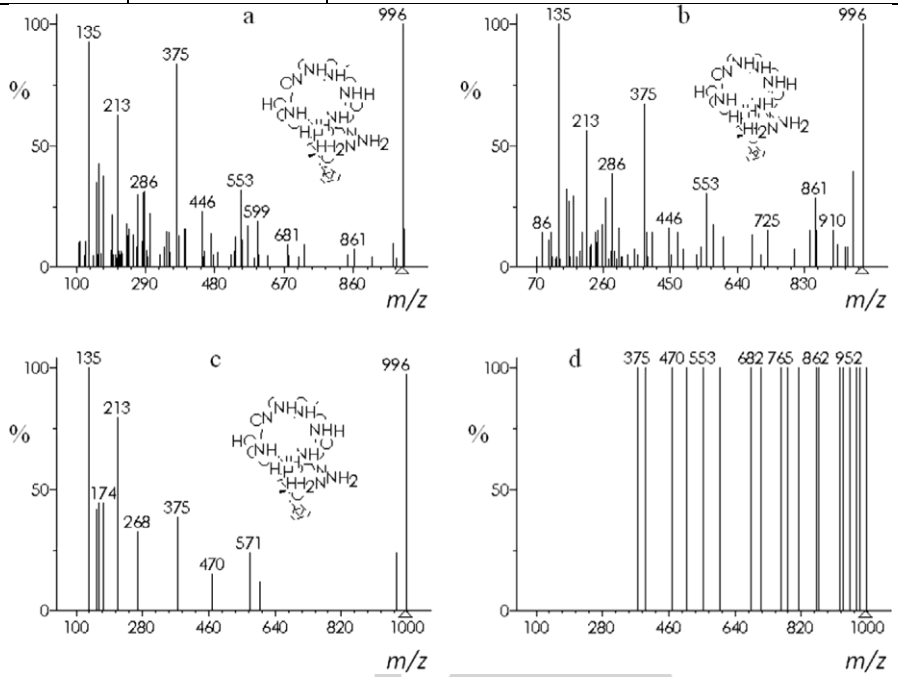

Fig. 1. Experimental ESI-MS² mass spectrum of microcystin-LR (a) and its the most similar spectra of the different types: experimental MALDI (b), literature (c), and literature 'one-dimension' (d) mass spectra. Specifications: (a) ESI-Q-ToF, collision energy 65 eV; (b) MALDI ToF/ToF, laser power 40–56%, MF 472 for matching the spectrum (a); (c) ESI-TQ, collision energy 20 eV, MF 209; (d) ESI-IT, relative collision energy 32%, MF 142.

1–2µl solution of alpha-cyano-4-hydroxycinnamic acid (Bruker, Germany) in 2:1 mixture of 0.1% trifluoric acid and acetonitrile. A 0.5 ll volume of the mixture was deposited onto a standard MALDI plate (Bruker Daltonics) for the forthcoming mass analysis. The UltrafleXtreme MALDI TOF/TOF mass spectrometer (Bruker Dal-tonics) was used. MALDI-ToF spectra were acquired by 500–2000 shots in the positive ion reflector mode at 1000 Hz UV laser frequency and 20% and 40% laser power. MALDI TOF/TOF fragment spectra acquired by means of the LIFT technique [36] were com-posed from precursor ion peaks recorded at the above conditions to which were added product ion spectra acquired by 500–2500 shots at the same laser frequency and 28% and 56% power, respectively. A collision gas was not specially supplied. Four spectra of each compound obtained at the combination of power of 20 + 28% and 40 + 56%, two spectra for the pair of powers, were entered into the library. In these, most mass peaks were the same for the unique compound with some variations of their relative intensities. For example, there were variations of the $[M+H]^+$ peak intensity within the range of approximately 50–100%.

From seven standard compounds, MALDI product-ion spectra were acquired for six ones. In the case of microcystins-LA, the sig-nal of poorly fragmenting $[M+Na]^+$ ions was only observed.

The water-methanol extracts of the two bloom-forming cyano-bacterial strains of *Microcystis aeruginosa* 972 and 973 (see [37]) with microcystin concentration of <1 lg/ml, were kindly provided by Dr. Nadezhda Medvedeva of Research Centre for Ecological Safety of Russian Academy of Sciences (Saint Petersburg, Russia). These samples were analyzed in the same conditions as the above standards.

1.1 Literature databases

To search pertinent publications on the compounds under consideration, a series of databases and retrieval systems were used. From them, Chemical Abstracts and Google Scholar were suitable to determine an abundance of these chemical compounds measuring by the number of their occurrences in a database, i.e. the number of relevant articles (see [10, p.141]).

1.2 Validation

A validation procedure is a series of tests that are library searches in which spectra of known compounds are

 DOI : 10.35291/2454-9150.2026.0007

considered as ones of unknown analytes (test compounds) and compared to the library reference spectra. Mass spectra of different precursor ions of the same cyclic peptides were considered as spectra of different compounds. We used the procedure of cross-validation in which library spectra in turn served as test ones. All the details of this procedure are described in the publication [12] on the first version of this library.

There were two kinds of validation tests relating to positive and negative control of the result trueness. In the first case, the library contained two or more mass spectra of unique compounds with the same *m/z* of precursor ions. Negative control implied that the only spectrum of a cyclopeptide was available in the library. In searches, if that had become a test spectrum, the search result belonged to foreign compounds.

The essential result of the validation was the threshold value of match factor ($MF_{cut\text{-}off}$). The $MF_{cut\text{-}off}$ threshold was estimated during the validation process by minimizing the overall percent of false results (see below) with the use of the ROC graph [11,12].

In positive control searches, the reference spectrum of the 1st rank in the hit list with its MF in excess of the $MF_{cut\text{-}off}$ value was considered as positive result. The latter was TP or FP if both test and reference spectra belonged to the same peptide or different compounds, respectively. For the case of $MF < MF_{cut\text{-}off}$, the 1st line search result was classified as negative one. In negative control, that was TN. With the inverse ratio of $MF > MF_{cut\text{-}off}$, that search result was determined as FP. Then the rates as percentages of TP (TPR) and TN (TNR) in relation to the number of all corresponding searches were estimated. Statistical features of the 1st rank results, e.g. corresponding MF values, irrespective from result trueness, could be independently analyzed. The result rates estimated for the entire library and its deferent parts will be dis-cussed below.

MS^3 spectra were not involved into test searches. The NIST spectra of linear peptides were only used as references. In the validation process, a few test results were uncertain and not taken into account.

2 Results and discussion

2.1 General results

The validation results are given in Table 3. It is appropriate to compare these to corresponding values of the previous research for the small library of literature spectra. The principal value was 73% TP [12]. Here the corresponding result is 75% for the literature sublibrary (Table 3) that is very close to the previous value. Two other sublibraries are new in the second version.

The experimental sublibrary includes 54 ESI and MALDI production mass spectra recorded in our laboratory for 6–7 cyclic pep-tides. This implies 7.7 replicate spectra available per a compound. The last value is relatively high one that is essential for obtaining true search results. Indeed, it is known that both the TPR and the mean MF usually tend to rise with the number of replicate 'unknown' and reference spectra acquired for a unique compound [10, p.186; 11]. It is explained by increasing a probability of matching spectra of the same compounds. Searches performed only for the spectral subset under consideration, i.e. when both test and reference spectra were experimental, led to 100% TP (Table 3). The average MF in these searches is within the range of 590–648 that is higher than the average value of 396 for the 1st rank spectra in all the test searches.

Both ESI and MALDI fragment spectra are obviously closer to each other in every of the two groups than between groups. The spectral resemblance seems to be due to a similarity in experimental conditions which are rather similar within the groups although collision energy and partly laser shots/power are not the same. Also, ESI product-ion mass spectra match rather well with MALDI ones and vice versa. As the result, in all the test searches reference spectra of these two groups may substitute for one another, retaining high outcome of 86–95% initial TP obtained for both groups without their elimination.

It should be supplemented that MALDI product-ion spectra are very seldom to be entered in mass spectral libraries. To our best knowledge, the only exclusion is the group of MALDI ToF–ToF spectra obtained by means of high-energy collision-induced dissociation and incorporated in MassBank [38]. Our library is the first one containing MALDI LIFT ToF/ToF spectra (see also [39]).

Another part of the library is 'one-dimension' spectra. Searches performed only within this type of data resulted in the low rate of 36% TP. We explain this low value by frequent false matching many 'one-dimension' spectra of different compounds because of the same *m/z* of many fragment ions with the same 100% intensity of their peaks.

Basic validation results refer to the sum of sublibraries and searches in the entire library. The overall TPR of 70% is not very high one. The rate is diminished due to the unconventional contribution of defective/'one-dimension' data.

 DOI : 10.35291/2454-9150.2026.0007

Without that, the correct results that are the 1st rank above-threshold matches appeared in 88% of cases (Table 3). However, incorporation of 'one-dimension' subset into only reference spectra set increases the rate of TP from 88% to 91% (Table 3). We think that this effect is due to rising total probability of spectral match of the same compounds with increasing the number of replicate spectra (see below). It is also clear that here probable false matches between 'one-dimension' spectra are impossible to occur.

Table 3

Spectra		The number of searches	TP (%)
Test	Reference		
All	All	233	70
Experimental and literature	All	157	91
Experimental and literature	Experimental and literature	157	88
Experimental	Experimental	51	100
Literature	Literature	106	75
Literature 'One-dimension'	Literature 'One-dimension'	76	36

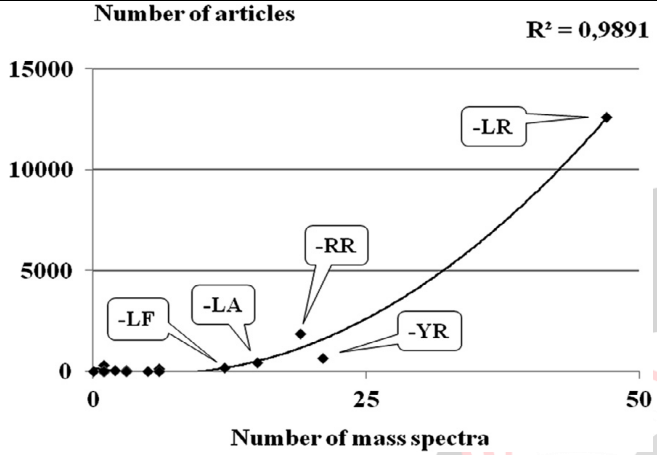

Fig. 3. The correlation between the overall number of mass spectra in the library and the number of papers issued before 2014 and retrieved by searches with the use of Google Scholar. The abundant microcystins are indicated

I consider the rate of 88–91% as the principal one for the second version of this library. The rate is higher than the previous one (73%) and corresponds to a good efficiency level of modern libraries [12]. The difference between the two rates for correct answers seems to be caused firstly by increasing an average number of library spectra per compound. The latter is so because the value of the spectra/compound ratio increased from 2.7 to 4.5 on the enlargement of the library from the initial version of 75 spectra to the current status of 263 ones.

3.2 Different compounds

Validation results and related identification potential of the library are not the same for different compounds under research. Taken as a whole, the trend of increasing TPR with the number of replicates (see above) is also expressed for individual com-pounds and this correlation is not strong (Fig 2). Test searches for compounds having 8 and more reference spectra of the same precursor ion (the $[M+H]^+$ ions of microcystins-LR, -LA, -YR, a- and b-amanitin, phalloidin and also the $[M+2H]^{2+}$ ion of microcystin-RR) without 'one-dimension' data led to high TPR of 96%. Here the correlation between the two quantities which are the number of spectra in the library and the abundance of chemical compound measured by the number of corresponding publications (Fig. 3), is advantageous to see. Thus identification of such abundant compounds based on this library is certainly reliable.

Target identification of those analytes with the use of HRMS[1] and measurement of accurate mass of protonated molecules may also reliable. Every abundant cyclopeptide under consideration has the unique molecular formula and therefore molecular mass discriminating it from other peptides of this group. Moreover, the number of known compounds with the same formulas is relatively poor as estimated based on Chemical Abstract database (Table 4).

 DOI : 10.35291/2454-9150.2026.0007

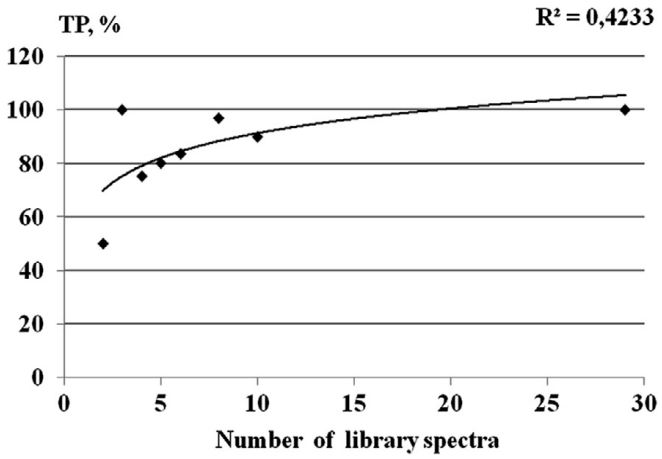

Fig. 2. The dependence of TPR on the number of mass spectra of individual compounds. The values were averaged for the same numbers. 'One-dimension' data were not considered.

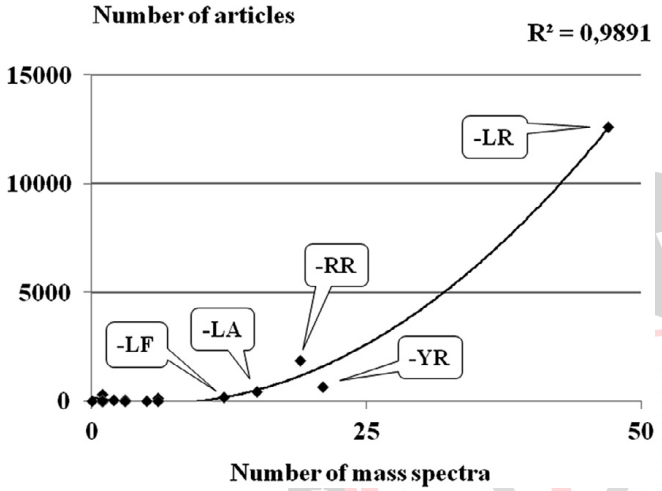

Fig. 3. The correlation between the overall number of mass spectra in the library and the number of papers issued before 2014 and retrieved by searches with the use of Google Scholar. The abundant microcystins are indicated.

However, the availability of only $[M+nH]^{n+}$ ion peaks in ESI-MS[1] mass spectra is not generally sufficient for identification at its confirmation level (this evidence is the equivalent of only two identification points instead of 3–4 ones required for certainly true results [10, p.123]). So combination of both library and HRMS approaches should provide reliable recognition of such molecules, particularly, in samples of unknown origin. For this purpose, the HRMS version of our library which could be easily built on the base of subset of mass spectra acquired on high resolution mass spectrometers would be especially effective.

On the other hand, relatively rare peptides with only two library spectra, i.e. a pair of one test spectrum and the only reference one, involved in test searches for a unique compound, did not find out very reliable identification. Here, TPR is not higher than 50%. The use of HRMS cannot also be *a panacea* for recognition of many rare cyclopeptides because of their isomerism. For example, there are not less than four known cyclic isomers of demethyl microcystin-LR, two of them occurring with comparable probability (Table 4). Those are not differentiated by HRMS[1] and are partly distinguished by tandem mass spectra [40]. Also, there may be linear isomers of those cyclopeptides (Table 4). Obviously, in order to be granted a chance to reliably identify isomer structures the library should be far enlarged with their replicate mass spectra.

3.3 False positives

The majority of FP results were due to different cyclic peptides similar in structure to 'true' compounds related to test spectra. It can be seen for the case of searches among literature spectra resulted in TPR of 75% (see Table 3). There were six FPs in positive control and four FPs in negative one recorded in test searches for this sublibrary. In the first group, the difference in structure between the test compound as TP and the falsely identified ones is the only amino acid, i.e. it is the minor difference (Table 5). In the second case, compared structures are the same, with different precursor ions, or are differed by the substructure ranging from hydroxyl

 DOI : 10.35291/2454-9150.2026.0007

to the three amino acid residue (Table 5).

Linear peptides sometimes occupied the 1st position in hit-lists. All corresponding MFs were low and under $MF_{cut\text{-}off}$ thresholds for positive results. Correspondingly, there were no linear compounds among total of 23 false positive results obtained at positive and negative control during the validation of the entire library. Thus false identification of an individual cyclic peptide implies true group recognition of cyclic peptides as the class of natural com-pounds and even its subclass (e.g. microcystins, see Table 5).

Table 4

Formula	The number of occurrences	Compound	Occurrences (%)
$C_{49}H_{74}N_{10}O_{12}$	1103	Microcystin-LR	98
$C_{49}H_{75}N_{13}O_{12}$	371	Microcystin-RR	97
$C_{35}H_{48}N_8O_{11}S$	307	Phalloidin	98
$C_{52}H_{72}N_{10}O_{13}$	205	Microcystin-YR	95
$C_{39}H_{54}N_{10}O_{14}S$	129	alpha-Amanitin	100
$C_{46}H_{67}N_7O_{12}$	84	Microcystin-LA	95
$C_{52}H_{71}N_7O_{12}$	75	Microcystin-LF	100
$C_{54}H_{72}N_8O_{12}$	67	Microcystin-LW	100
$C_{39}H_{53}N_9O_{15}S$	36	beta-Amanitin	92
		average	*98*
$C_{48}H_{72}N_{10}O_{12}$	61	[D-Asp3]microcystin-LR	31
		[Dha7]microcystin-LR	25
		[D-Asp3,Dhb7]microcystin-LR	7
		[DMadda5]microcystin-LR	2
		Demethyl microcystin-LR, other or uncertain	5
		17 linear peptides, total	31

Occurrence of formulas and related compounds in Chemical Abstracts, 2007–2013.

Table 5

Test compounds and falsely identified compounds. Literature spectra.

Test/TP	FP identification	Difference in structure
Positive control		
[Dha7]Microcystin-RR	Microcystin-RR	1 aa[a]
Microcystilide C	Microcystilide A	1 aa
Microcystin-LF	Microcystin-LR	1 aa
Microcystin-LR	[ADMAdda5]microcystine-LR	1 aa
Microcystin-RR	[Dha7]Microcystin-RR	1 aa
Microcystin-YR	Microcystin-LR	1 aa
Negative control		
		2 aa
[ADMAdda5]microcystine-Lhar	micro-cystine-LR [D-Asp3,ADMAdda5]	3 aa
7-Deoxycylindrospermopsin	Cylindrospermopsin[b]	OH
Microcystin-LR ($[M+H+NH_4]^{2+}$)	Microcystin-LR ($[M+2H]^{2+}$)	Precursor ion

[a] aa is amino acid residue.

[b] Cyanotoxin, polycyclic uracil derivative.

3.4 The supplementary identification criterion

For reliable identification, there should be matching *m/z* of pre-cursor ions of compared spectra. Here I did not use this criterion of identification because most library/test compounds are of different formula/molecular mass. Correspondingly, *m/z* value set in the ''Precursor Ion'' box (see above) unambiguously determined com-pound identity in most searches. So, test searches with the use this criterion should not led to results very different from obtained ones. Indeed, with the criterion of matching *m/z* of precursor ions of test and reference spectra the overall TP rate increased from 70% to only 71%. The test performed in larger libraries resulted in more increasing search performances [11].

3.5 Identification of cyanobacterial peptides

A library performance was also examined in the example of recognition of cyclopeptides in extracts from the two

 DOI : 10.35291/2454-9150.2026.0007

M. aeruginosa 972 and 973 cyanobacterial strains. This qualitative analysis was very briefly mentioned in our presentation [41]. Three identifications related to principal peaks in the MALDI spectra of extracts are considered below.

The 1st case refers to the strain 973 and the analyte corresponding to the ion with integral-valued *m/z* 996. The library search with the ToF/ToF spectrum, see Fig. 4b, resulted in the high MFs and microcystin-LR as the same identification candidate in top lines of the hit list leaves no room to doubt that this abundant microcystin is reliably recognized.

There were more difficult problems to identify main components of the extract of the strain 972 (Fig. 5a). Here, the nature of the ion *m/z* 1025 is relatively clear (Fig. 5a). Indeed, HRMS demonstrated that it is demethylmicrocystin-RR as its molecular formula

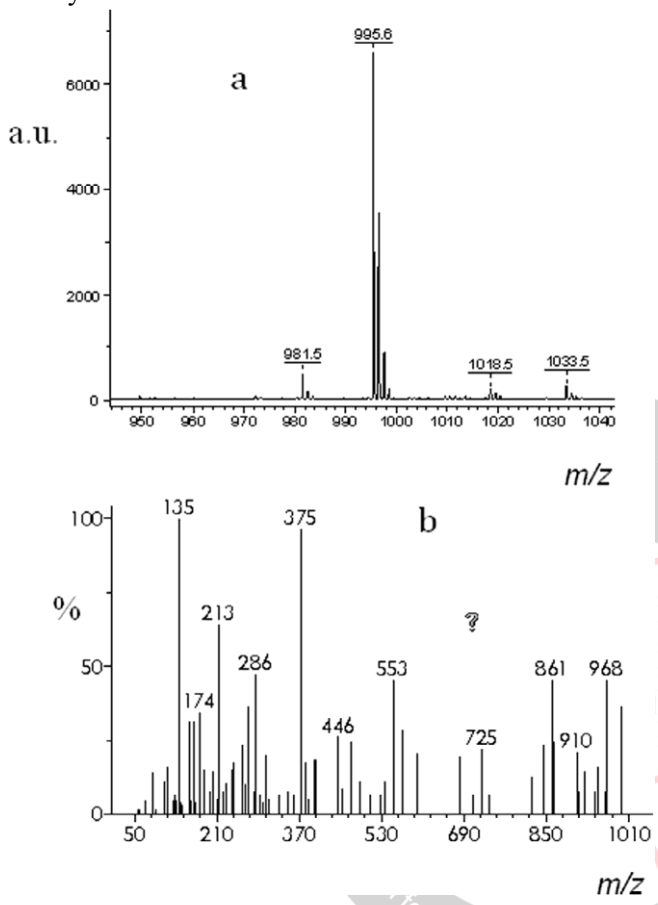

Fig4. MALDI ToF (a, the region of the principal peak) and ToF/ToF (b, the precursor ion *m/z* 996) spectra of the strain 973 extract. The result of the library search of the latter is the 1st rank library spectrum of microcystin-LR, Fig. 1a, MF 883, at $MF_{cut\text{-}off}$ 150. Also, the spectra of the rank from 2nd through 11th belong to the same microcystin-LR.

 DOI : 10.35291/2454-9150.2026.0007

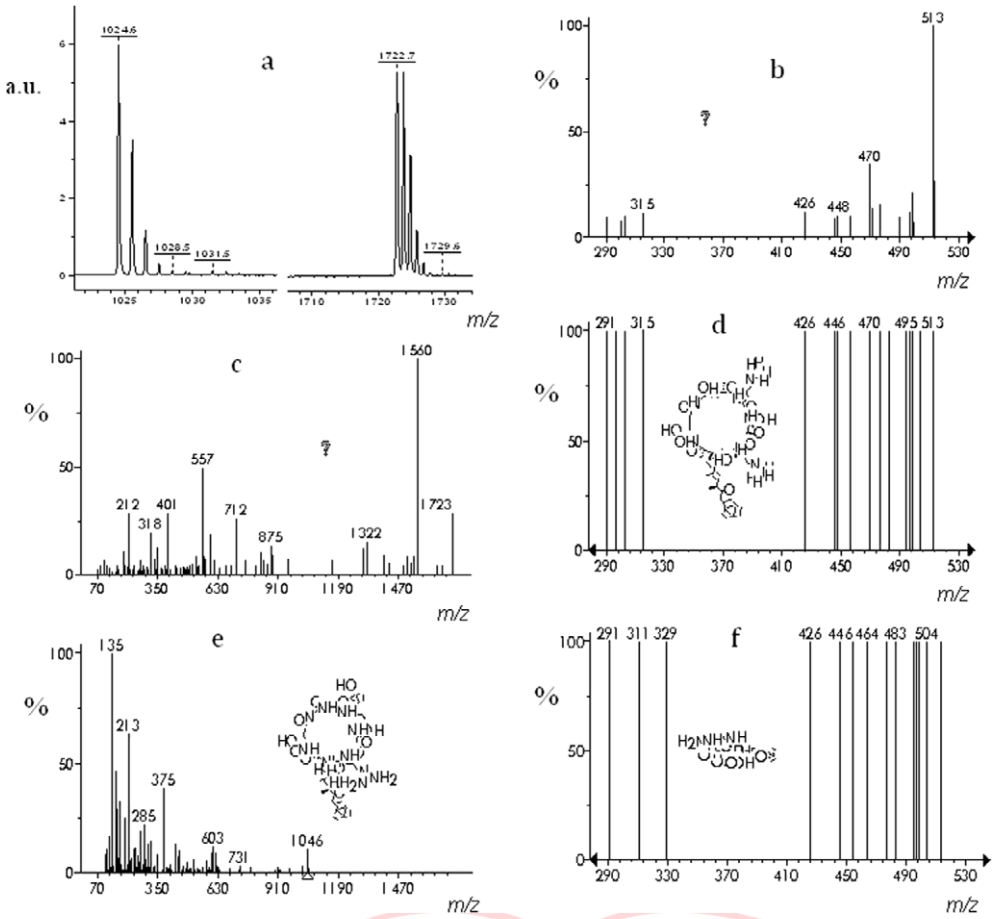

Fig. 5. MALDI ToF (a, principal peaks) and ToF/ToF (c, the precursor ion *m/z* 1723) spectra and ESI-MS^2 spectrum (b, the $[M+2H]^{2+}$ precursor *m/z* 513) for the strain 972 extract. The latter precursor corresponds to the $[M+H]^+$ singly-charged ion, *m/z* 1025, recorded in the conditions of MALDI (a). The library search for the spectrum (b) resulted in the 1st and 2nd rank library spectra belonging to [Dha7]microcystin-RR (MF 447, d) and its isomer [D-Asp3]microcystin-RR (MF 205, f), respectively. Here, $MF_{cut\text{-}off}$ is 150. In the case of the MALDI spectrum (c) of the precursor ion *m/z* 1723, the search did not provide similar spectra; the best-score reference spectrum showed the near-zero MF of 18 (microcystin-YR, e).

$C_{49}H_{73}N_{11}O_{13}$ had been confirmed with the accuracy of 3.1 ppm (the $[M+H]^+$ ion in MALDI) and 0.1 ppm (the $[M+2H]^{2+}$ ion, ESI-MS^2). Library searches with ESI spectrum (Fig. 5b) led to the two above-$MF_{cut\text{-}off}$ candidates which are isomeric [Dha7]microcystin-RR (Fig. 5d) and [D-Asp3]microcystin-RR (Fig. 5f). However, because of not very high MF values and a few reference spectra of different demethylmicrocystins-RR including [D-Asp3,E-Dhb7]microcystin-RR [42] available in the library, from one through three spectra, we cannot definitely identify the particular isomer and prefer ‘‘demethylmicrocystin-RR’’ as the result statement.

In the 3rd case, there can be no any identification of the ion *m/z* 1723 (Fig. 5a) based on library searches because of the absence of corresponding library spectra. Hit list contained only spectra (e.g. in Fig. 5e) not matching to the corresponding MALDI fragment spectra (Fig. 5c) and therefore resulted in only TNs. So, the approach to identification based on MS libraries is not applicable for especially rare/novel compounds without references.

One might assume on mass level grounds that there was one from microviridins, a group of cyclic depsipeptide (see [43]). Recently, ‘‘microviridin 1721’’ with *m/z* 1722.8 of the [M+H]+ ion was detected [43]. We measured the accurate mass of this ion (1722.7516) and also the $[M+2H]^{2+}$ precursor (*m/z* 861.8810) and its fragments in ESI-MS^2 and could not conclude on the structure of this molecule. The reason is a variety of plausible candidate for-mulas (tens) and sequences (hundreds) for these masses within the common mass tolerance (<5 ppm).

Conclusion

The library of tandem mass spectra originated from many publications and our own experiments and consisted of 263 MS^n spectra (mainly MS^2 spectra) of 59 cyclic peptides and a few related compounds, was build for identification of toxins of this group. The library was tested by means of cross-validation. Tests demonstrated that

 DOI : 10.35291/2454-9150.2026.0007

the percentage of correct identifications based on the best-match product ion mass spectra of 'unknowns' and references was 70% in a general case and 88% without defective/ 'one-dimension' spectra, i.e. simple lists of *m/z* values. However, incorporation of such spectra in the reference set of mass spectra in the course of the library validation somewhat increased TPR (up to 91%). So, the percentage of 88–91% is the principal estimate for the overall performance of this library.

ESI-MS^2 spectra appeared to be rather close to MALDI-ToF/ToF fragment spectra. Therefore reference spectra of these two groups may be to a large extent substitutable for one another. So, fragmention variants of MALDI such as the LIFT technique used in this research have a promise as general techniques of building mass spectral libraries.

The consideration of the nature of false positive results shows that this library provides not only a method of choice for identification of individual cyclopeptides and also one for the group recognition of individual classes of these toxins. The approach to identification of cyclopeptides based on mass spectral library matching should be the most effective for abundant 'known unknown' compounds. Here *de novo* identification methods and related ones taken from genomics/proteomics (see Introduction) are not required. Also, this database should be enlarged to reliably determine relatively rare peptides. If this is not the case, the use of the library in qualitative analysis of real-world samples leads to group/ambiguous identification results or even no results.

References

1. H.K. Hudnell (Ed.), Cyanobacterial Harmful Algal Blooms: State of the Science and Research Needs, Springer, New York, 2008.
2. T. Wieland, Peptides of Poisonous Amanita Mushrooms, Springer-Verlag, New York, 1986.
3. J.E. Velásquez, W.A. van der Donk, Genome mining for ribosomally synthesized natural products, Curr. Opin. Chem. Biol. 15 (2011) 11–21.
4. T. Krishnamurthy, L. Szafraniec, D.F. Hunt, et al., Structural characterization of toxic cyclic peptides from blue-green algae by tandem mass spectrometry, Proc. Natl. Acad. Sci. USA 86 (1989) 770–774.
5. H. Mazur-Marzec, L. Spoof, J. Kobos, M. Plin´ ski, J. Meriluoto, Cyanobacterial hepatotoxins, microcystins and nodularins, in fresh and brackish waters of the Pomeranian Province, northern Poland, Oceanol. Hydrobiol. Stud. 37 (2008) 3–21.
6. H. Mohimani, W.T. Liu, J.S. Mylne, et al., Cycloquest: identification of cyclopeptides via database search of their mass spectra against genome databases, J. Proteome Res. 10 (2011) 4505–4512.
7. C.W. Diehnelt, S.M. Peterman, W.L. Budde, Liquid chromatography–tandem mass spectrometry and accurate *m/z* measurements of cyclic peptide cyanobacteria toxins, Trends Anal. Chem. 24 (2005) 622–634.
8. W.T. Liu, J. Ng, D. Meluzzi, et al., Interpretation of tandem mass spectra obtained from cyclic nonribosomal peptides, Anal. Chem. 81 (2009) 4200–4209.
9. T.H. Niedermeyer, M. Strohalm, MMass as a software tool for the annotation of cyclic peptide tandem mass spectra, PloS One 7 (2012) e44913, http:// dx.doi.org/10.1371/journal.pone.0044913.
10. B.L. Milman, Chemical Identification and its Quality Assurance, Springer, Berlin, 2011.
11. B.L. Milman, I.K. Zhurkovich, Towards a full reference library of MSn spectra. II: a perspective from the library of pesticide spectra extracted from the literature/Internet, Rapid Commun. Mass Spectrom. 25 (2011) 3697–3705.
12. B.L. Milman, I.K. Zhurkovich, Tandem mass spectral library of microcystins and related compounds, J. Anal. Chem. 68 (2013) 1188–1194.
13. M. Namikoshi, K.L. Rinehart, R. Sakai, et al., Identification of 12 hepatotoxins from a Homer Lake bloom of the cyanobacteria *Microcystis aeruginosa*, *Microcystis viridis*, and *Microcystis wesenbergii*: nine new microcystins, J. Org. Chem. 57 (1992) 866–872.
14. K. Sivonen, M. Namikoshi, W.R. Evans, M. Fardig, W.W. Carmichael, K.L. Rinehart, Three new microcystins, cyclic heptapeptide hepatotoxins, from *Nostoc* sp. strain 152, Chem. Res. Toxicol. 5 (1992) 464–469.
15. M. Namikoshi, B.W. Choi, R. Sakai, et al., New nodularins: a general method for structure assignment, J. Org. Chem. 59 (1994) 2349–2357.
16. M. Namikoshi, F. Sun, B.W. Choi, K.L. Rinehart, W.W. Carmichael, W.R. Evans, V.R. Beasley, Seven more Microcystins from Homer Lake Cells: application of the general method for structure assignment of peptides containing alpha, beta-dehydroamino acid unit (s), J. Org. Chem. 60 (1995) 3671–3679.
17. M. Yuan, M. Namikoshi, A. Otsuki, K. Sivonen, Effect of amino acid side-chain on fragmentation of cyclic peptide ions: differences of electrospray ionization/ collision-induced decomposition mass spectra of toxic heptapeptide microcystins containing ADMAdda instead of Adda, Eur. Mass Spectrom. 4 (1998) 287–298.
18. M. Yuan, M. Namikoshi, A. Otsuki, K.L. Rinehart, K. Sivonen, M.F. Watanabe, Low-energy collisionally activated decomposition and structural characterization of cyclic heptapeptide microcystins by electrospray ionization mass spectrometry, J. Mass Spectrom. 34 (1999) 33–43.
19. K. Karlsson, V. Sipiä, H. Kankaanpää, J. Meriluoto, Mass spectrometric detection of nodularin and desmethylnodularin in mussels and flounders, J. Chromatogr. B 784 (2003) 243–253.
20. C. Dell'Aversano, G.K. Eaglesham, M.A. Quilliam, Analysis of cyanobacterial toxins by hydrophilic interaction liquid chromatography–mass spectrometry, J. Chromatogr. A 1028 (2004) 155–164
21. I. Oksanen, J. Jokela, D.P. Fewer, M. Wahlsten, J. Rikkinen, K. Sivonen, Discovery of rare and highly toxic microcystins from

 DOI : 10.35291/2454-9150.2026.0007

lichen-associated cyanobacterium *Nostoc* sp. strain IO-102-I, Appl. Environ. Microbiol. 70 (2004) 5756–5763.

22. P.M. Ortea, O. Allis, B.M. Healy, M. Lehane, A. Nı´ Shuilleabháin, A. Furey, K.J. James, Determination of toxic cyclic heptapeptides by liquid chromatography with detection using ultra-violet, protein phosphatase assay and tandem mass spectrometry, Chemosphere 55 (2004) 1395–1402.
23. C. Kubwabo, N. Vais, F.M. Benoit, Characterization of microcystins using in-source collision-induced dissociation, Rapid Commun. Mass Spectrom. 19 (2005) 597–604.
24. C.W. Diehnelt, N.R. Dugan, S.M. Peterman, W.L. Budde, Identification of microcystin toxins from a strain of Microcystis aeruginosa by liquid chromatography introduction into a hybrid linear ion trap-fourier transform ion cyclotron resonance mass spectrometer, Anal. Chem. 78 (2006) 501–512.
25. O. Allis, J. Dauphard, B. Hamilton, A. Ni Shuilleabhain, M. Lehane, K.J. James, A. Furey, Liquid chromatography-tandem mass spectrometry application, for the determination of extracellular hepatotoxins in Irish lake and drinking waters, Anal. Chem. 79 (2007) 3436–3447.
26. M.S. Filigenzi, R.H. Poppenga, A.K. Tiwary, B. Puschner, Determination of a-amanitin in serum and liver by multistage linear ion trap mass spectrometry, J. Agric. Food Chem. 55 (2007) 2784–2790.
27. M.G. Antoniou, J.A. Shoemaker, A.A. De la Cruz, D.D. Dionysiou, LC/MS/MS structure elucidation of reaction intermediates formed during the TiO_2 photocatalysis of microcystin-LR, Toxicon 51 (2008) 1103–1118.
28. B. Puschner, C. Pratt, E.R. Tor, Treatment and diagnosis of a dog with fulminant neurological deterioration due to anatoxin-a intoxication, J. Vet. Emerg. Crit. Care 20 (2010) 518–522.
29. J. Rittgen, M. Pütz, U. Pyell, Identification of toxic oligopeptides in Amanita fungi employing capillary electrophoresis–electrospray ionization–mass spectrometry with positive and negative ion detection, Electrophoresis 29 (2008) 2094–2100.
30. L. Wu, P. Xie, J. Chen, D. Zhang, G. Liang, Development and validation of a liquid chromatography–tandem mass spectrometry assay for the simultaneous quantitation of microcystin-RR and its metabolites in fish liver, J. Chromatogr. A 1217 (2010) 1455–1462.
31. F.A. Dörr, D. Oliveira-Silva, N.P. Lopes, J. Iglesias, D.A. Volmer, E. Pinto, Dissociation of deprotonated microcystin variants by collision-induced dissociation following electrospray ionization, Rapid Commun. Mass Spectrom. 25 (2011) 1981–1992.
32. J. Jokela, L. Oftedal, L. Herfindal, P. Permi, M. Wahlsten, S.O. Døskeland, K. Sivonen, Anabaenolysins, novel cytolytic lipopeptides from benthic anabaena cyanobacteria, PloS One 7 (2012) e41222, http://dx.doi.org/10.1371/ journal.pone.0041222.
33. C.O. Miles, M. Sandvik, S. Haande, H. Nonga, A. Ballot, LC–MS analysis with thiol derivatization to differentiate [Dhb7]-from [Mdha7]-Microcystins: analysis of cyanobacterial blooms, planktothrix cultures and european crayfish from lake Steinsfjorden, Norway, Environ. Sci. Technol. 47 (2013) 4080–4087.
34. NIST MS Search Program, ver. 2.0d, NIST, 2005.
35. NIST Libraries of Peptide Tandem Mass Spectra. Available from: <http:// peptide.nist.gov>. (last accessed January 14, 2014).
36. D. Suckau, A. Resemann, M. Schuerenberg, P. Hufnagel, J. Franzen, A. Holle, A novel MALDI LIFT-TOF/TOF mass spectrometer for proteomics, Anal. Bioanal. Chem. 376 (2003) 952–965.
37. Y. Polyak, T. Zaytseva, N. Medvedeva, Response of toxic cyanobacterium *Microcystis aeruginosa* to environmental pollution, Water Air Soil Pollut. 224 (2013) 1–14.
38. MassBank. Record Index Results. Available from: <http://www.massbank.jp/jsp/Result.jsp?sortAction=1&exec=page&sorKey=name&totalPageNo=2&type =rcdidx&srchkey=MALDI-TOFTOF&idxtype=inst&pageNo=1>. (last accessed April 23, 2014).
39. L.B. Piotrovsky, B.L. Milman, D.N. Nikolaev, N.V. Lugovkina, E.V. Litasova, M.A. Dumpis, Characterization of fullerene derivatives by MALDI LIFT ToF/ToF mass spectrometry, Nanosyst. Phys. Chem. Math. 5 (2014) 76–80. Available from: <http://nanojournal.ifmo.ru/en/wp content/uploads/2014/02/NPCM51_P76-80.pdf>. (last accessed April 23, 2014).
40. B.L. Mil'man, Y.V. Russkikh, L.V. Nekrasova, Z.A. Zhakovskaya, An approach to the mass spectrometry identification of cyanobacterial peptides. The case of demethylmicrocystin-LR, J. Anal. Chem. 66 (2011) 1423–1431.
41. I.K. Zhurkovich, N.G.Medvedeva, I.A.Merkusheva, B.L. Mil'man, A.O.Rudenko, V.V.Chelovechkova, Analytical mass spectrometry of microcystins and related compounds, in: The Fifth Conference on Mass Spectrometry and its Applied Problems, Moscow, 2013.
42. C.O. Miles, M. Sandvik, S. Haande, H. Nonga, A. Ballot, LC–MS analysis with thiol derivatization to differentiate [Dhb7]-from [Mdha7]-microcystins: analysis of cyanobacterial blooms, Planktothrix cultures and European crayfish from Lake Steinsfjorden, Norway, Environ. Sci. Technol. 47 (2013) 4080–4087.
43. O. Czarnecki, M. Henning, I. Lippert, M. Welker, Identification of peptide metabolites of Microcystis (Cyanobacteria) that inhibit trypsin-like activity in planktonic herbivorous Daphnia (Cladocera), Environ. Microbiol. 8 (2006) 77–87.

 DOI : 10.35291/2454-9150.2026.0007