New Functions of MassBank: Mass Spectral Database for Metabolome Analysis
- Peak Search by Formula, Web API and Database Maintenance -

Hisayuki Horai
horai@ttck.keio.ac.jp
Masanori Arita
arita@cb.k.u-tokyo.ac.jp
Nihei Yoshito
ynihei@ttck.keio.ac.jp

Ikeda Tasuku
tasukuik@ttck.keio.ac.jp
Yuya Ojima
ojima@ttck.keio.ac.jp
Takaaki Nishioka
takaaki@sfc.keio.ac.jp

1 Institute for Advanced Bioscience, Keio University, 14-1 Babacho, Tsuruoka, Yamagata 997-0035, Japan
2 Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba 277-8561, Japan
3 RIKEN Plant Science Center, 1-7-22 Suehirocho, Tsurumi, Yokohama, 230-0045, Japan

Keywords: mass spectral database, metabolome analysis, web service

1 Introduction

Mass spectrometry (MS) is widely used as a standard technology for exhaustive identification and quantification of metabolites. The recent technological progress of MS with high resolution and high sensitivity, as well as various combination with separation technologies such as liquid chromatography and capillary electrophoresis, enables to analyze a quite large number of metabolites [1]. Nevertheless, a large portion of the obtained spectra have been left without chemically identifying metabolites, because such identification is usually a manual task relying on analyst's individual expertise. A comprehensive and public database contributes to solve the bottleneck. As an implementation of such kind of database, we are developing a mass spectral database (MSDB) called MassBank (http://www.massbank.jp). Currently, MassBank contains more than 23,000 spectra of more than 12,000 chemical compounds including metabolites. MassBank provided several search services including a similarity search of spectra and a peak search by \( m/z \) values. This software demonstration focuses on the newest functions of MassBank.

2 New Functions of MassBank

2.1 Advanced Peak Search

Peak Search Advanced is a new search service of MassBank. It is a peak search by formulae of product ions or neutral losses. A peak search by formulae is more precise search from the viewpoint of chemo-bio informatics than a peak search by \( m/z \) values which cannot distinguish ions that have close molecular weight. For example, the formula and the exact mass weight of the protonated ion of cytosine is \( C_4H_6N_3O \) and 112.052 respectively. The results of Peak Search Advanced by \( C_4H_6N_3O \) are the spectra of 4 compounds which commonly have a cytosine skeleton as a substrate. On the other hand, a peak search by 112.052 results in the spectra of 21 compounds which includes the above mentioned 4 compounds and 17 compounds that do not have a cytosine skeleton. It means that Peak Search Advanced, which results in less false positives, are more accurate than a peak search by \( m/z \) value for searching spectra which include a specific ion peak. In this software demonstration, we show the difference of Peak Search Advanced and a peak search by \( m/z \) value.

In order to realize Advanced Peak Search, the molecular formula of the product ion should be identified for each peak in a spectrum. Current Peak Search Advanced is based on the results of our peak annotation task that predicts the molecular formulae and the chemical bonds cleaved in their precursor ions for the mass spectra measured on QqTOF-MS2 in MassBank. For example, the peak annotation for the \( m/z \) 59 peak of 2,3-diaminopropionate and glucosaminate results in different formulae \( C_2H_7N_2 \) and \( C_3H_7O \) respectively.
2.2 SOAP Web API

MassBank is a distributed database on the internet, and anyone can access the database from a web browser freely. We started a SOAP Web API [2] service of MassBank and MassBank is accessible by a user application software. An existing mass spectral analysis such as Mass++ [3] can be modified to search MassBank by user's spectrum directly. Current available functions of the MassBank SOAP API include similarity search of spectra, a peak search by m/z value and a peak pair search by the difference of m/z value. A user can combine them into a program which communicates with MassBank many a time and achieves the user's objectives. As same as MassBank Web service, a user application software can obtain results from all distributed MassBank servers by a single method call of MassBank Web API. Current available functions are limited to a part of Web services. We will enlarge the range of services of the SOAP API in near future.

2.3 Database Maintenance

We freely provide not only the MassBank internet database service but also the MassBank server itself. A user can download it from MassBank homepage and create the user's original MassBank database. MassBank system is so scalable that has capability to build a wide range of mass spectral databases including a personal spectral database, a laboratory-wide shared mass spectral library in a LAN environment, and an internet mass spectral database of a special interest group. Currently we provide both of the Linux and the Windows installers of MassBank system.

The current distribution of MassBank system includes not only MassBank server but also a database maintenance subsystem. The database maintenance subsystem consists of MassBank Server Administration Tool and MassBank Record Management Tool. MassBank Server Administration Tool is a Web-based tool for registration and removal of MassBank records. MassBank Record Management Tool is an MS-Excel based tool for creating MassBank records. MassBank record includes the chemical information of the measured compound such as formula, exact mass, SMILES code and InChI code. These information is automatically generated from a user specified molfile and inserted into a MassBank record by MassBank Record Management Tool.

A pipeline of spectra from experiment to database can be realized by the combination of MassBank database maintenance subsystem and an existing mass spectral analysis tool such as Mass++ which can extract measured data and analytical conditions from a mass spectrometry equipment and generate a peak list.

References