

Development of a client tool for a mass spectra database

Satoshi Tanaka¹, Shigeki Kajihara¹, Shinichi Utsunomiya¹, Tsuyoshi Tabata², Ken Aoshima², Yoshiya Oda², Yoshito Nihei³, Takaaki Nishioka³, Koichi Tanaka¹

¹SHIMADZU CORPORATION, Nishinokyo-kuwabaracho, Nakagyo-ku, Kyoto 604-8511, Japan ²Eisai Co., Ltd., Tokodai 5-1-3, Tsukuba, 300-2635, Japan ³BIRD, JST and Keio University, Baba-cho 14-1, Tsuruoka, 997-0035, Japan



1: Introduction

Commercial software attached to an MS instrument is usually used to deal with MS raw data after acquisition by a mass spectrometer. However, it has some of the following problems:

- It cannot read file formats from other mass spectrometers. So the same analysis methods might not be able to be used with different file formats.
- It must be used with a valid license. So when dealing with a huge amount of data, multiple tasks cannot be executed without extra licenses.
- It cannot be controlled by third-party software. Therefore analysis using original methods or algorithms is not easy.

To solve these problems, we have developed a new software product named Mass++.

MassBank is a public repository database of mass spectra, currently containing 30,857 data. To search the data on a large scale, users had to manually repeat extracting peak data from raw data and submitting them to MassBank. To relieve researchers from these tedious manual tasks, we cooperated to add new functions to Mass++ and MassBank.

2: Mass++

Mass++ is software for viewing and manipulating mass spectra, which supports various data formats as listed below.

Table 1 Supported data formats

Software	Company	extension	Input / Output
LCMS solution	Shimadzu	.lcl	Input
GCMS solution	Shimadzu	.qgd	Input
Analyst, Analyst QS	Applied Biosystems	.wiff	Input
Xcalibur	Thermo Fisher	.raw	Input
MassLynx	Waters	.raw	Input
mzXML / mzML		.mzXML .mzML	Input / Output
MSB	(Mass++ Original Format)	.msb	Input / Output

Mass++ is a plug-in style software application. So users can customize it depending on their purposes such as addition of new functions and deleting unnecessary functions to increase performance, without editing Mass++ source code.



Figure 1 plug-in

Mass++ supports new plug-ins written in the C/C++, C++/CLI, C#.NET or VB.NET programming languages, to expand its functionality.

Also, Mass++ has many functions implemented as plug-ins such as Profile View, 3D View, Overlapping, Peak Detection, Smoothing, Baseline Subtraction, Background Subtraction, Data Fusion, Quantitation and so on.

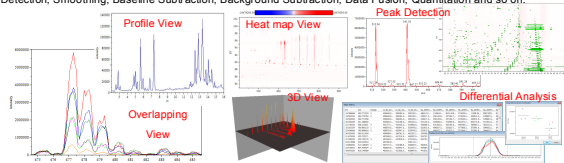


Figure 2 Mass++ Functions

Mass++ can be freely downloaded from the web site: <http://masspp.jp/>

3: MassBank^[1]

MassBank currently contains 30,857 mass spectral data provided by twenty laboratories. In addition, MassBank provides functions for database search via a web application. So anyone can use them freely with a web browser such as Internet Explorer or Firefox.

Table 2 Database services of MassBank

Service	Details
Spectrum Search	Search similar spectra on a peak-by-peak basis
Quick Search	Keyword search of chemical compounds
Peak Search	Search spectra by m/z values and molecular formulae
Substructure Search	Search chemical compounds by substructures
Advanced Search	Search similar spectra on a neutral loss-to-neutral loss basis
Spectral Browser	3D viewer of user's spectra
Batch Service	Similarity search of MSn spectra in a batch process
Browse Page	Hierarchical browsing of all data
Record Index	Categorized list of spectra

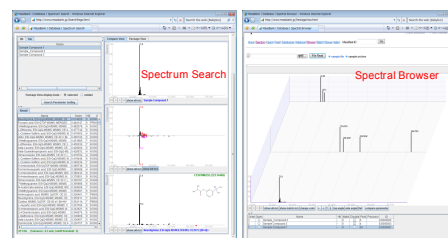


Figure 3 Database services of MassBank

Contributors to MassBank provide PCs as their own data servers for publishing their data. The MassBank system and installer for Windows and Linux is available as open source software. Additionally the system is useful for building personal or group mass spectra databases in the laboratory.

MassBank is available on the website: <http://www.massbank.jp/>

4: Linking Mass++ and MassBank

Mass++ is plug-in style software, so functions related to MassBank are also implemented as a plug-in. MassBank provides a SOAP (Simple Object Access Protocol) API (Application Programming Interface) as well as a web application, which enables applications to be written without requiring a web browser. Mass++ can perform a search in MassBank through its SOAP API.

4-1: Database Search

Mass++ supports some database search functions of MassBank.

Table 3 Database Search Services of MassBank supported in Mass++

Database Search	Details
Spectrum Search	Search similar spectra on a peak-by-peak basis
Peak Search	Search spectra by m/z values
Peak Difference Search	Search spectra by m/z differences
Batch Search	Search similar spectra in a batch process. (This function will be supported in the next version of Mass++)

Previously users had to extract peak data in text format and paste it into the browser because raw data is not accepted as a query for spectral search in MassBank. Using new Mass++, users can search MassBank data by simply selecting a raw data spectrum for a MassBank query.

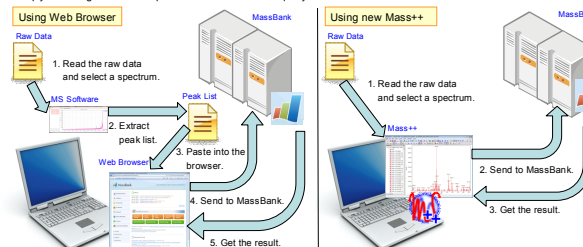


Figure 4 Data flow in MassBank search

After searching, the search results are displayed and peaks in MassBank can be overlaid onto the waveform of spectra displayed by Mass++ to confirm how similar they are.

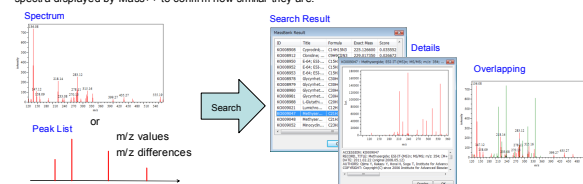


Figure 5 MassBank search results shown in Mass++

4-2: MassBank Record

To register MS spectra in MassBank, text files called "MassBank records" are required. Contributors usually manually prepare MassBank records from their raw data analyzed on various kinds of MS instruments. It was very time-consuming to extract sample information, set-up parameters and peak data. Furthermore, the way of extracting them is different according to the software. On the other hand, Mass++ can export MassBank records semi-automatically by extracting information from raw data and detecting peaks. These records can then be easily registered in MassBank.

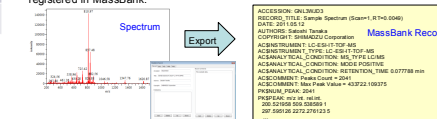


Figure 6 Creating a MassBank record using Mass++

5: Summary

Mass++ is plug-in style all-purpose software for Mass Spectrometry.

MassBank is a powerful database for MS spectra.

Mass++ and MassBank make it easy to search similar spectra with a large query dataset.

-Future Plans

- Higher performance/accuracy peak detection algorithm

- A more user-friendly user interface

- Development of MassBank REST API

Mass++ ver. 2.0.0.

The latest version of Mass++ is 1.7.4.

We are preparing to distribute Mass++ 2.0.0.

It will be released this September. Please check the Mass++ website.

-Reference

[1] MassBank: A public repository for sharing mass spectral data for life sciences

H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Yokota Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsura, T. Soga, R. Taguchi, K. Saito and T. Nishioka, J. Mass Spectrom., 45, 703-714(2010)

-Acknowledgment

This research is granted by the Japan Society for the Promotion of Science (JSPS) through the "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program)," initiated by the Council for Science and Technology Policy (CSTP).