Overview of METLIN and MassBank libraries

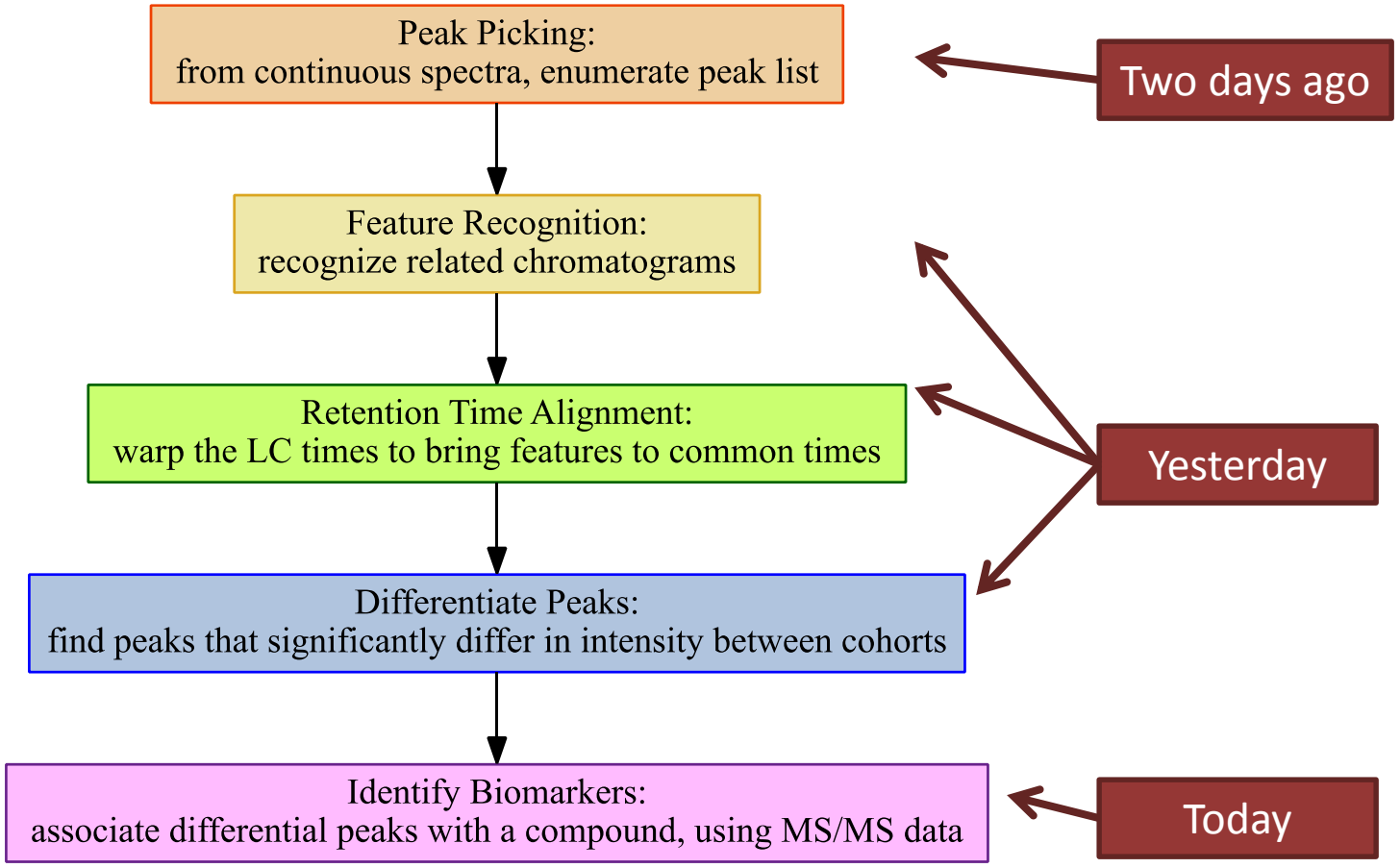
David L. Tabb, Ph.D.

dtabb1973@gmail.com

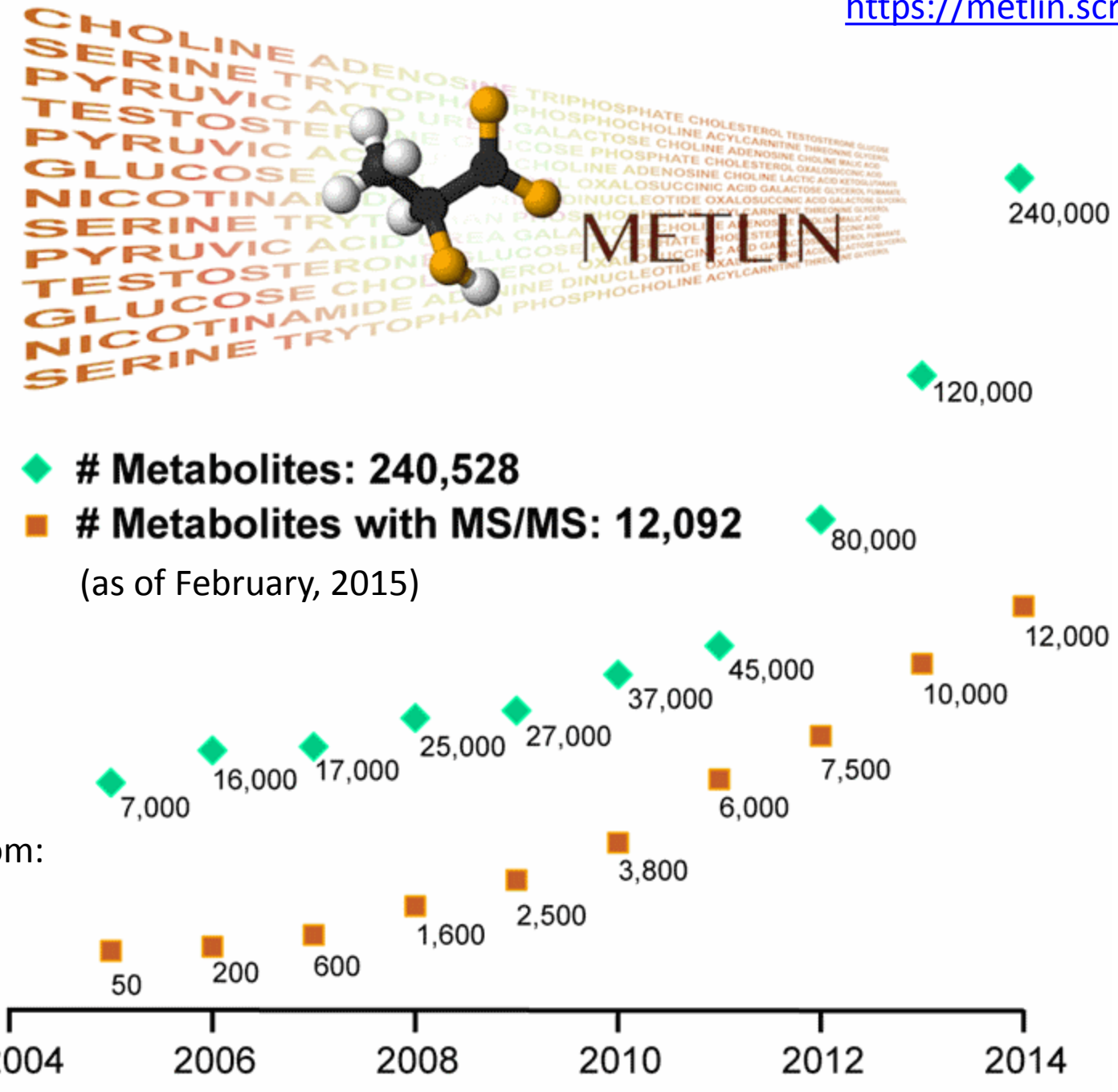
Overview

- Investigating search types available within the METLIN database at Scripps Research Institute
- Examining the international federated database at MassBank

Metabolome informatics



https://metlin.scripps.edu/



Chemicals from:
Agilent
SIGMA
Cayman
Cambridge
Chromadex

METLIN origins

Its name derives from “Metabolite Link,” connecting information from these sources:

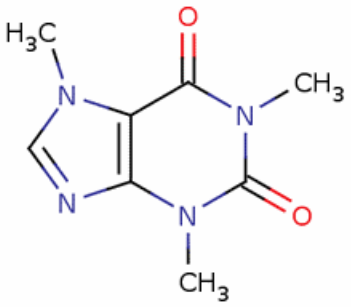
- Structural and physical data
- FTMS data from reference samples
- MS/MS data from reference samples
- LC-MS profiles from human and other samples

Smith et al. *Ther. Drug Monit*. (2005) 6: 747

Where did these spectra come from?

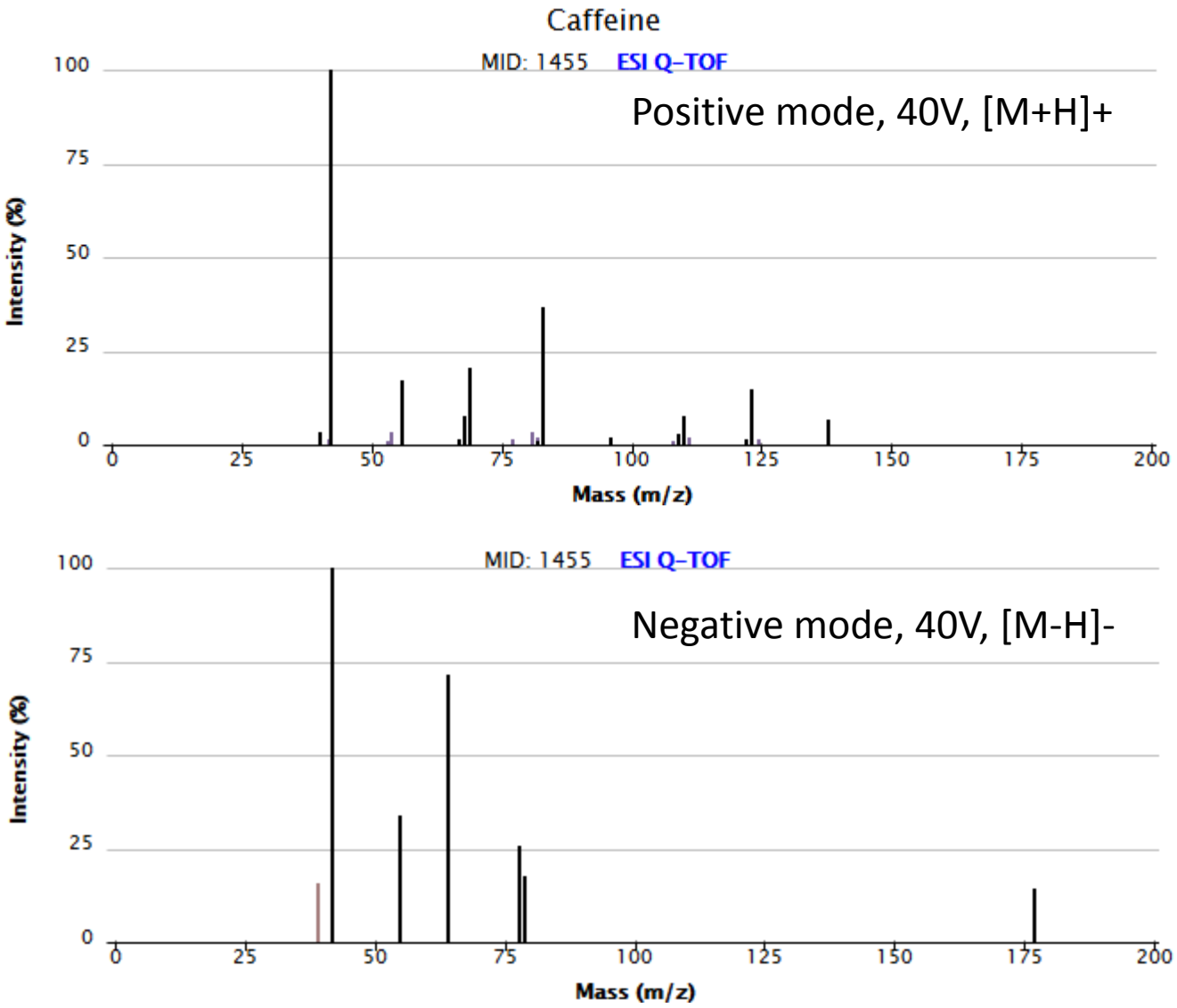
- Siuzdak Lab formed partnerships with chemical companies, who provided reference compounds to his team.
- At Scripps, these samples were subjected to multiple collision energies in both polarities to produce MS/MS scans and high-resolution FTMS scans.
- The library is accessed only via website.

An entry for our friend

MID	1455
Mass	194.080376 m/z calculator
Name	**Caffeine**
Synonym	1,3,7-Trimethylxanthine; 7-methyl Theophylline; Thein; 1-methyl-Theobromine; Methylxanthine theophylline; 1,3,7-Trimethyl-3,7-dihydro-1H-purine-2,6-dione; Lanorinal; 1,3,7-Trimethyl-2,6-dioxopurine; 3,7-Dihydro-1,3,7-trimethyl-1H-purine-2,6-dione; Methyltheobromide; Guaranine; Anhydrous caffeine (JP15); Monohydrate Caffeine
Systematic Name	1,3,7-trimethylpurine-2,6-dione
Formula	$C_8H_{10}N_4O_2$
CAS	58-08-2
Purchase Option	**Sigma-Aldrich: C1778** SIGMA-ALDRICH **Chromadex: ASB-00003032-010** ChromaDex
LMID	
KEGG	**C07481**
HMDB	**HMDB01847**
PubChem	**2519**
Notes	naturally occurring alkaloid Dollery, Colin Therapeutic Drugs, 2nd Ed. 1999 p. C4
Updated	2014-07-29 15:58:37
Drug	N
Structure	 Structure View

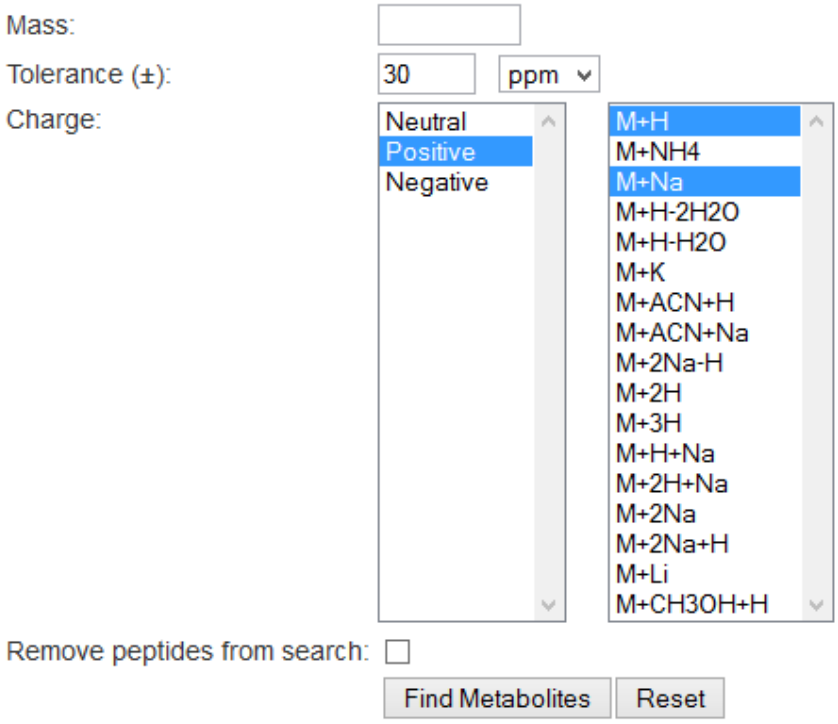
Visualizing MS/MS

In both positive and negative ion mode, energies from 0 to 40V are available for display.

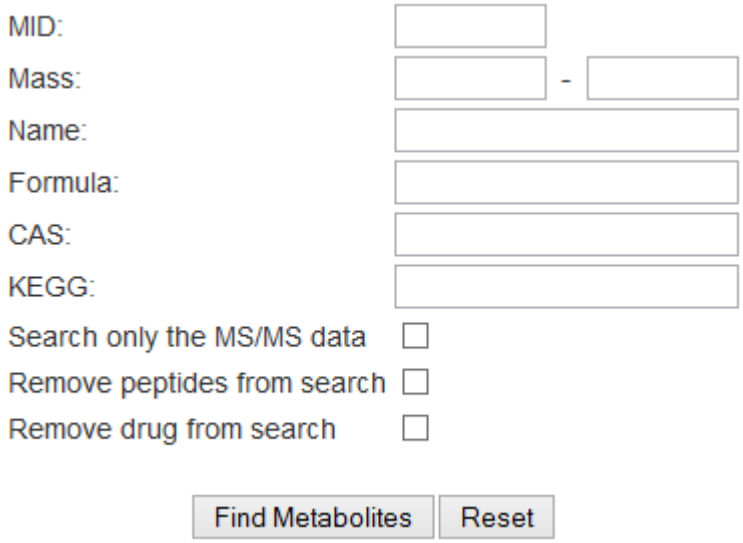


Search possibilities

Simple

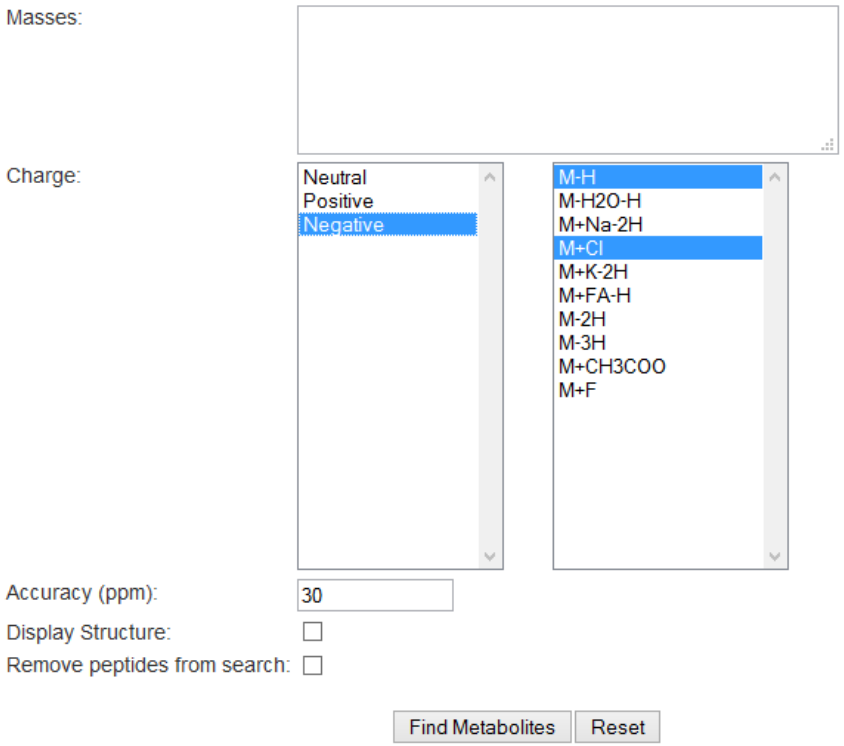


Advanced

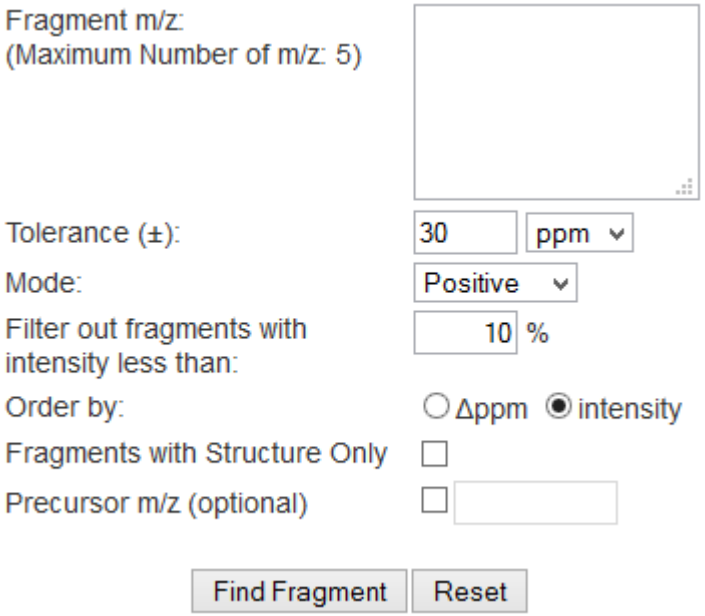


Working from your data

Batch



Fragment



Highly specific matches by MS/MS

MS/MS Spectrum Match

m/z, intensity

Peaks:
(MAX: 30 peaks)

```
177.040, 956
79.029, 569
78.022, 520
178.047, 441
121.028, 308
64.007, 128
```

Mode:

Positive	M-H
Negative	M-H2O-H
	M+Na-2H
	M+Cl
	M+K-2H
	M+FA-H
	M-2H
	M-3H
	M+CH3COO
	M+F

*To select multiple Adducts:
- Hit Ctrl + Adducts
- Hit Command + Adducts

Collision Energy (eV): 20eV

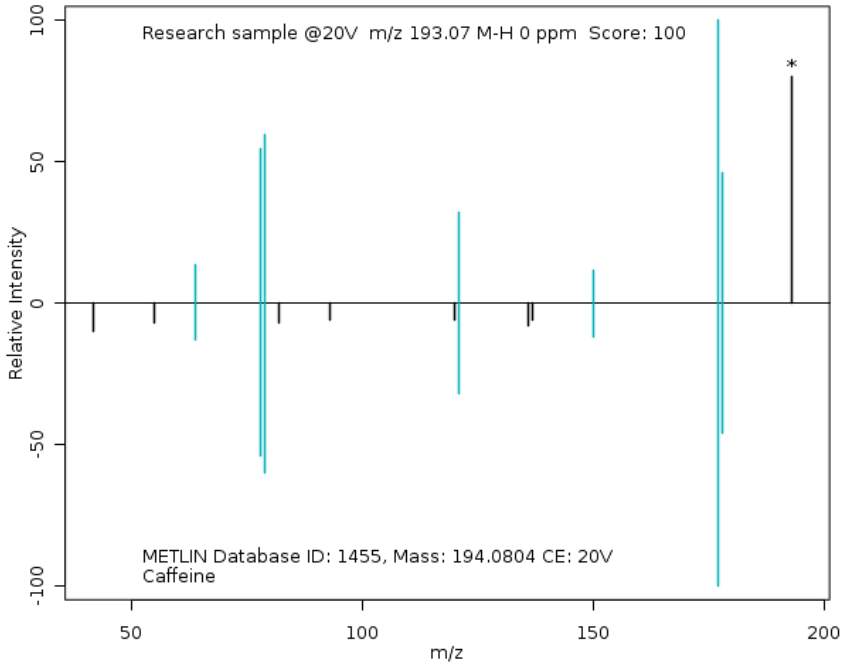
Tolerance MS/MS (Da): 0.01

Tolerance Precursor (ppm): 20

Precursor m/z: 193.0731

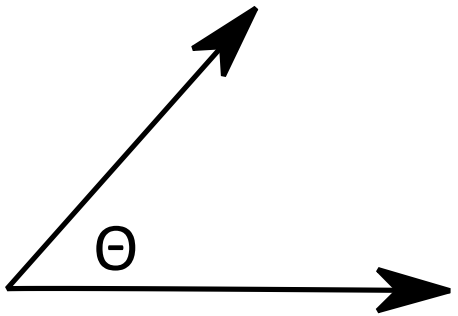
Find Metabolites | Reset

Response to query

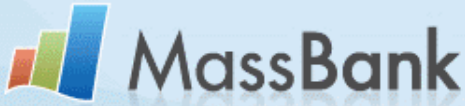


Dot products: MS/MS comparison

- Computes a similarity score between spectra:
  - Each spectrum becomes a vector in *n* dimensions, where n is the number of m/z values included. The intensity values are the coordinates.
  - Comparing two spectra computes the cosine of angle between the two corresponding vectors.
- To emphasize minor peaks, many systems use square root of intensity instead.



Stein and Scott. *J. Amer. Soc. Mass Spectrom*. (1994) 5:859



High Quality Mass Spectral Database

- Database Service
- Statistics
- Publications
- Download
- Manuals
- About MassBank
- Contact
- Consortium Members

- Site Map
- Use Restrictions
- Copyright
- Japanese Site

News

Feb 23, 2015 Java applets were updated with Code Signing. new
Feb 05, 2015 For updating CodeSigning, MassBank Java applets are unavailable few days. new
Dec 22, 2014 MassBank service will stop 0:00 – 7:00, 27 Dec (UTC) for electricity maintenance. new
Oct 16, 2014 MassBank service will stop 9:00, 17 Oct - 0:00, 20 Oct (GMT) for maintenance.
Oct 03, 2014 MassBank service will stop on Oct. 18-19 (UTC+9) for electricity maintenance.

All news

<<About the security checks of Java blocked MassBank applets>> fixed

To all the MassBank users who installed Java 7 Update 51, we apologize for their inconvenience that mass spectra are not displayed because the security checks of Java blocked MassBank applets.
See more details

Database Service

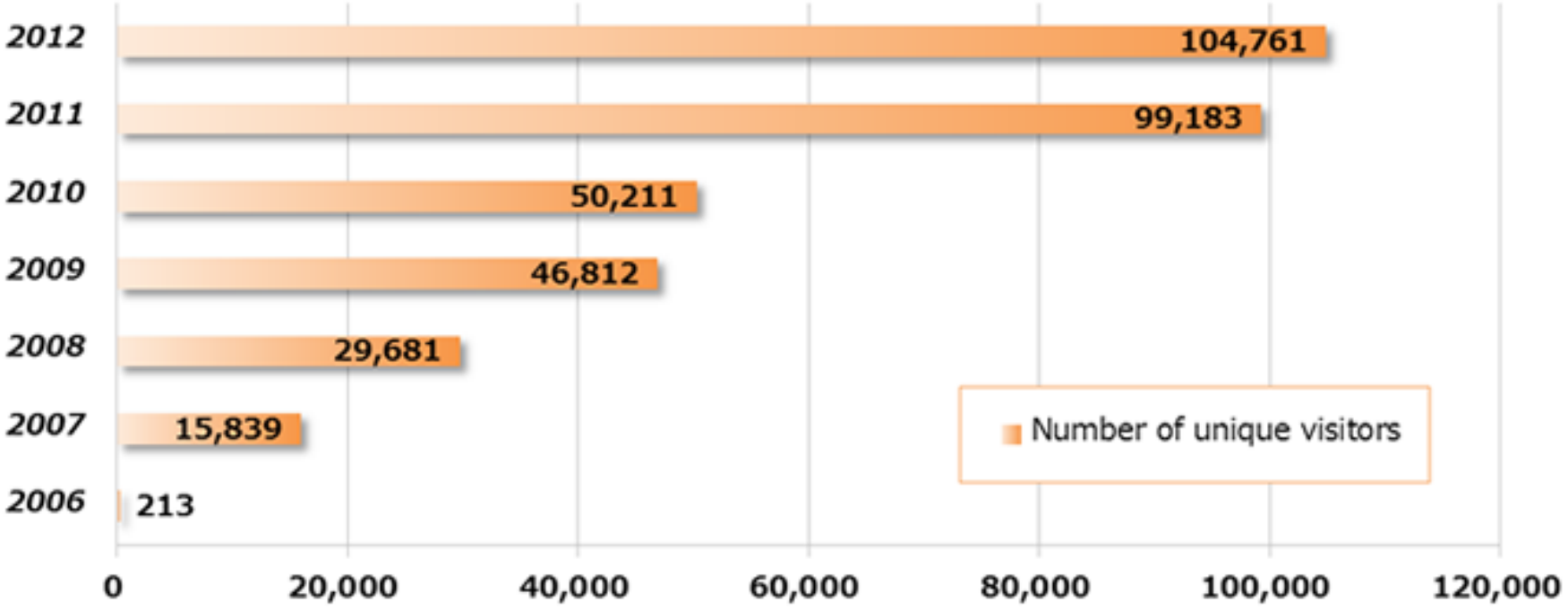
Spectrum Search | Quick Search | Peak Search | Substructure Search | Metabolite Prediction

Spectral Browser | Batch Service | Browse Page | Record Index

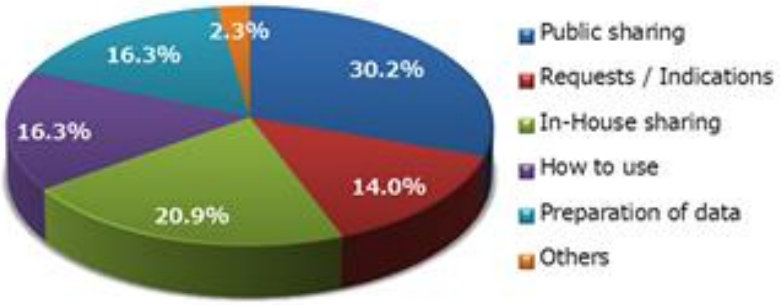
MassBank is financially suported from National Bioscience Database Center, Japan Science and Technology Agency (2011-2013).
The Mass Spectorometry Society of Japan officially supports MassBank.
Please cite the article (DOI) when using MassBank.

http://www.massbank.jp/

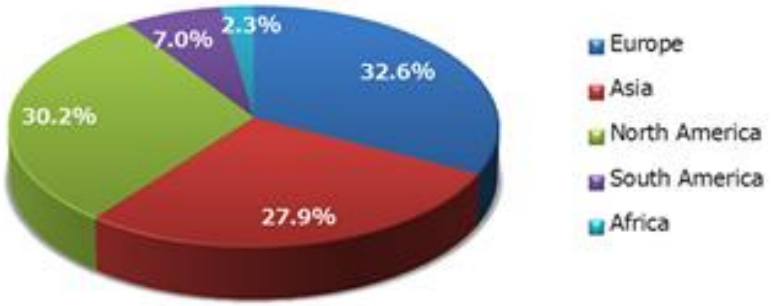
The number of unique visitors to MassBank



The contents of the inquiry



Contact by Region



MassBank origins

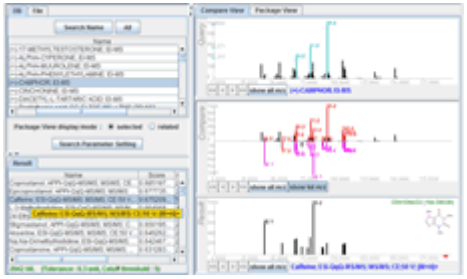
- MassBank is a distributed database, with twenty-eight research groups contributing data through local servers.
- The Institute for Advanced Biosciences led publications, and the School of Engineering at the University of Tokyo has contributed EI spectra for more than 11,000 compounds.
- MassBank collates spectra rather than collecting them.

Horai et al. *J. Mass Spectrom*. (2010) 45: 703

Database services

Spectrum Search

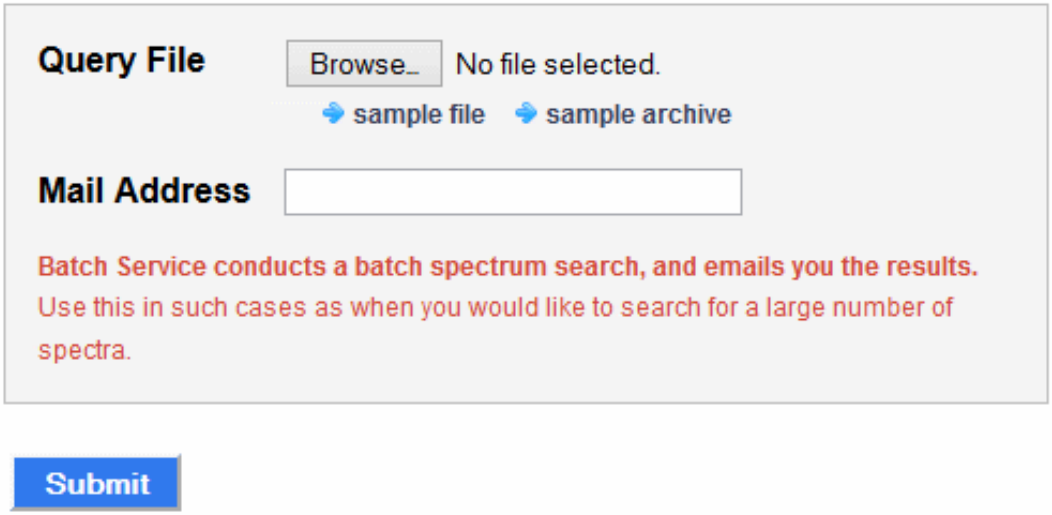
- Retrieves spectra similar to user's spectrum in terms of the m/z value. This search is helpful to identify chemical compound by comparing similar spectra on a 3D-display.

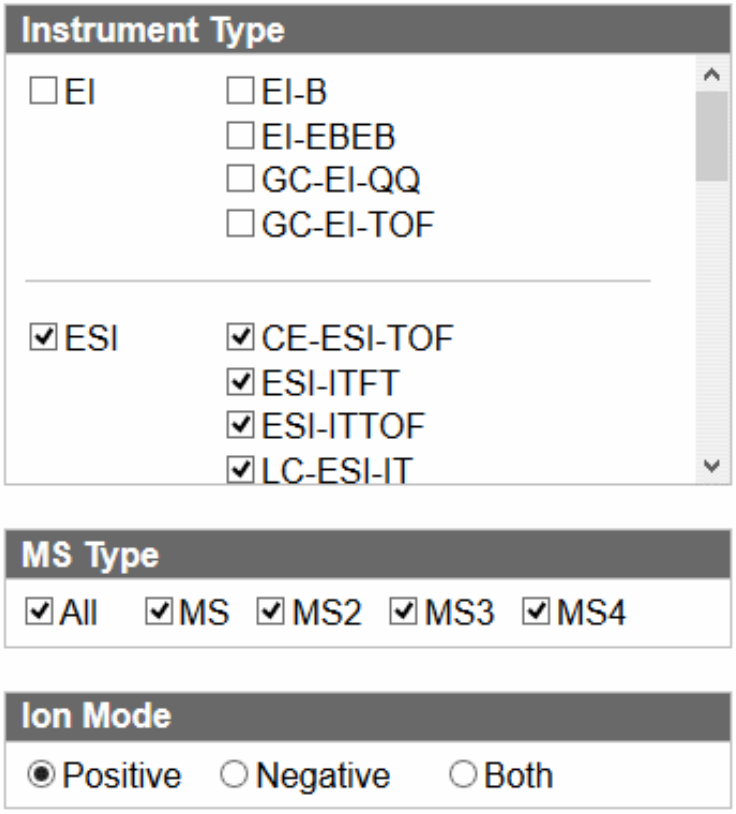


Substructure Search

- Retrieves chemical compounds for which the chemical structure contains the substructures specified by users and displays their spectra.

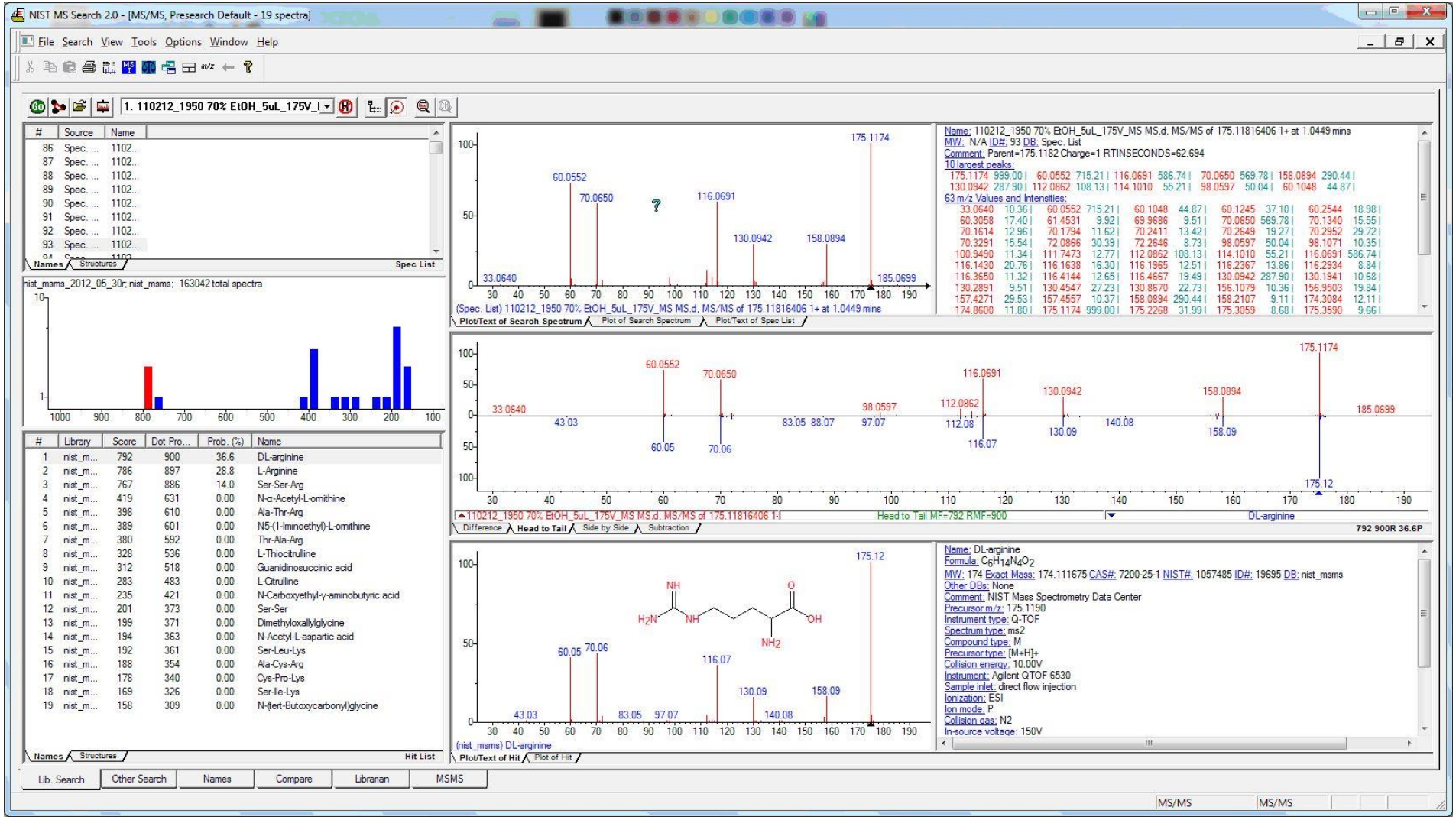
MassBank Batch Service





Text report includes only best 20 hits. HTML report can include larger number of analytes, and it contains links to KEGG for context.

NIST 12 MS/MS Database and Search Program



http://chemdata.nist.gov/mass-spc/msms-search/

Current NIST contents

Database Contents:

6,999	Compounds
15,180	Precursor Ions
121,586	Spectra
~90%	Positive Ion Spectra
~10%	Negative Ion Spectra

Instrument Type	**Precursor Ions**
Ion Trap	12,047
Collision Cell (QTOF and QQQ)	9,232

Takeaway messages

- Repositories for EI and ESI spectra have become a major effort during the last decade.
- If an unknown compound appears in many contexts, its MS/MS may already be available.
- Library matching generally makes use of the dot product strategy.