Analytical Chemistry Research 1 (2014) 8–15

Contents lists available at ScienceDirect

Analytical Chemistry Research

journal homepage: www.elsevier.com/locate/ancr

Identification of toxic cyclopeptides based on mass spectral library matching

Boris L. Milman [a,*], Inna K. Zhurkovich [b]

[a] Institute of Experimental Medicine, ul. Akad. Pavlova 12, 197376 Saint Petersburg, Russia
[b] Institute of Toxicology, ul. Bekhtereva 1, 192019 Saint Petersburg, Russia

ARTICLE INFO

Keywords:
Mass spectral library
Cyclic peptides
Microcystins
Chemical identification
Rates of qualitative analysis
Cyanobacterial strains

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 literature 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 spectra 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 fingerprinting (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

* Corresponding author. Tel.: +7 9217665296.
E-mail addresses: bmilman@mail.rcom.ru (B.L. Milman), zhurkovich@toxicology.ru (I.K. Zhurkovich).

http://dx.doi.org/10.1016/j.ancr.2014.06.002
2214-1812/

[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 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).

2. Building mass spectral library

2.1. 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 β-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 program, 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.

2.2. Experimental

Standard solutions of 10 μg/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 μl/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 compounds (microcystins-LR, -RR, -YR, and beta-amanitin).

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