

MassBank : Mass Spectral Database for Metabolome Analysis

Hisayuki Horai^{1,2}, Masanori Arita^{1,2,3,4}, and Takaaki Nishioka^{1,2}

¹ Institute for Advanced Biosciences, Keio University

² JST-BIRD

³ Graduate School of Frontier Sciences, University of Tokyo

⁴ RIKEN Plant Science Center

Database of comprehensive, high-resolution mass spectra of metabolites

URL: [//www.massbank.jp/](http://www.massbank.jp/)

e-mail: massbank@iab.keio.ac.jp

Overview

- Database of comprehensive mass spectra of metabolites.
- Freely accessible public distributed database via internet.
- Includes high-resolution ESI-MS and (MS)ⁿ.
- Three dimensional comparison of spectra.
- Batch service of database search.

Introduction

Mass spectra of metabolites are analyzed on different types of mass analyzers with various ionization methods. This makes difficult to share and exchange mass spectral data.

In 2006, MassBank started to collect mass spectra of metabolites. MassBank defines a standard record format to describe a set of detail information on mass spectral experiments that is along the guidelines and standards for sharing and reporting experimental data proposed by working groups of Human Proteome Organization (HUPO) and Metabolomics Society.

Data and services of MassBank have grew up rapidly. We report the current status of MassBank, especially the spectral data added and search functions developed newly in a year.

Methods

MassBank is a distributed database in the internet so that many researchers can contribute their spectra on their own servers easily (Figure 1). It is expected to enlarge the number of participants and spectra and to collect mass spectra of metabolites comprehensively. From the beginning, MassBank supplies spectral search web services based on the similarity in m/z values and on the difference between a pair of m/z values.

To compare a query and search results easily, we introduced a new view which can align them graphically in three dimensions.

Furthermore, to search the database with many queries at one time, we started a batch service which accepts a file containing mass spectra and sends all results to the user by e-mail.

Results

Growth of spectral data: All data are based on the facts in the last ASMS annual meeting (on May 21st, 2007) and on April 30th, 2008 (Table 1). The number of spectra increased to 13,563 from 11,182. Concerning MS methodologies, spectra measured in IT-(MS)ⁿ, LC/QqTOF-MS/MS, QqIT-MS/MS and FAB-MS were newly added. The number of metabolites grew up from 1,428 to 1,851. Especially, mass spectra of 106 carotenoids and 42 lipids were new metabolite classes added. The number of contributors grew up from 3 to 6 laboratories, and 3 others are preparing to contribute their data. The servers are installed in laboratories not only for the public service but also for their internal use.

Three-dimensional view (3D view): It is a newly developed display function which is useful to compare more than four mass spectra in a usual computer display (Figure 2). Its functions include changing the angle of sight and the magnification smoothly, displaying the coexistence of peaks among spectra graphically, and searching the database with the peaks selected on the view as a query. The 3D view show the mass spectra of a query as well as those of the search results and also those in a local file.

Batch service: The system delegates every mass spectrum to all distributed servers concurrently, and sends an e-mail including the top 20 lists for all query mass spectra in a text format and a HTML format with the anchors to the database records (Figure 3). The performance depends upon factors including the number of peaks in the query and the number of the query files processed concurrently. In case that only one query file is processed and the file includes 100 mass spectra, each of which has approximately 15 peaks on average, the system takes about 5 minutes from receiving the file to sending the results.

Table 1: Contents Overview

Contributor	Method	Spectra	Compounds
Keio IAB	QqTOF-MS/MS	4,510	695
	QqQ-MS/MS	4,275	
	IT-MS/MS	515	
Riken PSC	GC/TOF-MS	241	341
	LC/TOF-MS	85	
	LC/QqQ-MS/MS	87	
	CE/TOF-MS	20	
	QTOF-MS/MS	261	
Nippon Waters K.K.	LC/Q-MS	2721	577
	QqQ-MS	273	
Kyoto U.	FAB-MS/MS	106	106
U. of Tokyo	QqIT-MS/MS	378	42
Kazusa DNA Res.Inst.	GC/TOF-MS	91	90
Leibniz IPB		(to be appeared)	
Tottori U.		(to be appeared)	
Fukuyama U.		(to be appeared)	

(30 April 2008)

Figure 1: Overview of MassBank Distributed Database

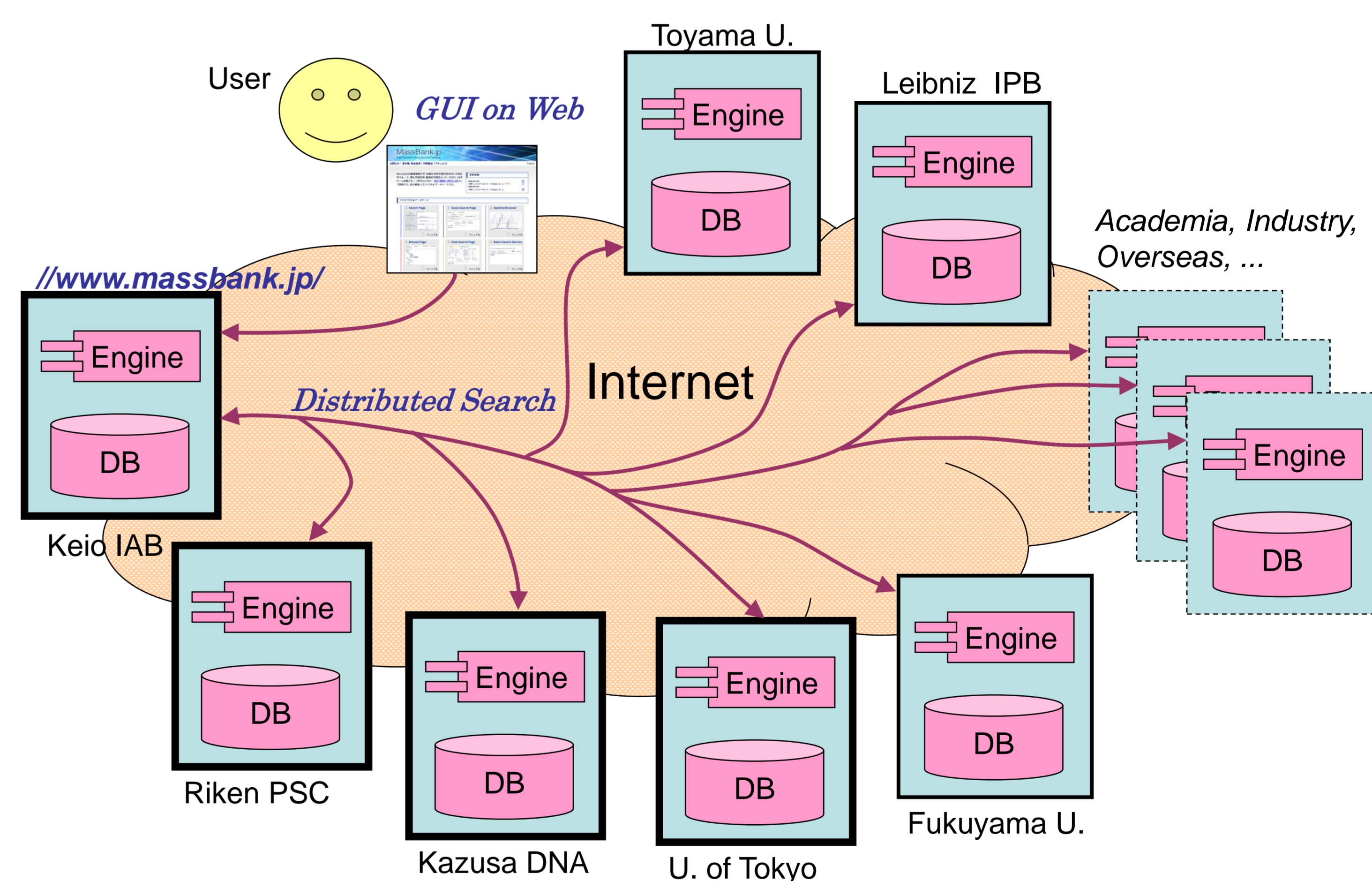


Figure 2: 3D view of Spectral Search in MassBank

Figure 3: Batch Service

Batch Search Service

Home | Search | Browse | Peak Search | Quick Search | Spectral Browser | Batch | Record List | MassBank Record No: Go

This service will appear as a part of Quick Search Page.

Query File: **File name input field**

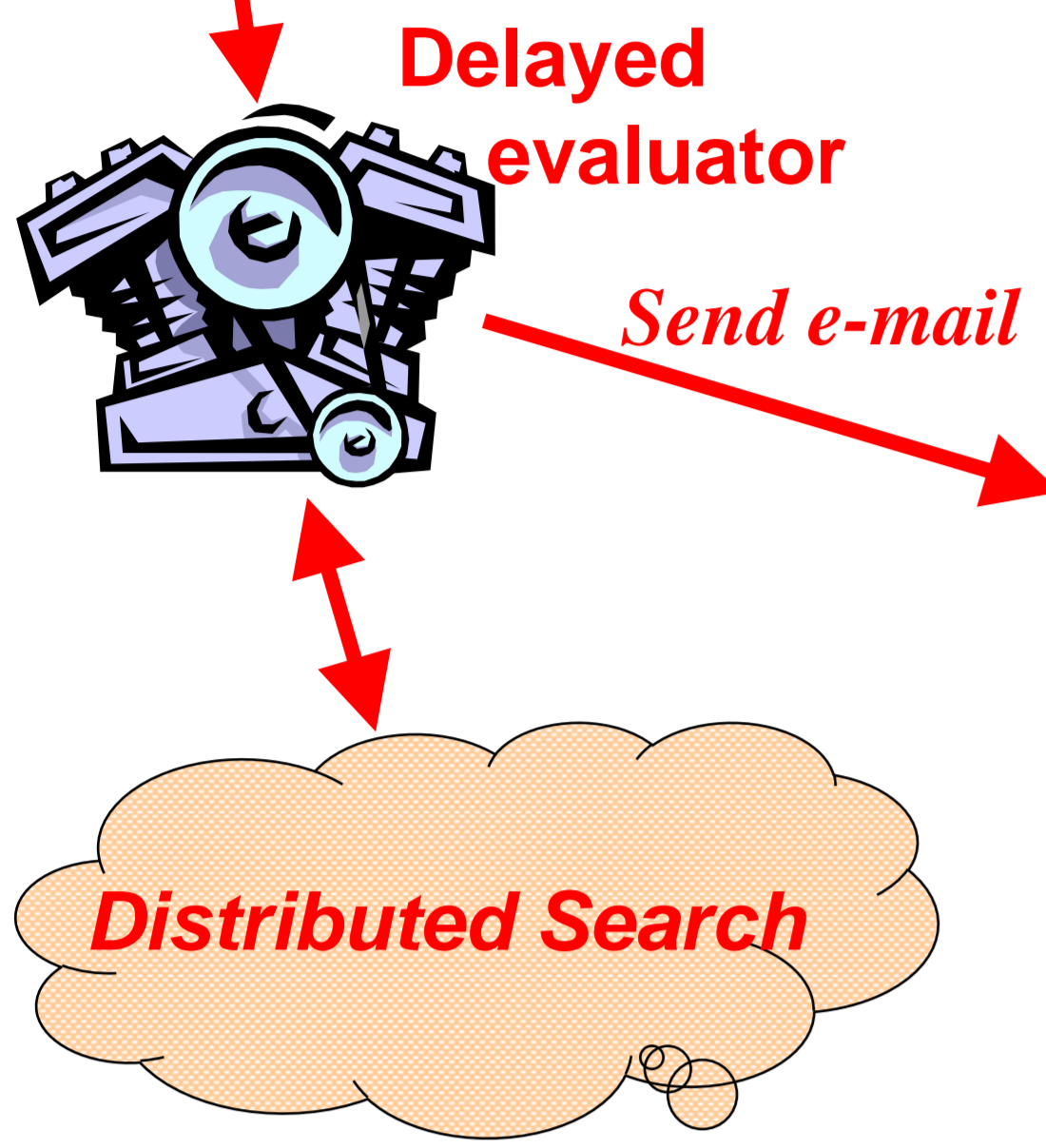
Mail Address: **E-mail address input field**

Immediate acceptance notice in browser

Batch Search Service

Home | Search | Browse | Peak Search | Quick Search | Spectral

[2008/05/20 13:09:38]
Your batch search is accepted.
The results will be sent to horai@ttck.keio.ac.jp later.



E-mail client software

受信トレイ

送信者: massbank@iab.keio.ac.jp 宛先: horai@ttck.keio.ac.jp
件名: MassBank Batch Search Results

Dear Users,

Thank you for using MassBank Batch Service.

The results for your request dated '2008/05/20 13:09:38 JST' are attached to this e-mail.

MassBank.jp - High Resolution Mass Spectral Database
URL: <http://www.massbank.jp>
E-mail: massbank@iab.keio.ac.jp

Attachments: MassBankResults.txt (117 KB), MassBankResults.htm (340 KB)

Flat text & HTML

Attached HTML document

MassBank Batch Search Results

Request Date : 2008/05/20 13:09:38 JST

Query 1 **Top 20 list for each query**

Name: Compound 1
Hit: 3855
Top 20 List

Accession	Val	MS/MS	QqTOF	CE40 V	[M+H] ⁺	Formula	Ion	Score	Hit
KO005733	Val	MS/MS	QqTOF	CE40 V	[M+H] ⁺	C6H11NO2	[P]	0.8262	35
KO006722	2-Amino-2-methylbutanoate	MS/MS	QqTOF	CE40 V	[M+H] ⁺	C6H11NO2	[P]	0.7819	34
KO004255	Val	MS/MS	QqQ	CE50 V	[M+H] ⁺	C6H11NO2	[P]	0.7330	23
KO006723	2-Amino-2-methylbutanoate	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C6H11NO2	[P]	0.7180	21
KO006662	5-Aminopentanoate	MS/MS	QqTOF	CE40 V	[M+H] ⁺	C6H11NO2	[P]	0.7091	33
KO008590	Thr	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C4H9NO3	[P]	0.6917	25
KO007816	l-Methylserine	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C4H9NO3	[P]	0.6898	19
KO002392	2-Amino-2-methylbutanoate	MS/MS	QqQ	CE50 V	[M+H] ⁺	C6H11NO2	[P]	0.6820	14
KO006732	2-Amino-2-hydroxymethyl-1,3-propanediol	MS/MS	QqTOF	CE40 V	[M+H] ⁺	C4H11NO3	[P]	0.6785	25
KO006663	5-Aminopentanoate	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C6H11NO2	[P]	0.6774	27
KO006733	2-Amino-2-hydroxymethyl-1,3-propanediol	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C4H11NO3	[P]	0.6765	24
KO007286	l-Leucine	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C6H13NO2	[P]	0.6718	19
KO007616	lle	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C6H13NO2	[P]	0.6705	26
KO008606	allo-threonine	MS/MS	QqTOF	CE50 V	[M+H] ⁺	C4H9NO3	[P]	0.6641	22
KO004254	Val	MS/MS	QqQ	CE40 V	[M+H] ⁺	C6H11NO2	[P]	0.6629	21
KO002311	5-Aminopentanoate	MS/MS	QqQ	CE40 V	[M+H] ⁺	C6H11NO2	[P]	0.6562	23
KO008589	Thr	MS/MS	QqTOF	CE40 V	[M+H] ⁺	C4H9NO3	[P]	0.6556	24
KO006661	5-Aminopentanoate	MS/MS	QqTOF	CE30 V	[M+H] ⁺	C6H11NO2	[P]	0.6541	28
KO009126	lle	MS/MS	QqQ	CE50 V	[M+H] ⁺	C6H13NO2	[P]	0.6517	20

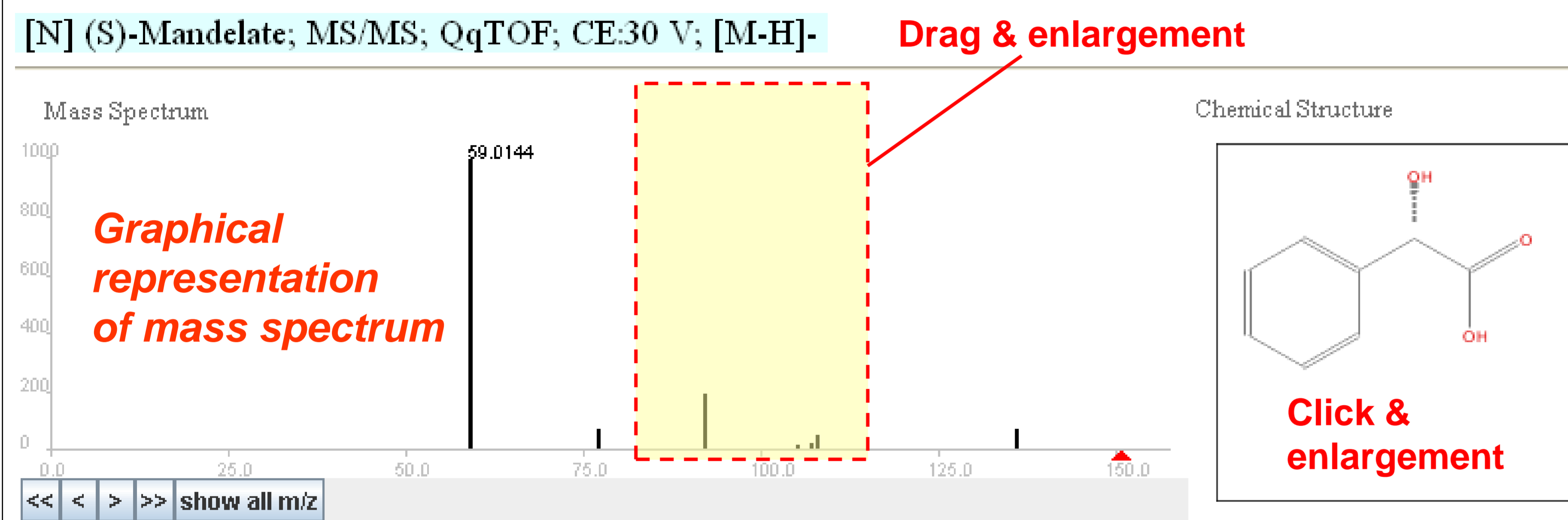
Query 2

Open

Link to MassBank Record

Example of MassBank Database Record

MassBank Record: KO005733



We would appreciate it very much if you could contribute spectra of metabolites and natural products to www.massbank.jp.

MassBank Database Record

Each MassBank database record corresponds to one-to-one with a spectrum, and the version of a record is managed by the update date (DATE). Each database record includes the following information: chemical information of compound (CH\$), experimental conditions (AC\$), those of the mass spectrum measurement, such as its precursor ion and base peak (MS\$), and spectral data (PK\$).

As much experimental condition as the contributor discloses are described in detail. In case of chromatographic separation on LC, GC and CE, experimental conditions of separation are also described.

In case of (MS)ⁿ, the sequence of precursor ions which appear in each generation of MS is described in the field of MS\$FOCUSED_ION.

The fundamental data of the mass spectra follow. They consist of the triplet of *m/z*, actual measured intensity and relative intensity of each peak, which is described in the field of PK\$PEAK. The chemical formula of product ions is predicted and described optionally in the field of PK\$ANNOTATION. Currently, it is predicted only to the spectra measured on high-resolution MS. Predicted chemical formulae are assigned to the peaks with in 50 ppm error from the measured *m/z*. For instance, C₇H₇O (exact mass: 107.04969) is the only one candidate for *m/z* 107.0500 of C₈H₈O₃. The error of exact mass is 3 ppm, and the unsaturation is 4.5. Feature studies include more precise prediction and extend applicable compounds.

The display of a database record consists of the graphical representation of the mass spectrum and the text representation of the record. The graphical mass spectrum can be enlarged. Metabolites are linked to other database on the internet by their IDs described in the field of CH\$LINK.

Chemical structure of the compound is graphically displayed. Substructure search service will be opened soon.

Research Collaboration

- Keio U. IAB, U.Tokyo Grad Sch. Frontier Sci., NAIST Grad.Sch.Info.Sci., Riken PSC, U.Tokyo Grad.Sch.Med., Nara W.U. Fac.Sci, Kazusa DNA Res. Inst., Leibniz Inst. Plant Biochem.
- This project is financially supported by JST-BIRD project.

Copyright(C) 2006-2007 Institute for Advanced Biosciences, Keio University

ACCESSION: KO005733
RECORD TITLE: (S)-Mandelate; MS/MS; QqTOF; CE:30 V; [M-H]⁻
DATE: 2007.07.07
AUTHORS: Kakazu Y, Horai H, Institute for Advanced Biosciences, Keio Univ.
COPYRIGHT: Copyright(C) 2006-2007 Institute for Advanced Biosciences, Keio University

ACCESSION: Identifier of the record
RECORD_TITLE: Title
DATE: Date of last update
AUTHORS: Authors of the record

CH\$NAME: (S)-Mandelic acid
CH\$NAME: (S)-2-Hydroxy-2-phenylacetic acid
CH\$NAME: (S)-Mandelate
CH\$NAME: (S)-2-Hydroxy-2-phenylacetate
CH\$FORMULA: C8H8O3
CH\$EXACT_MASS: 152.04734
CH\$SMILES: OC(=O)[C@H](O)c1ccccc1
CH\$IUPAC: 1/C8H8O3/c9-7(8(10)11)6-4-2-1-3-5-6/h1-5,7,9H
CH\$LINK: CAS [90-64-2](#) [611-72-3](#)
CH\$LINK: CHEBI [32800](#)
CH\$LINK: ChEMBL [SMN](#)
CH\$LINK: KEGG [C01984](#)
CH\$LINK: KEIO M057
CH\$LINK: PUBCHEM SID: [5081](#)

Information of Compound
CH\$NAME: Name and synonyms
CH\$FORMULA: Formula
CH\$EXACT_MASS: exact mass
CH\$SMILES: SMILES code
CH\$IUPAC: InChI code
CH\$LINK: ID of other database with link

AC\$INSTRUMENT: Qstar, Applied Biosystems
AC\$ANALYTICAL_CONDITION: MS TYPE MS/MS
AC\$ANALYTICAL_CONDITION: MODE NEGATIVE
AC\$ANALYTICAL_CONDITION: PRECURSOR_SELECTION Q
AC\$ANALYTICAL_CONDITION: FRAGMENTATION_EQUIPMENT Q
AC\$ANALYTICAL_CONDITION: SPECTRUM_TYPE TOF
AC\$ANALYTICAL_CONDITION: COLLISION_ENERGY 30 V

Experimental Condition
AC\$INSTRUMENT: Equipment
AC\$ANALYTICAL_CONDITION: Condition
Mode, collision energy, etc.
Information of separation (retention time, name of column, etc.)

MS\$FOCUSED_ION: PRECURSOR_TYPE [M-H]⁻
MS\$FOCUSED_ION: PRECURSOR_M/Z 151

MS\$FOCUSED_ION; Information of characteristic peaks (precursor ion, base peak, etc.)

PK\$NUM_PEAK: 8
PK\$ANNOTATION_METHOD: Combinatory of compositional formula; threshold = 50 ppm
PK\$ANNOTATION: *m/z* num { formula unsat. mass error(ppm) } * **Number of candidates**

<i>m/z</i>	num	formula	unsat.	mass error(ppm)	Number of candidates
107.0500	1				
C7H7O	4.5	107.04969	3		
105.0385	1				
C7H5O	5.5	105.03404	42		
77.0379	1				
C6H5	4.5	77.03913	-16		
59.0144	1				
C2H3O2	1.5	59.01330	19		
59.0144	650.000	999			
77.0379	47.000	72			
92.0291	125.000	192			
105.0385	10.000	15			
107.0500	14.000	22			
107.0592	10.000	15			
108.0189	33.000	51			
136.0182	47.000	72			

Observed *m/z* (107.0500) **Actual intensity** (14.000) **Relative intensity** (22)

Candidate (C7H7O) **Unsaturation** (4.5) **Error in ppm** (107.04969)