Interactive Demonstration

NORMAN-MassBank

Emma Schymanski (Eawag) emma.schymanski@eawag.ch

Contents of Demonstration

- NORMAN MassBank
  - Record Structure and the Record Index
  - Quick Search
  - Peak Search
  - Spectrum Search
  - Substructure Search
  - Administration of NORMAN MassBank

- MassBank in NIST

Mass Spectral Database Searching

Some Definitions

- Compound Database (e.g. ChemSpider, PubChem):
  - A collection of structures, basic properties and associated information[1]
  - Generally, no spectral data – but >32,000,000 structures

- Mass Spectral Databases (or Libraries)
  - A collection of structures, their mass spectra and associated information
  - NIST and Wiley Mass Spectral Libraries for GC-EI-MS
    - NIST14: 276,259 spectra; Wiley 10th: >719,000 spectra[2]
  - MS/MS databases are growing; none are yet “established”
    - MassBank: 40,899 spectra (mixed origin);
    - METLIN: 61,872 HR-ESI-MS/MS, Agilent TOF-MS of 12,057 cmpds
    - NIST MS/MS: 193,164 MS/MS spectra of ~8,000 compounds
    - mzCloud.org: 134,029 HR-MS/MS spectra of 1,840 compounds

[1]http://www.chemspider.com/About.aspx. [2]http://www.sisweb.com/software/ms/wiley.htm

Using NORMAN MassBank

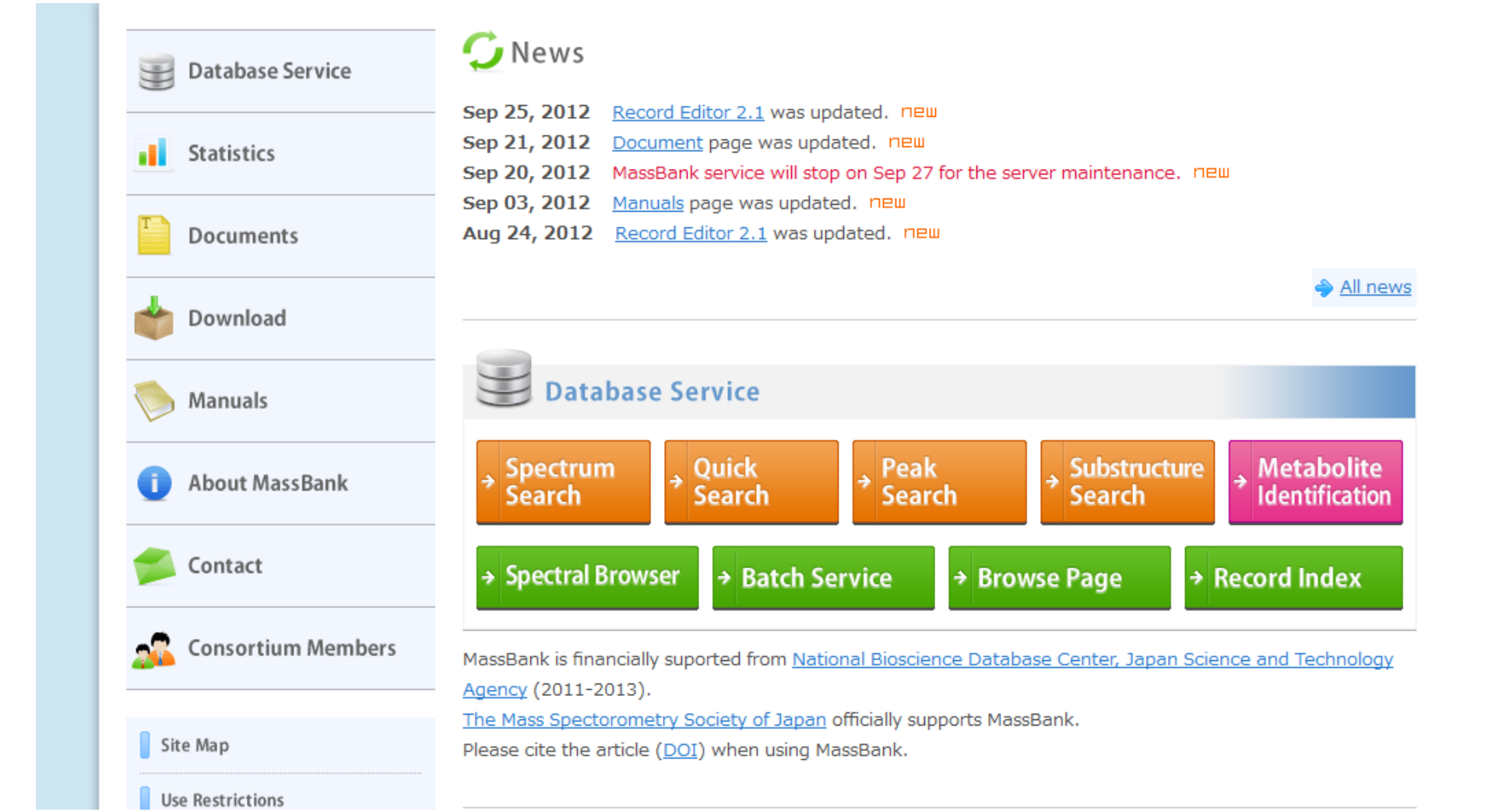
What do you really want to achieve?

- Flexible exchange of data within NORMAN
  - Upload of data from different instruments, different resolutions, …
  - Reduce standard purchase via sharing of data
  - Extension to tentative / unknown spectra

- Purpose of a Mass Spectral Search
  - I have a compound of interest
    - Are there any spectra? What do they look like?
  - I want to identify an unknown
    - Features of MassBank that may be useful
    - Using MetFusion for identification
  - I want to perform routine confirmation and quantification of targets
    - MassBank is not the best stand-alone solution for you (yet?!)
    - *Talk to vendors to value-add instrument software with MassBank*

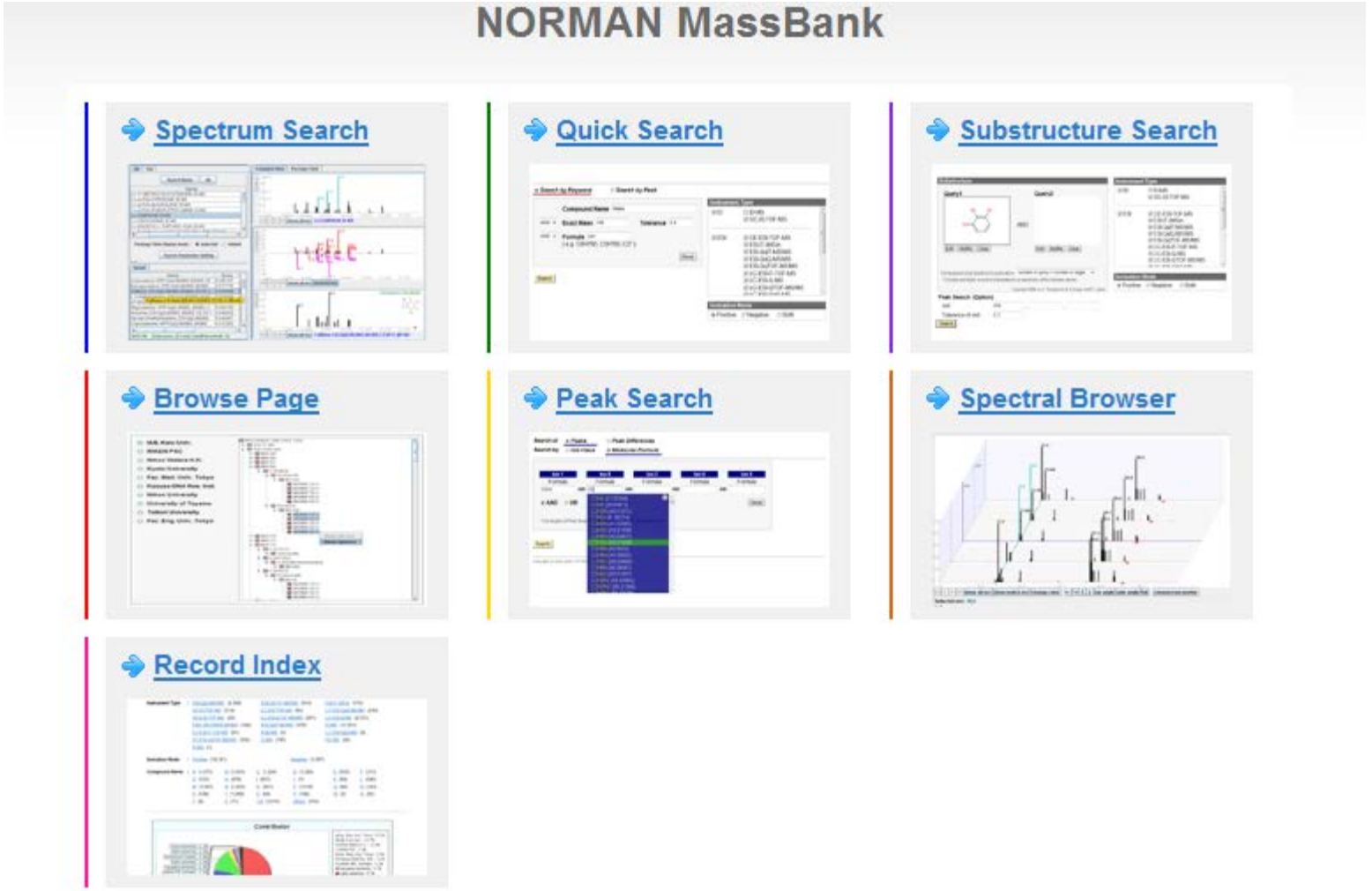
MassBank: www.massbank.jp

Horai et al. 2010: DOI 10.1002/jms.1777



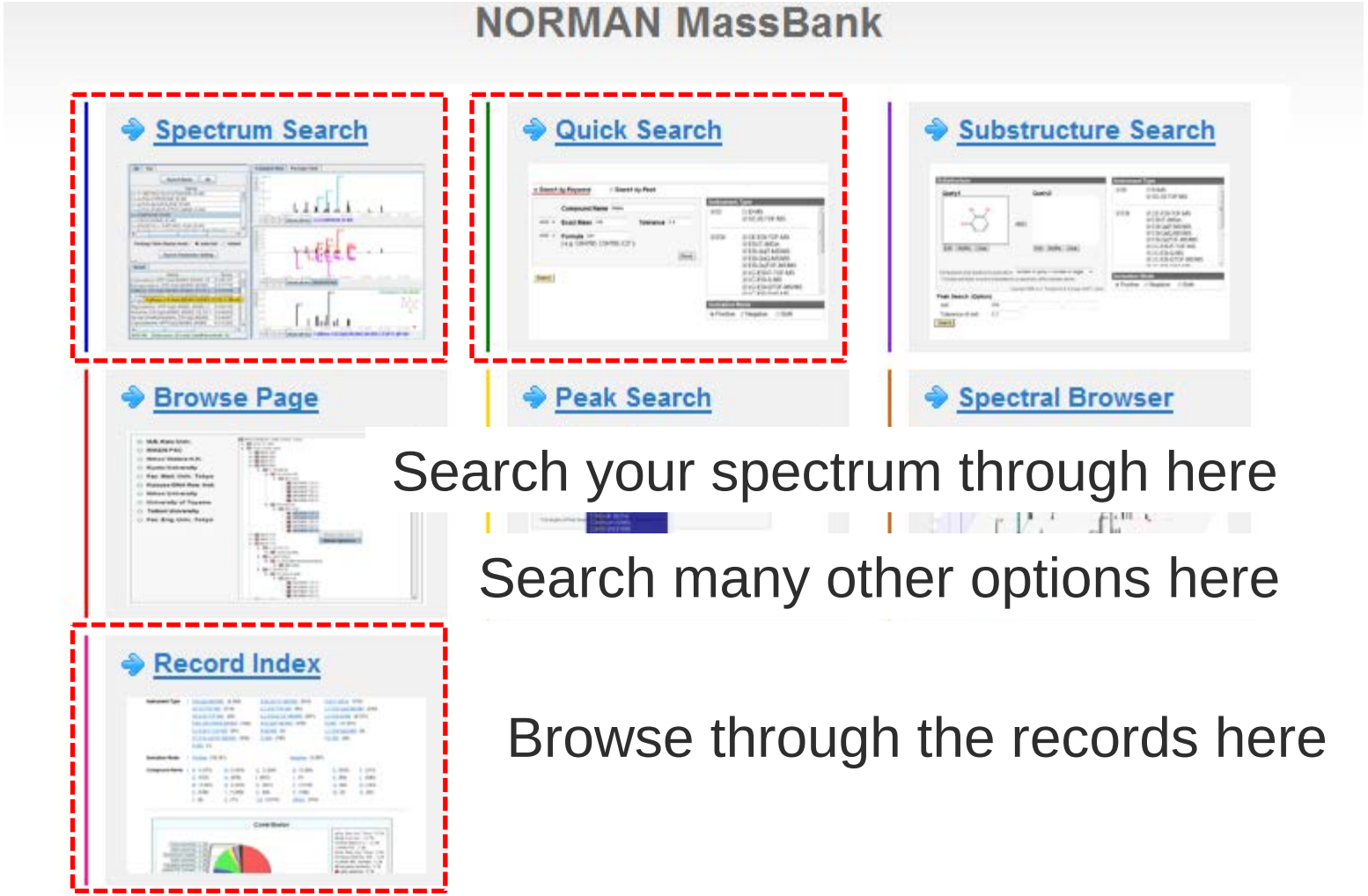
NORMAN MassBank

http://massbank.normandata.eu/MassBank/



NORMAN MassBank

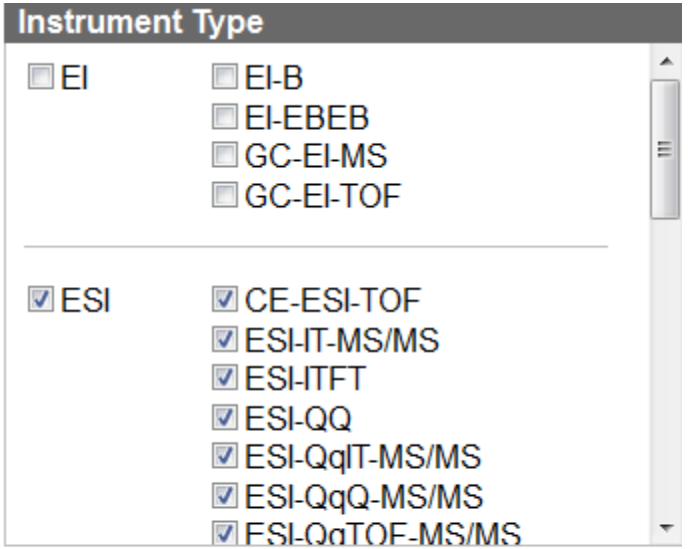
http://massbank.normandata.eu/MassBank/

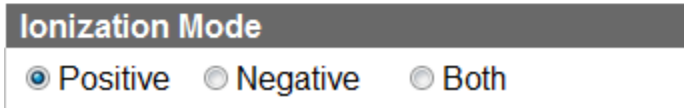


Searching MassBank

What do you want to achieve?

- MassBank is a mixed database
  - Be careful to pick and mix your settings
  - Adjust the thresholds to the data you have
  - …but also to the data in MassBank
  - Tolerance is in Da (i.e. one mass unit)
- Quality of spectra is quite mixed
  - …as is the information provided in them





More details in the next slides!

MassBank Record Format

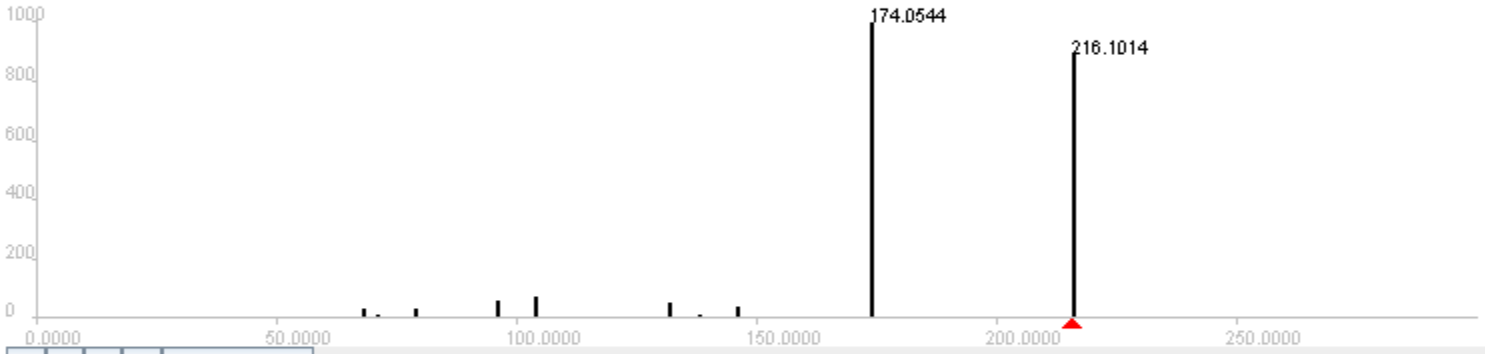
MassBank Record: EA028804

Home | Spectrum | Quick | Peak | Substructure | Prediction | Browser | Batch | Browse | Index

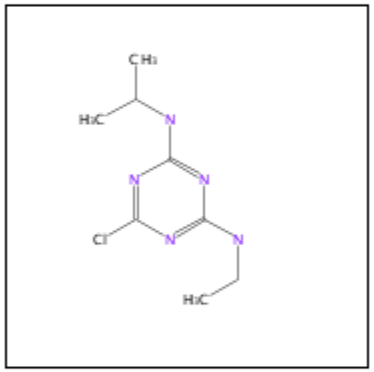
MassBank ID: Go

Atrazine; LC-ESI-ITFT; MS2; 45%; R=7500; [M+H]+

Mass Spectrum



Chemical Structure



```
ACCESSION: EA028804
RECORD_TITLE: Atrazine; LC-ESI-ITFT; MS2; 45%; R=7500; [M+H]+
DATE: 2012.08.02
AUTHORS: Stravs M, Schymanski E, Singer H, Dep
LICENSE: http://massbank.ufz.de/MassBank/files/
COPYRIGHT: Copyright (C) 2011 Eawag, Duebendorf, Switzerland
COMMENT: CONFIDENCE standard compound
COMMENT: EAWAG_UCHEM_ID 288
```

Authors, licence, copyright, etc

```
CH$NAME: Atrazine
```

MassBank Record Format

Compound Information

```
CH$NAME: Atrazine
CH$NAME: 6-chloro-N-ethyl-N'-isopropyl-1,3,5-triazine-2,4-diamine
CH$NAME: 6-chloranyl-N4-ethyl-N2-propan-2-yl-1,3,5-triazine-2,4-diamine
CH$COMPOUND_CLASS: N/A; Environmental Standard
CH$FORMULA: C8H14Cl1N5
CH$EXACT_MASS: 215.0932
CH$SMILES: c1(nc(nc(n1)Cl)NCC)NC(C)C
CH$IUPAC: InChI=1S/C8H14ClN5/c1-4-10-7-12-6(9)13-8(14-7)11-5(2)3/h5H,4H2,
CH$LINK: CAS 1912-24-9
CH$LINK: CHEBI 15930
CH$LINK: KEGG C06551
CH$LINK: PUBCHEM CID:2256
CH$LINK: INCHIKEY MXWJVTOOROXGIU-UHFFFAOYSA-N
CH$LINK: CHEMSPIDER 10774593
```

MassBank Record Format

Instrument and Measurement Information

```
AC$INSTRUMENT: LTQ Orbitrap XL Thermo Scientific
AC$INSTRUMENT_TYPE: LC-ESI-ITFT
AC$MASS_SPECTROMETRY: MS_TYPE MS2
AC$MASS_SPECTROMETRY: IONIZATION ESI
AC$MASS_SPECTROMETRY: ION_MODE POSITIVE
AC$MASS_SPECTROMETRY: FRAGMENTATION_MODE HCD
AC$MASS_SPECTROMETRY: COLLISION_ENERGY 45 % (nominal)
AC$MASS_SPECTROMETRY: RESOLUTION 7500
AC$CHROMATOGRAPHY: COLUMN_NAME XBridge C18 3.5um, 2.1x50mm, Waters
AC$CHROMATOGRAPHY: FLOW_GRADIENT 90/10 at 0 min, 50/50 at 4 min, 5/95 at
AC$CHROMATOGRAPHY: FLOW_RATE 200 ul/min
AC$CHROMATOGRAPHY: RETENTION_TIME 8.3 min
AC$CHROMATOGRAPHY: SOLVENT A water with 0.1% formic acid
AC$CHROMATOGRAPHY: SOLVENT B methanol with 0.1% formic acid
```

```
MS$FOCUSED_ION: BASE_PEAK 216.1012
MS$FOCUSED_ION: PRECURSOR_M/Z 216.101
MS$FOCUSED_ION: PRECURSOR_TYPE [M+H]+
MS$DATA_PROCESSING: DEPROFILE Spline
MS$DATA_PROCESSING: RECALIBRATE loess on assigned fragments and MS1
MS$DATA_PROCESSING: WHOLE RMassBank
```

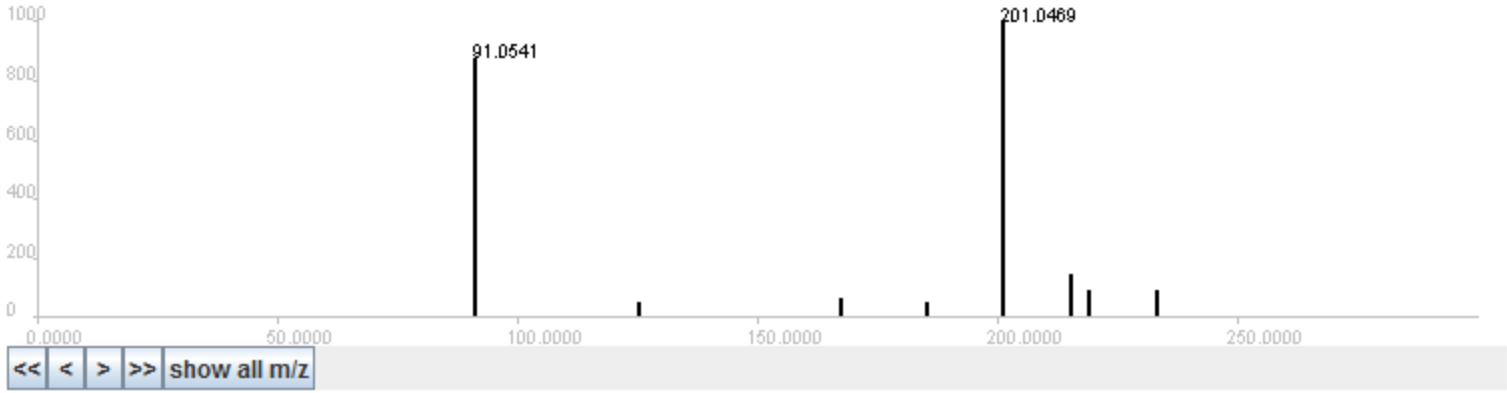
MassBank Record Format - Peaks

```
PK$ANNOTATION: m/z num {formula mass error(ppm)}
  68.0243 1 C2H2N3+ 68.0243 0.24
  71.0603 1 C3H7N2+ 71.0604 -0.35
  79.0058 1 CH4ClN2+ 79.0058 0.35
  96.0557 1 C4H6N3+ 96.0556 0.48
  104.001 1 C2H3ClN3+ 104.001 0.28
  132.0324 1 C4H7ClN3+ 132.0323 0.37
  138.0779 1 C5H8N5+ 138.0774 3.1
  138.1029 1 C7H12N3+ 138.1026 2.36
  146.0228 1 C3H5ClN5+ 146.0228 0.01
  146.0481 1 C5H9ClN3+ 146.048 0.95
  174.0544 1 C5H9ClN5+ 174.0541 1.9
  216.1014 1 C8H15ClN5+ 216.101 1.81
PK$NUM_PEAK: 12
PK$PEAK: m/z int. rel.int.
  68.0243 949141.4 28
  71.0603 167093.1 5
  79.0058 1014555.8 30
  96.0557 1956578 59
  104.001 2256202.5 68
  132.0324 1805661.3 54
  138.0779 349329.9 10
  138.1029 280901.6 8
  146.0228 1313633.2 39
  146.0481 263134.4 7
  174.0544 32958196.3 999
  216.1014 29362258.7 890
//
```

MassBank Record – Example Unknown (so far)

CASMI2012 LC Challenge 13; APCI-ITFT; MS2; CE:45 CID;

Mass Spectrum



Chemical Structure

Not Available

```
ACCESSION: SMI00131
RECORD_TITLE: CASMI2012 LC Challenge 13; APCI-ITFT; MS2; CE:45 CID;
DATE: 2012.08.31 (Created 2012.08.31)
AUTHORS: S. Neumann: IPB-Halle, Germany & E. Schymanski: Eawag, Switzerland
COPYRIGHT: CASMI2012
COMMENT: Unknown
COMMENT: http://casmi-contest.org/challenges-cat1-2.shtml

CH$NAME: CASMI2012 LC Challenge 13
CH$COMPOUND_CLASS: N/A; Unknown for CASMI
CH$FORMULA: N/A
CH$EXACT_MASS: N/A
CH$SMILES: N/A
CH$IUPAC: N/A

AC$INSTRUMENT: LTQ-Orbitrap
AC$INSTRUMENT_TYPE: APCI-ITFT
AC$ANALYTICAL_CONDITION: MODE POSITIVE
AC$MASS_SPECTROMETRY: MS_TYPE MS2
AC$MASS_SPECTROMETRY: ION_MODE POSITIVE

MS$FOCUSED_ION: PRECURSOR_M/Z N/A
MS$FOCUSED_ION: PRECURSOR_TYPE N/A

PK$NUM_PEAK: 8
PK$PEAK: m/z int. rel.int.
  91.0541 2211410.3 873
  125.0152 130666.3 52
  167.0859 161086.1 64
  185.0519 125915.7 50
  201.0469 2531123 999
  215.0626 371698.1 147
  219.0576 223408.9 88
  233.0733 226928.3 90
//
```

MassBank Record Index

Have a look at what is in MassBank:

Contributor:

Instrument Type: (Orbitrap = ITFT)

Compound Name: watch out for numbers…

Contributor :

NORMAN EMPOMASS (2,081)	CASMI (42)	Eawag (944)
UFZ (2,509)	Keio Univ. (5,629)	RIKEN (1,721)
Waters (2,993)	Kyoto Univ. (185)	Chubu Univ. (2,628)
Kazusa (273)	Nihon Univ. (75)	Univ. Toyama (253)
Tottori Univ. (16)	IPB Halle (528)	Fukuyama Univ. (340)
Metabolon (149)	UOEH (35)	NAIST (680)
Osaka MCHRI (20)	Univ. Connecticut (510)	Osaka Univ. (502)
IMM, CAMS & PUMC, China (192)	PFOS research group (365)	MPI for Chemical Ecology (691)

nstrument Type :

GC-EI-MS (2,097)	LC-ESI-QTOF (2,750)	LC-APCI-ITFT (10)
APCI-ITFT (1,206)	LC-ESI-ITFT (4,504)	ESI-ITFT (1,310)
LC-ESI-IT (515)	LC-ESI-QQ (5,038)	CE-ESI-TOF (20)
GC-EI-TOF (1,016)	LC-ESI-Q (2,720)	EI-EBEB (12)
FAB-EBEB (173)	LC-ESI-QIT (378)	EI-B (91)
LC-ESI-ITTOF (253)	FAB-EB (5)	ESI-IT-MS/MS (149)
CI-B (1)	FAB-B (26)	FD-B (3)
MALDI-TOF (17)	ESI-QqTOF-MS/MS (510)	ESI-QqIT-MS/MS (140)
ESI-QqQ-MS/MS (52)	APPI-QqQ-MS (2)	APPI-QqQ-MS/MS (27)
ESI-QQ (78)	LC-APPI-QQ (258)	

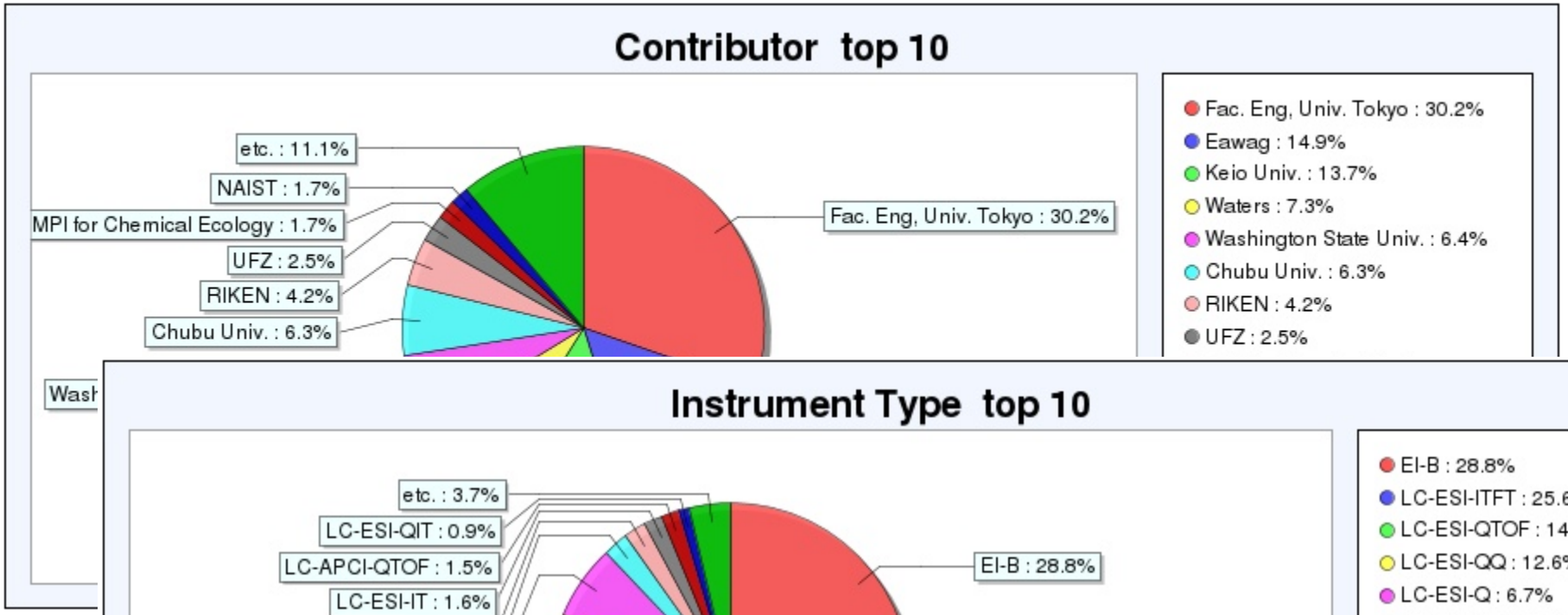
onization Mode : Positive (16,197) Negative (7,164)

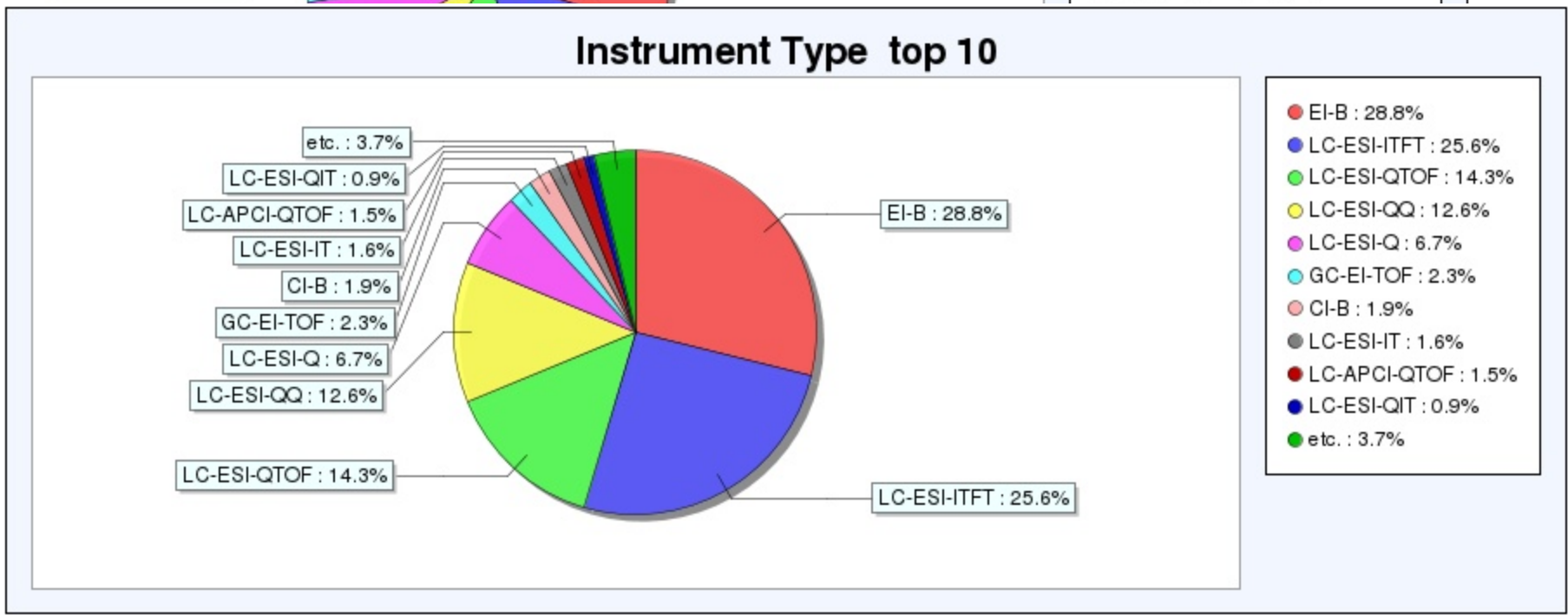
Compound Name :

A (1,185)	B (1,122)	C (1,372)	D (1,559)	E (426)	F (421)
G (793)	H (429)	I (657)	J (2)	K (223)	L (1,254)
M (996)	N (1,005)	O (388)	P (3,687)	Q (184)	R (287)
S (941)	T (1,249)	U (887)	V (128)	W (3)	X (51)
Y (5)	Z (72)	1-9 (3,844)	Others (191)		

MassBank Record Index

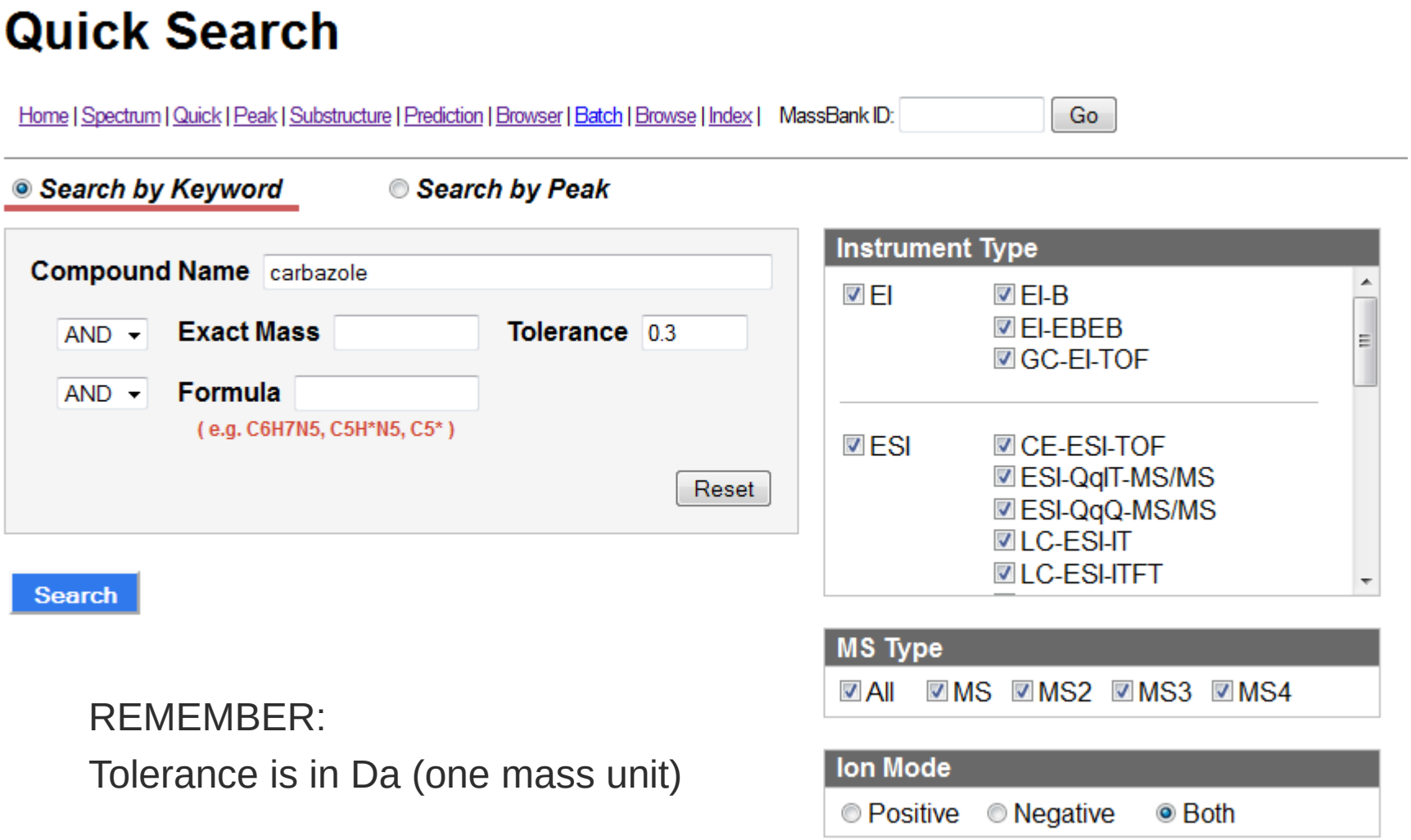
Scroll to see the contents (and see if your institute is in the top 10!)





MassBank Quick Search

Search compounds or peaks



REMEMBER:

Tolerance is in Da (one mass unit)

Quick Search Results

Search Parameters :

Compound Name: **carbazole**

Instrument Type: **EI-B , EI-EBEB , GC-EI-TOF , CE-ESI-TOF , ESI-QqIT-MS/MS , ESI-QqQ-MS/MS , LC-ESI-IT , LC-ESI-ITFT , LC-ESI-ITTOF , LC-ESI-Q , LC-ESI-QIT , LC-ESI-QQ , LC-ESI-QTOF , CI-B , FAB-B , FAB-EB , FAB-EBEB , FD-B , FI-B , LC-APPI-QQ , MALDI-TOF , MALDI-TOFTOF**

MS Type: **All**

Ion Mode: **Both**

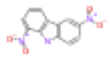
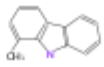
Edit / Resubmit Query

Results : 34 Hit. (1 - 34 Displayed)

Open All Tree | Multiple Display | Spectrum Search

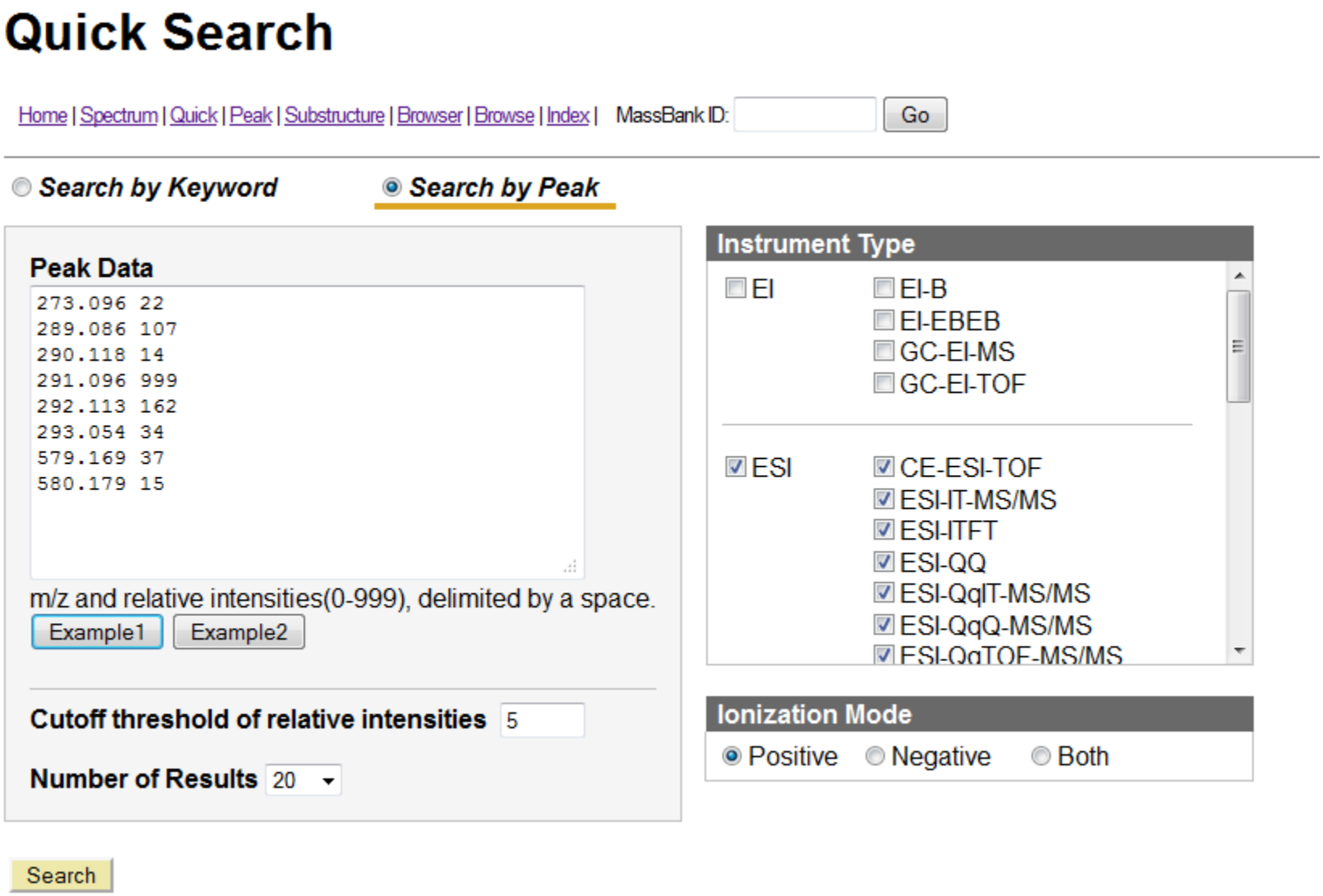
First Prev *1* Next Last (Total *1* Page)

▼ Results End

Name	Formula / Structure	ExactMass	ID
1,6-DINITROCARBAZOLE (1 spectrum)	C12H7N3O4 	257.04366	
1-METHYLCARBAZOLE (1 spectrum)	C13H11N 	181.08915	
2-METHYLCARBAZOLE	C13H11N	181.08915	

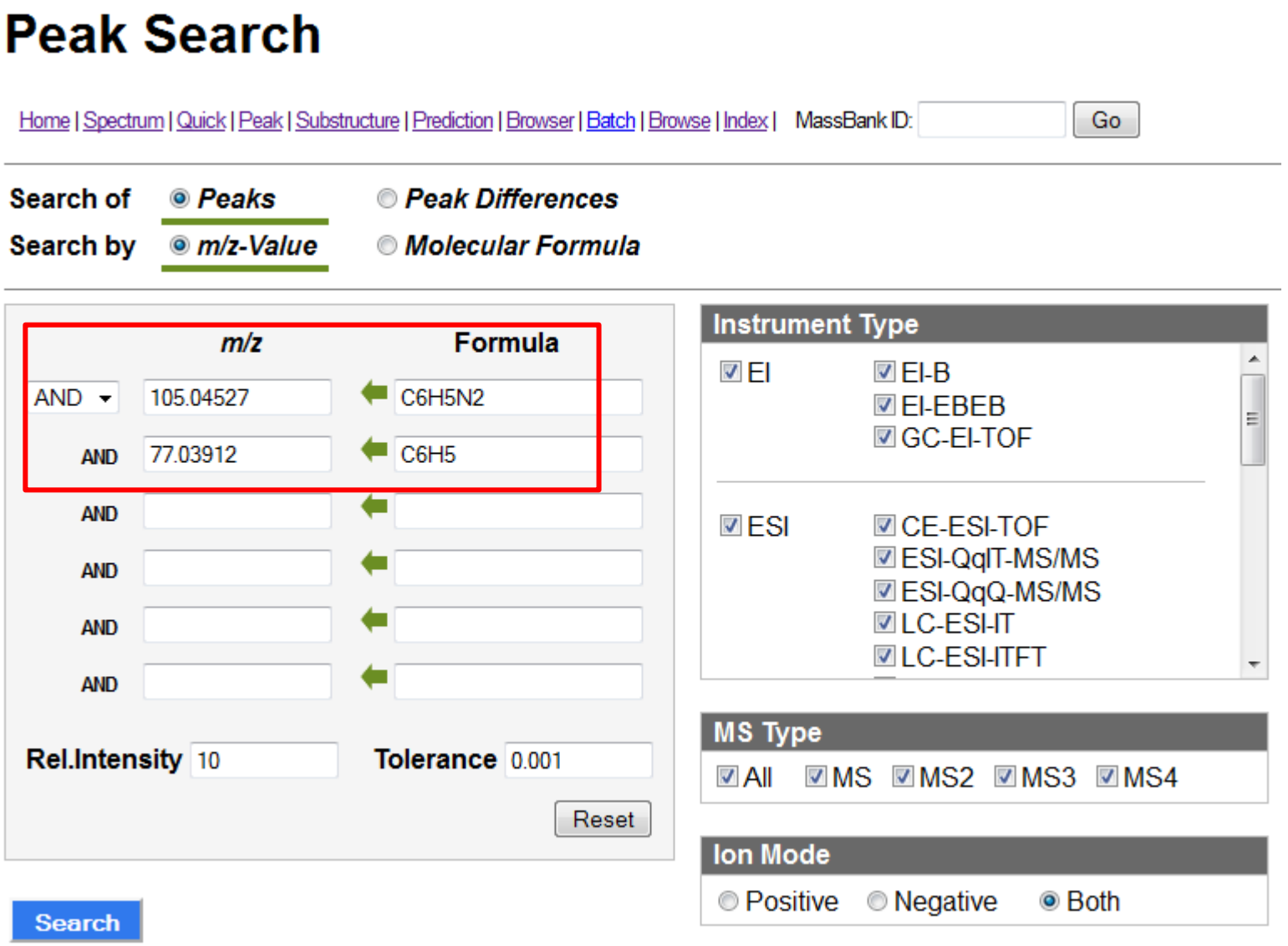
MassBank Quick Search

Search compounds or peaks



MassBank Peak Search

Different to Quick Search → Search by Peak!



Peak Search Results (Peaks by *m/z* value)

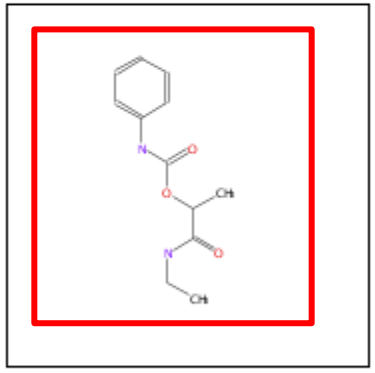
MassBank Record: EA013612

Home | Spectrum | Quick | Peak | Substructure | Prediction | Browser | Batch | Browse | Index | MassBank ID: Go

Carbetamide; LC-ESI-ITFT; MS2; 75%; R=15000; [M+H]+

Mass Spectrum

Chemical Structure

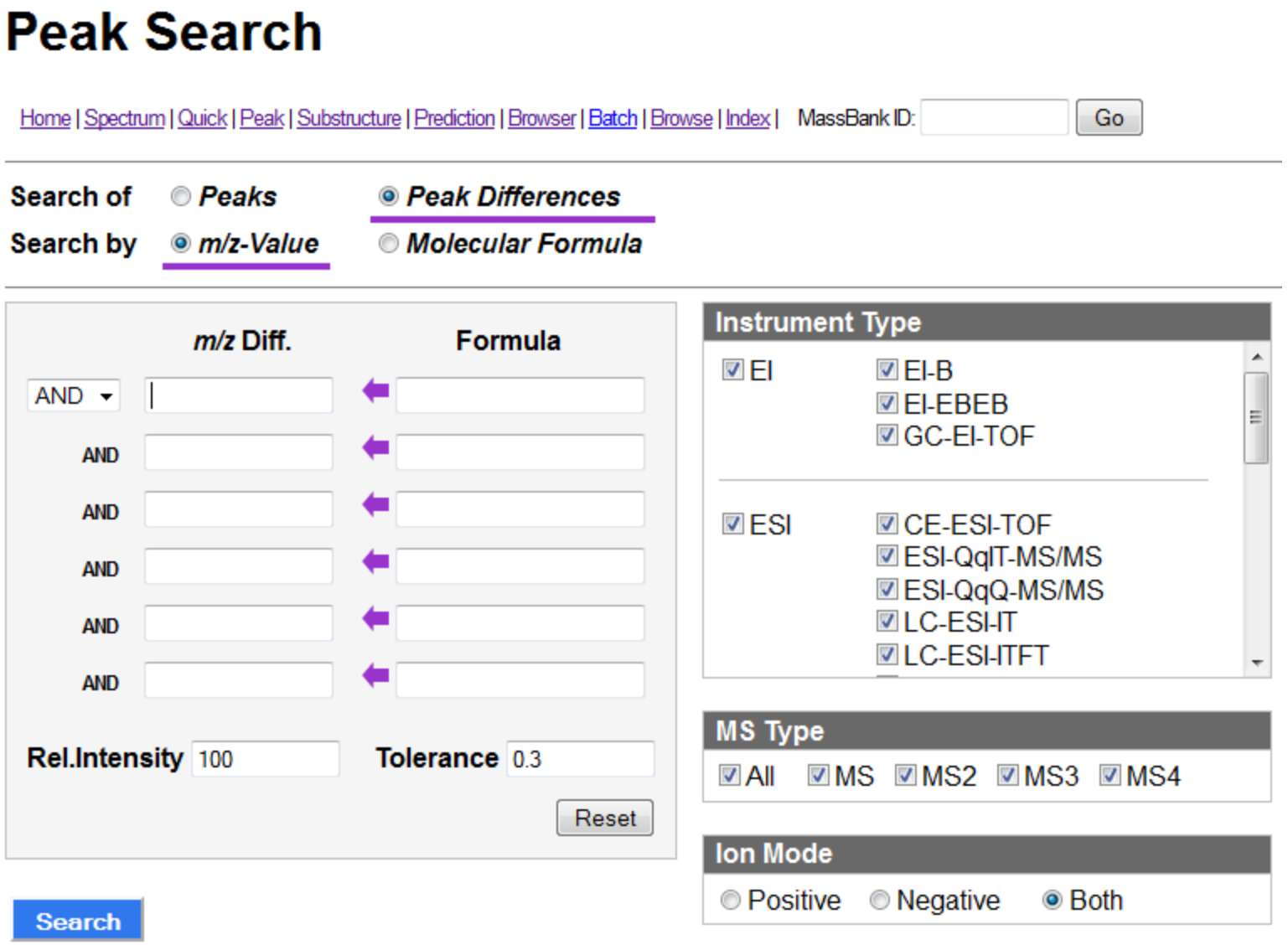


ACCESSION: EA013612
RECORD_TITLE: Carbetamide; LC-ESI-ITFT; MS2; 75%; R=15000; [M+H]+
DATE: 2012.08.02
AUTHORS: Stravs M, Schymanski E, Singer H, Department of Environmental Chemistry, Eawag

☐	⊞ **Aminocaproic acid** 1 spectrum	**C6H13NO2**	**131.09463**

MassBank Peak Search

By Peak Differences and m/z value



MassBank Peak Search

By Peak Differences and Molecular Formula

Peak Search

Home | Spectrum | Quick | Peak | Substructure | Prediction | Browser | Batch | Browse | Index | MassBank ID: Go

Search of ○ *Peaks* ◉ *Peak Differences*

Search by ○ *m/z-Value* ◉ *Molecular Formula*

Neutral Loss 1 Formula		Neutral Loss 2 Formula		Neutral Loss 3 Formula		Neutral Loss 4 Formula		Neutral Loss 5 Formula
	AND		AND		AND		AND	

◉ AND ○ SEQUENCE Reset

* The targets of Peak Search Advanced are only Keio and Riken data.

Search

Instrument Type

☑ ESI ☑ LC-ESI-QTOF

MS Type

☑ All ☑ MS2

Ion Mode

◉ Positive ○ Negative ○ B

MassBank Spectrum Search

http://www.massbank.jp/sample/sample.txt

```
// Specification of file format
// - A line started by '//' is a comment line.
// - A peak is denoted by its m/z and intensity
     separated by one or more spaces.
// - Delimiter of spectra: one or more empty lines is
     needed between spectra.
// - A line started by 'ID:' specifies the value of the
     ID column in the spectrum table for each spectrum.
// - A line started by 'Name:' specifies the value of the
     name column in the spectrum table for each spectrum.
// - A "Nist Format" file can be used without any
     modification.
```

If you saved the data before the workshop, you can try this example with “SpectrumSearch_EnvEgs.txt”

MassBank Spectrum Search Results

Spectrum Search

Home | Spectrum | Quick | Peak | Substructure | Prediction | Browser | Batch | Browse | Index | MassBank ID: Go

Browse... File Read sample file sample archive

DB File

No.	Name	ID
1	EA013612	US000001
2	EA008812	US000002

Click on file name to search

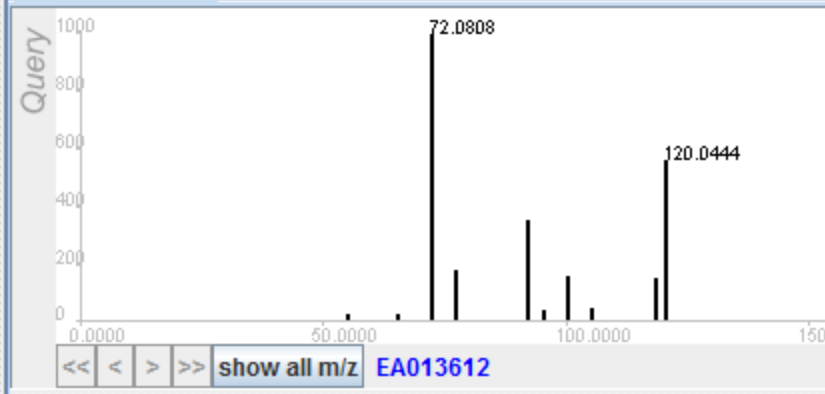
Package View display mode : selected related

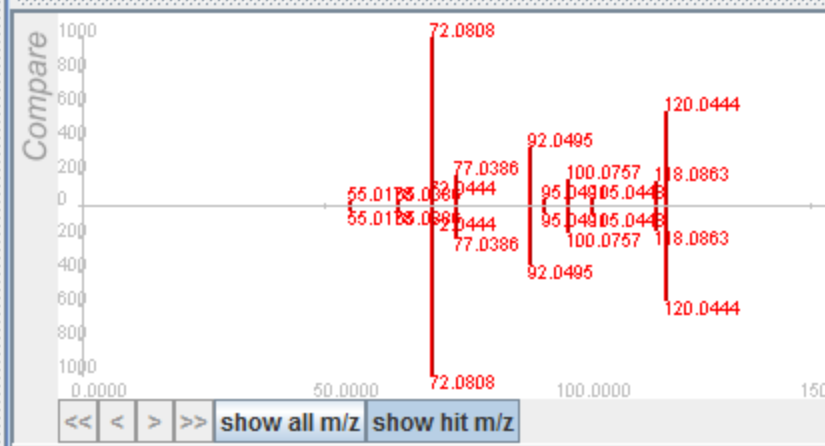
Search Parameter Setting

Result

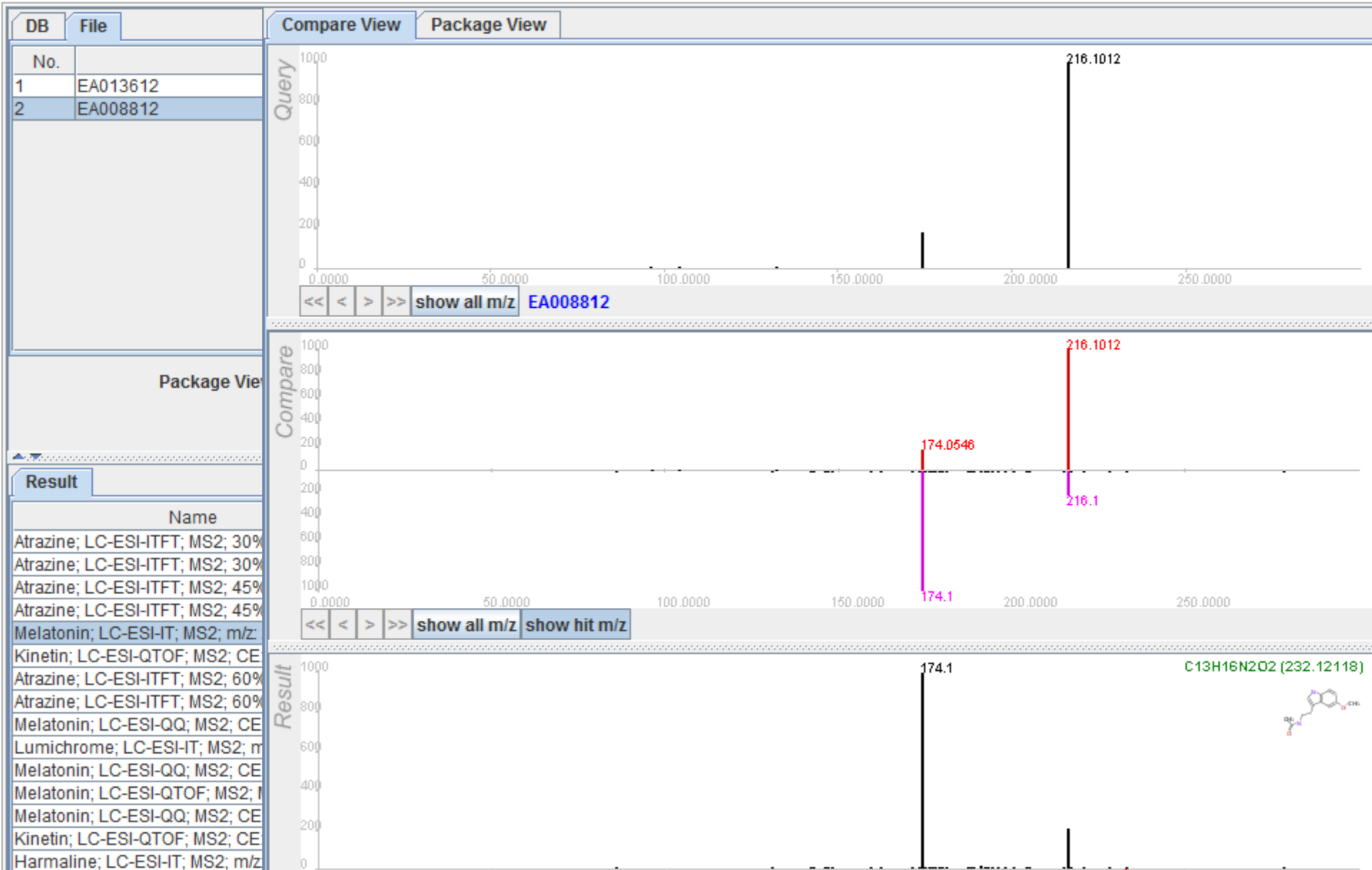
Name	Score	Hit	ID	Ion	Contributor	
Carbetamide; LC-ESI-ITFT; MS2; 75%; R=1...	0.941860...	12	EA013612	P	Eawag	1
Carbetamide; LC-ESI-ITFT; MS2; 75%; R=7...	0.937374...	12	EA013606	P	Eawag	2
Carbetamide; LC-ESI-ITFT; MS2; 60%; R=7...	0.908825...	11	EA013605	P	Eawag	3
Carbetamide; LC-ESI-ITFT; MS2; 60%; R=1...	0.903718...	10	EA013611	P	Eawag	4
Carbetamide; LC-ESI-ITFT; MS2; 90%; R=3...	0.893760...	12	EA013613	P	Eawag	5

Compare View Package View



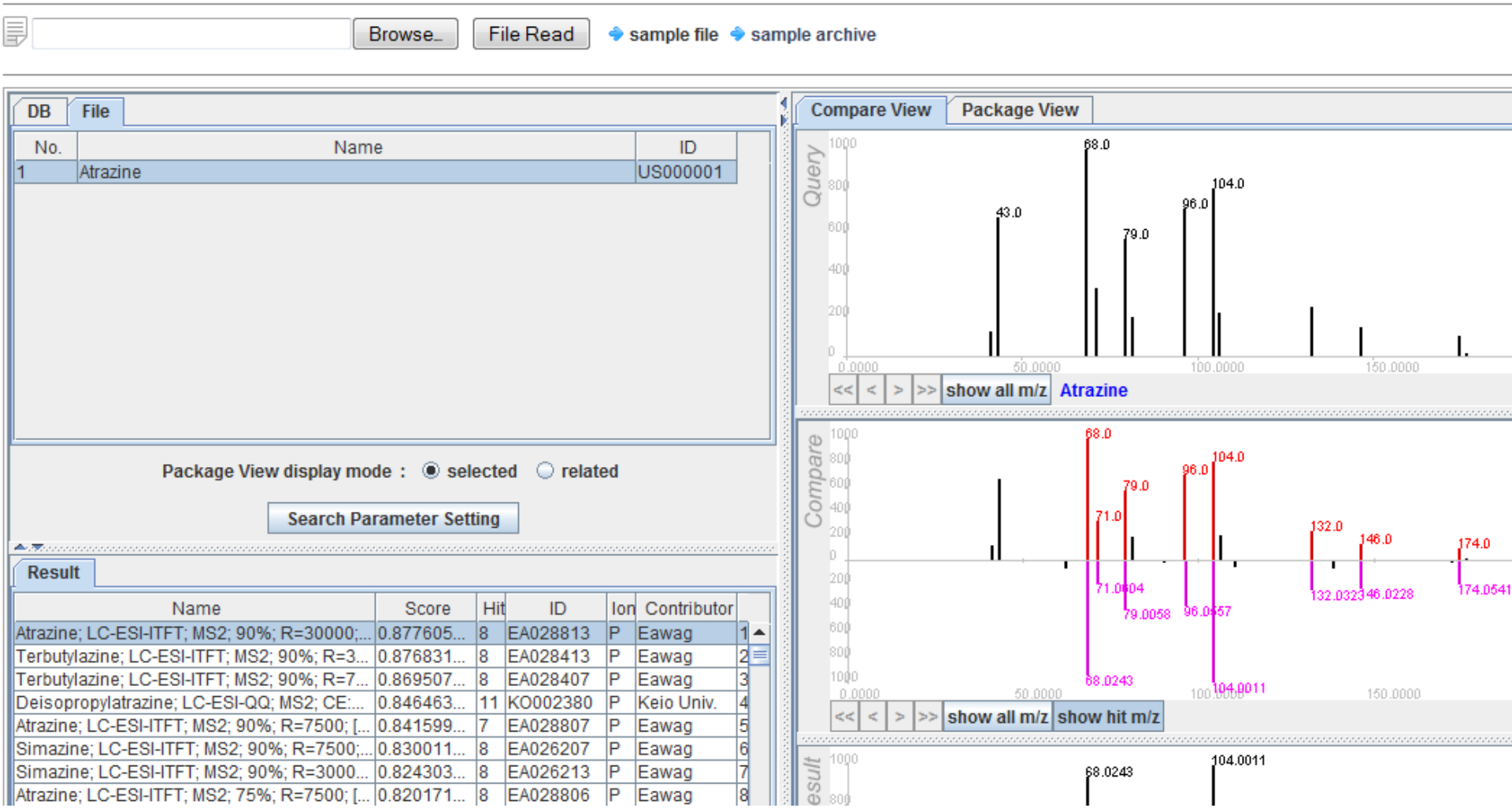


MassBank Spectrum Search Results



MassBank Spectrum Search with NIST Format

Example Spectrum “atrazine_msms_1563.msp”



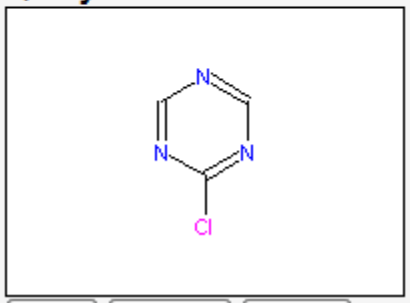
MassBank Substructure Search

How many spectra in MassBank have the atrazine backbone?

Search MassBank | Search KNApSAcK

Substructure

Query1



Edit | Molfile | Clear

AND

Query2

Edit | Molfile | Clear

Comparison of pi-electron for each atom

number in query = number in target

* Double and triple bound is translated to pi-electrons of the bonded atoms.

Instrument Type

- ☑ EI
  - ☑ EI-B
  - ☑ EI-EBEB
  - ☑ GC-EI-TOF
- ☑ ESI
  - ☑ CE-ESI-TOF
  - ☑ ESI-QqIT-MS/MS
  - ☑ ESI-QqQ-MS/MS
  - ☑ LC-ESI-IT
  - ☑ LC-ESI-ITFT

MS Type

☑ All ☑ MS ☑ MS2 ☑ MS3 ☑ MS4

Ion Mode

○ Positive ○ Negative ◉ Both

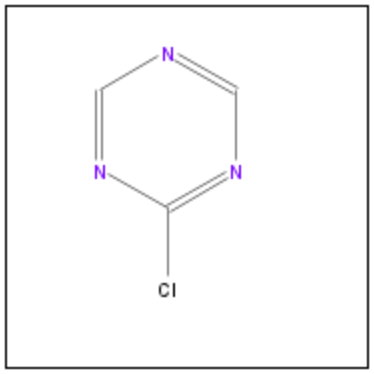
Peak Search (Option)

m/z , ,

Tolerance of *m/z* 0.3

Search

Query1



MassBank Substructure Search Results

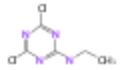
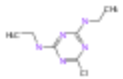
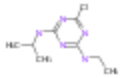
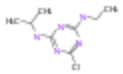
Edit / Resubmit Query

Results : 59 Hit. (1 - 59 Displayed)

Open All Tree | Multiple Display | Spectrum Search

First Prev **1** Next Last (Total **1** Page)

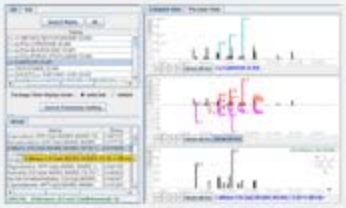
▼ Results End

Name	Formula / Structure	ExactMass	ID
2,6-DICHLORO-4-(N-ETHYL)AMINO-1,3,5-TRIAZINE 1 spectrum	C5H6Cl2N4 	191.99695	
2-CHLORO-4,6-BIS(ETHYLAMINO)-1,3,5-TRIAZINE 1 spectrum	C7H12ClN5 	201.07812	
6-CHLORO-N-ETHYL-N'-(1-METHYLETHYL)-1,3,5-TRIAZINE... 1 spectrum	C8H14ClN5 	215.09377	
Atrazine 15 spectra	C8H14Cl1N5 	215.09320	
Deisopropylatrazine	C5H8ClN5	173.04682	

Internal MassBank Administration

Eawag Uchem-MassBank

Spectrum Search



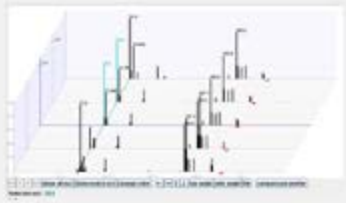
Quick Search

Substructure Search

Browse Page

Peak Search

Spectral Browser



Record Index

- MassBank Administration Tool → **User and Password…**
- WEB-API WSDL

Internal MassBank Administration

MassBank Administration Tool

Main Menu

Record Validator | **Validator**

ADMIN MENU: Record Validator | Record Registration | Structure Registration | File Upload | Sql File Generator | Version Information | Instrument Editor | Record List | Structure List | Validator | Record List Generator | Dababase Manager | HOME

Record Validator

Database : ---------- **Record Version :** 2 1 *(old record version)*

Record Archive : Browse... Validation

** please specify your [**recdata.zip**] or [*.msbk].*

Internal MassBank Administration

Database Manager

No.	DB Name	Short Label	Long Label
2	UchemA	Eawag Uchem Adducts	Uchem EZ records of adduct spectra

URL Type: (•) internal ( ) external

URL: http://uchem-massbank/MassBank/

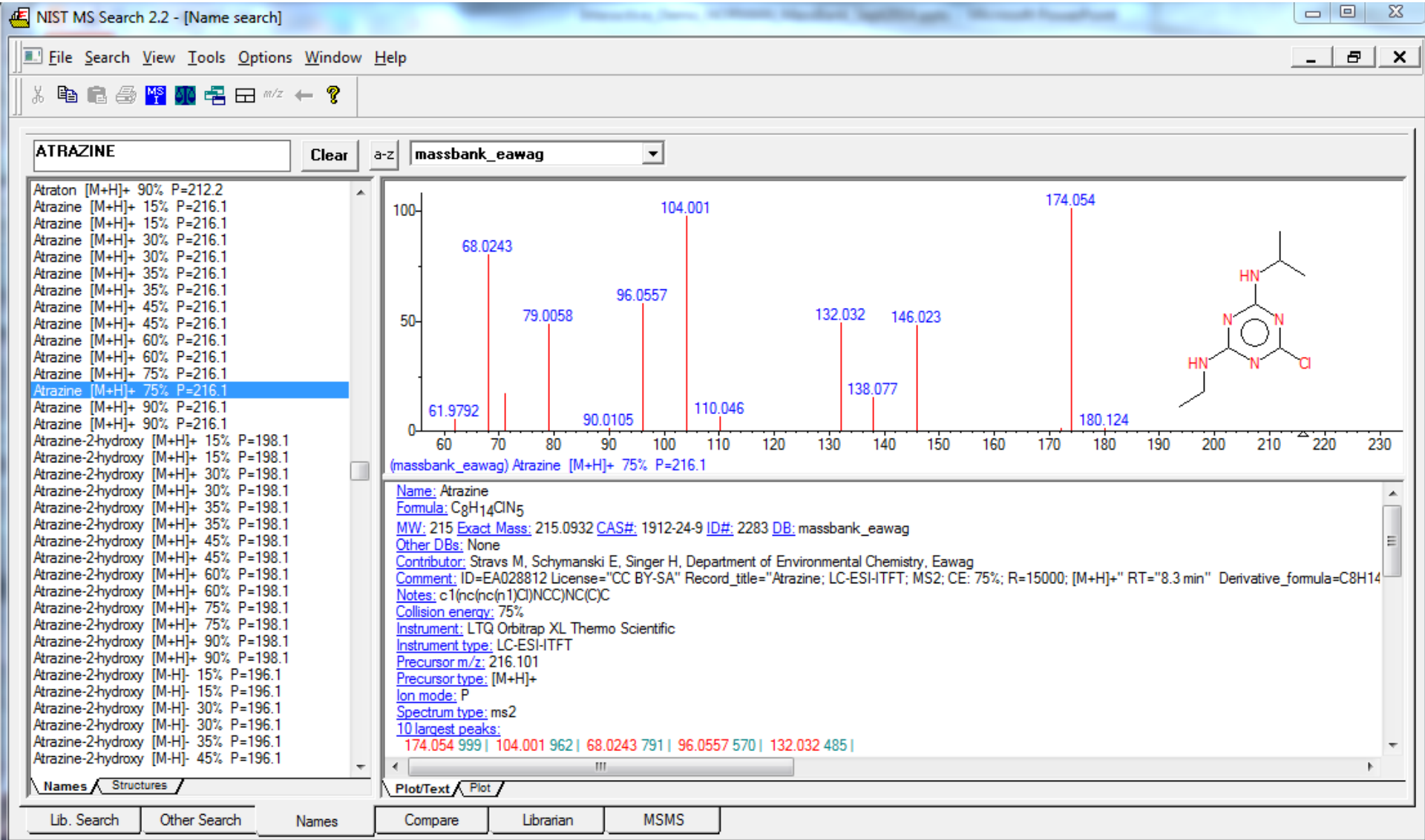
Add | Edit | Delete

26 database (20 external database)

No.	DB Name	URL	Short Label	Long Label	Status	Details
0	MassBank	*http://uchem-massbank/MassBank/*	EQ Uchem Q Ex	Eawag Uchem-MassBank	ok	
1	UchemEZ	*http://uchem-massbank/MassBank/*	EA Uchem Orbi Test	Annotated Uchem Standard Spectra for Approval	ok	
2	UchemA	*http://uchem-massbank/MassBank/*	Eawag Uchem Adducts	Uchem EZ records of adduct spectra	ok	
3	EA	*http://uchem-massbank/MassBank/*	EA Uchem Orbi	Annotated Uchem Standard Spectra	ok	
4	Waters	*http://www.massbank.jp/*	Waters	Nihon Waters K.K.	ok	external database.
5	Kyoto	*http://www.massbank.jp/*	Kyoto Univ.	Kyoto University	ok	external database.
6	MassBank	*http://157.110.6.77/MassBank/*	Chubu Univ.	College. Life and Health Sci, Chubu U	ok	external database.

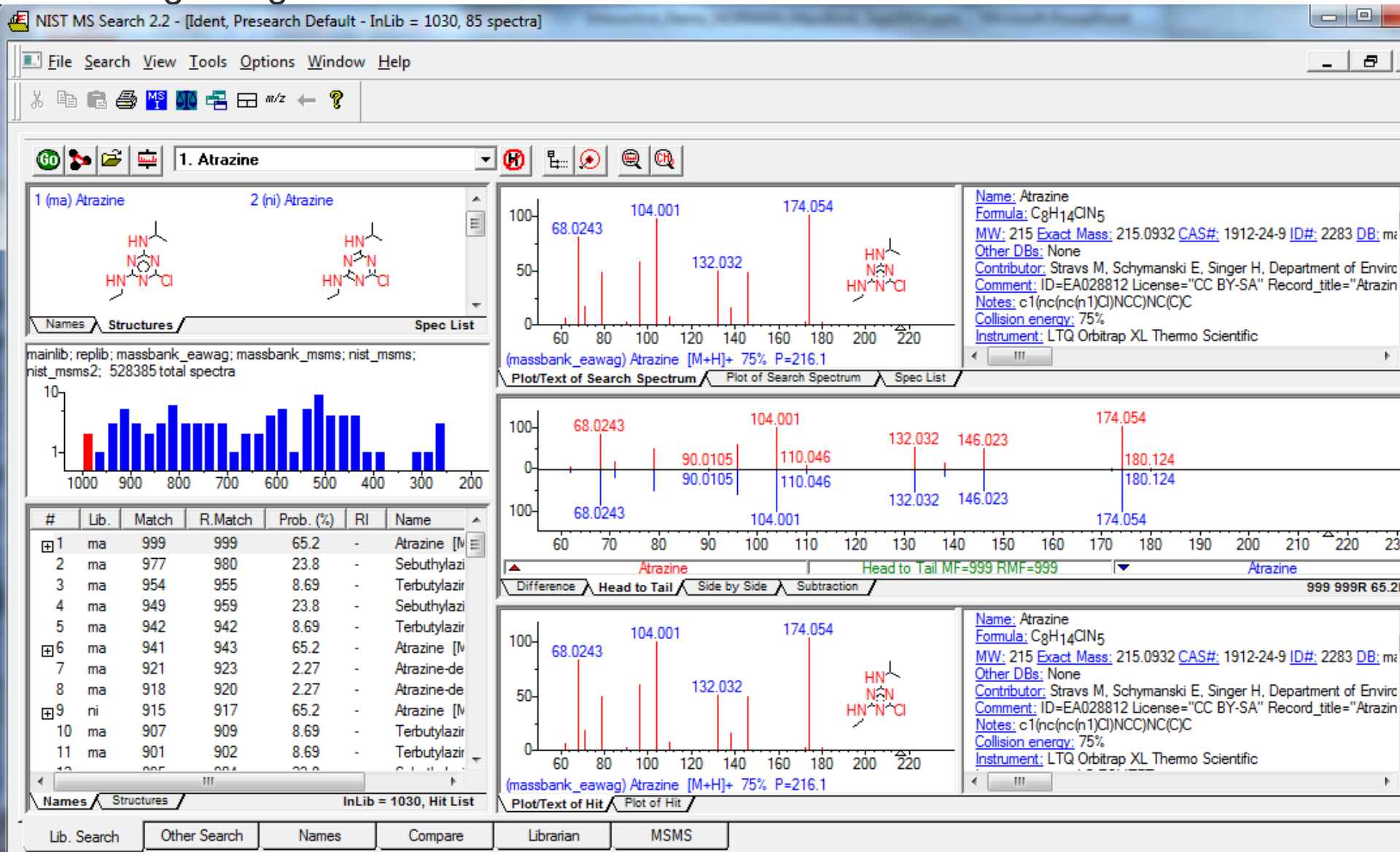
MassBank in NIST

The beginning of a collaboration…



MassBank in NIST

The beginning of a collaboration…



MassBank in NIST

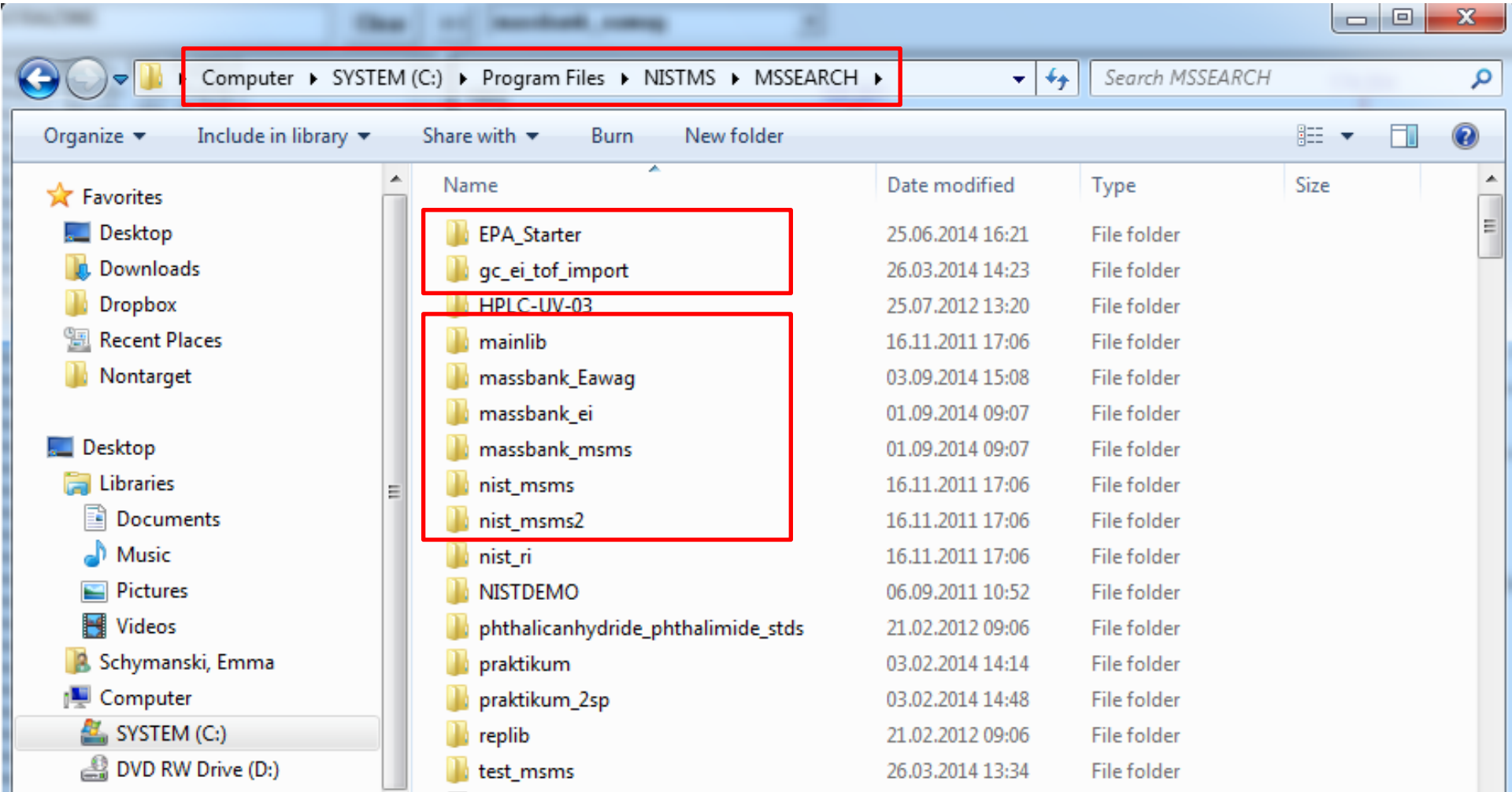
The beginning of a collaboration

- Currently a few sub-sets of MassBank available
  - massbank_ei – all EI spectra up to July 2012
  - massbank_msms – all MS/MS spectra up to July 2012
  - massbank_eawag – all Eawag MS/MS spectra
- NIST do not want to add our spectra into NIST distribution
  - “Standard spectral library” – reference laboratory
  - License is creative commons, but NIST is commercial => incompatible
  - BUT, NIST want to improve access to alternative databases within NIST
    - Several other libraries also available to download
- Very easy install for NIST users
  - Download zip file, unzip, copy to NIST folder and it’s there!
  - Can browse straight away
  - Need to add to search path manually

MassBank in NIST

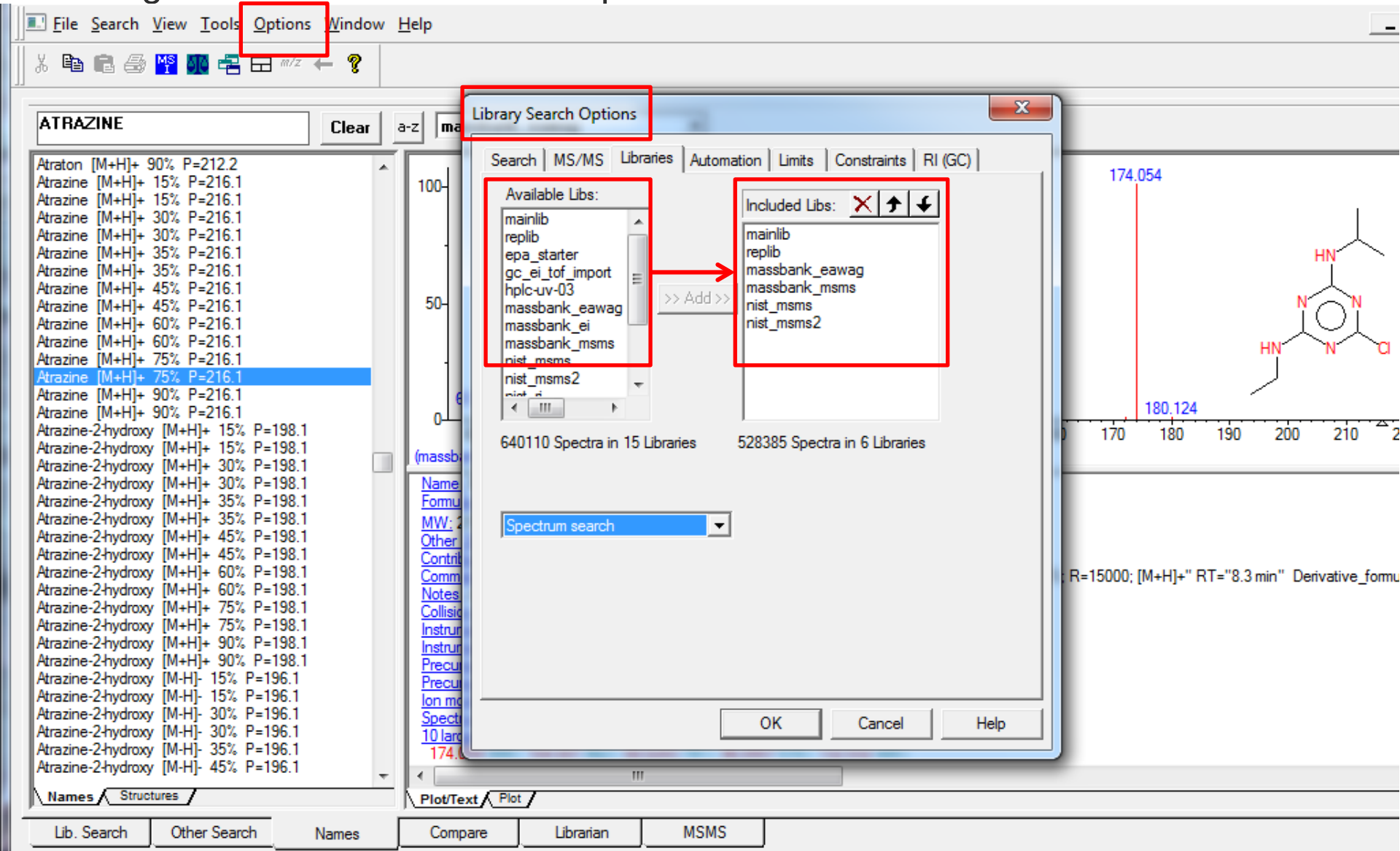
“Installing” MassBank

Simply copy unzipped folder into NISTMS/MSSEARCH path



MassBank in NIST

Adding MassBank to the search path



MassBank in NIST

The beginning of a collaboration

- Details under discussion
  - Easy install for NIST users => hard for “MassBank” to generate files
    - NIST will provide this service for the moment
    - Available to download via MassBank website
      - How to (best) split up the databases
  - Information in MassBank records not available in NIST 1:1
    - Current files are “work in progress”
    - Conversion will need to be optimized in the coming months before “going public”
  - Increasing the accessibility will improve usability
    - Web interface of MassBank complementary to offline MSSearch
    - Future developments to MassBank to improve functionality

Thank you for listening

Any Questions?

Emma Schymanski (Eawag) emma.schymanski@eawag.ch

Emma & Michele (Eawag) massbank@eawag.ch

Tobias Schulze (UFZ) tobias.schulze@ufz.de

Steffen Neumann (IPB) sneumann@ipb-halle.de