METABOLOMICS SOCIETY
EARLY- CAREER MEMBERS NETWORK

METABOLOMICS
SOCIETY

INTERNATIONAL METABOLOMICS SOCIETY

EARLY-CAREER MEMBERS' NETWORK (EMN)

WEBINAR SERIES - 10TH SESSION

Justin van der Hooft
EMN – Committee member

Glasgow Polyomics, University of Glasgow, UK

6th October, 2016

Connect with us online!

facebook.com/MetabolomicsSociety

@MetabolomicsSoc

Metabolomics Society

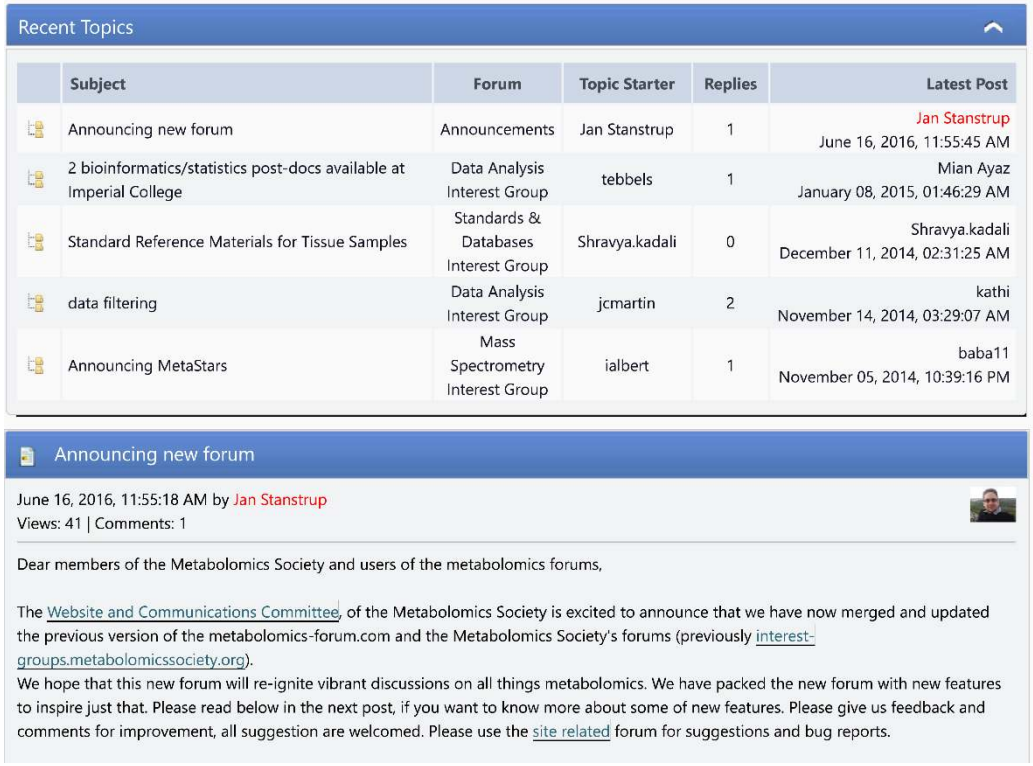
info.emn@metabolomicssociety.org

METABOLOMICS SOCIETY
EARLY- CAREER MEMBERS NETWORK

Metabolomics Forum

METABOLOMICS SOCIETY

Want to talk about metabolomics?



Recent Topics

Subject	Forum	Topic Starter	Replies	Latest Post
Announcing new forum	Announcements	Jan Stanstrup	1	Jan Stanstrup June 16, 2016, 11:55:45 AM
2 bioinformatics/statistics post-docs available at Imperial College	Data Analysis Interest Group	tebbels	1	Mian Ayaz January 08, 2015, 01:46:29 AM
Standard Reference Materials for Tissue Samples	Standards & Databases Interest Group	Shravya.kadali	0	Shravya.kadali December 11, 2014, 02:31:25 AM
data filtering	Data Analysis Interest Group	jcmartin	2	kathi November 14, 2014, 03:29:07 AM
Announcing MetaStars	Mass Spectrometry Interest Group	ialbert	1	baba11 November 05, 2014, 10:39:16 PM

Announcing new forum

June 16, 2016, 11:55:18 AM by Jan Stanstrup
Views: 41 | Comments: 1

Dear members of the Metabolomics Society and users of the metabolomics forums,

The Website and Communications Committee, of the Metabolomics Society is excited to announce that we have now merged and updated the previous version of the metabolomics-forum.com and the Metabolomics Society's forums (previously interest-groups.metabolomicssociety.org).
We hope that this new forum will re-ignite vibrant discussions on all things metabolomics. We have packed the new forum with new features to inspire just that. Please read below in the next post, if you want to know more about some of new features. Please give us feedback and comments for improvement, all suggestion are welcomed. Please use the site related forum for suggestions and bug reports.

metabolomics-forum.com and MetSoc interest-groups are now one!

A new forum is now available with a new friendly portal with latest posts!

metabolomics-forum.com

Mass Spectral Libraries for Small Molecules

Dr. Emma Schymanski

Eawag – Environmental Chemistry
Swiss Federal Institute for Aquatic Science and Technology
Switzerland

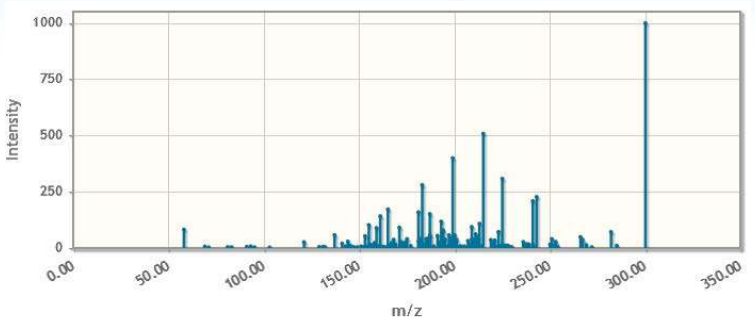
Mass Spectral Libraries for Small Molecules

EMN Webinar Series – Session 10

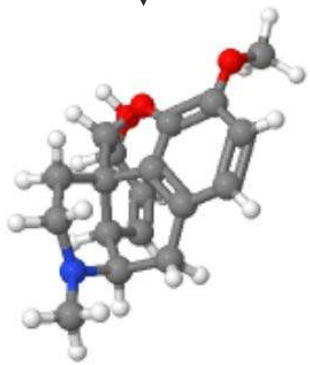
Emma Schymanski
Eawag: Swiss Federal Institute of Aquatic Science and Technology
emma.schymanski@eawag.ch
www.eawag.ch/~schymaem

October 6th, 2016

Plan for this webinar

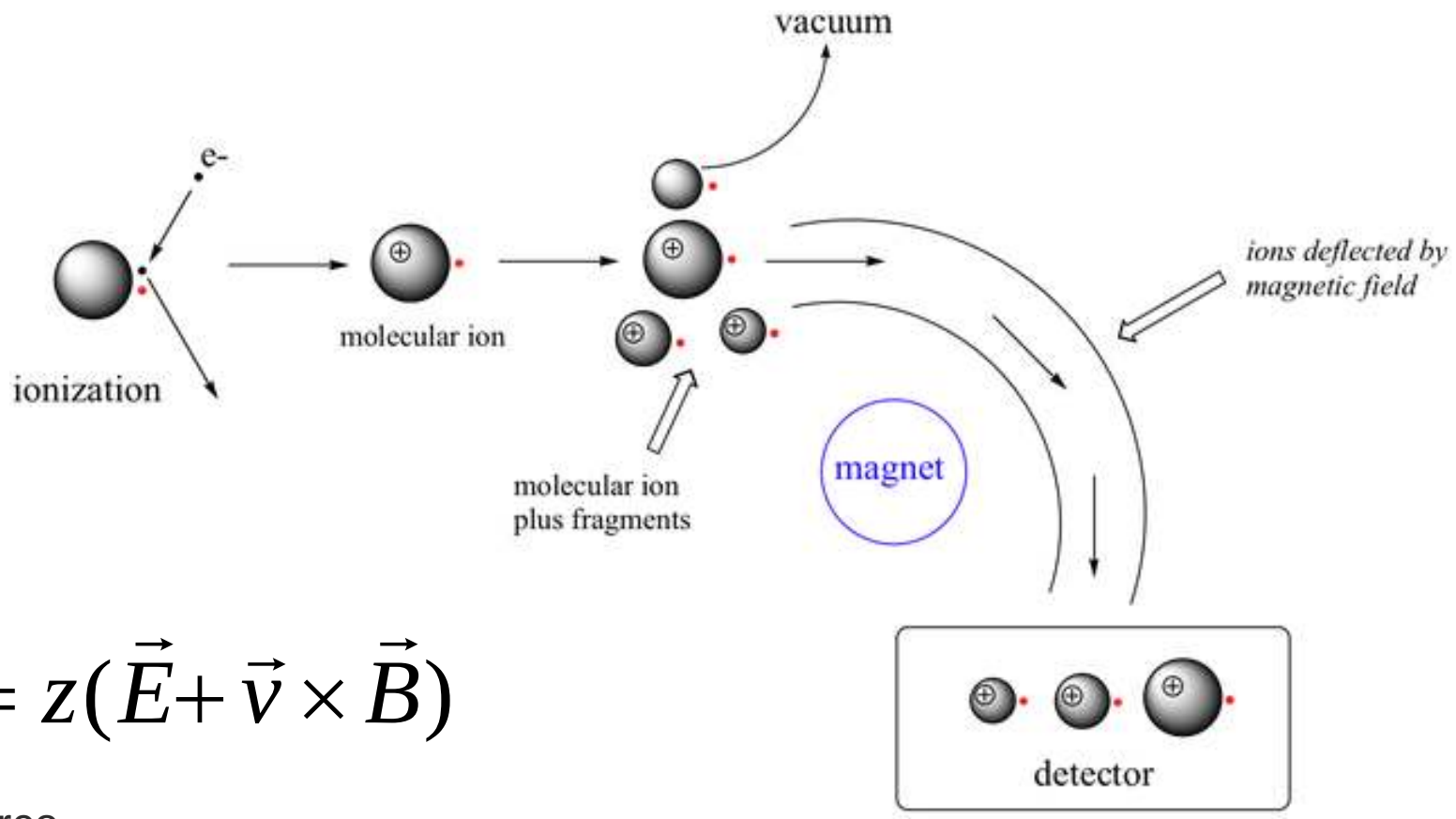


- Introduction to Mass Spectra
  - Electron impact mass spectrometry (EI-MS)
  - Softer ionization - tandem mass spectrometry (MS/MS)
- Overview of (GC-)EI-MS libraries
- Overview of (LC-)MS/MS libraries
- *Why do we need so many different libraries?*
- *Are you really sure? What to look out for … and …*
- *What to do if your spectrum isn't in the library!*
- Questions – via host (10 min)



EA278005 (massbank.eu) plotted in MetFragBeta
Structure image from ChemSpider, ID 4447447

Introduction – Mass Spectra



$$\vec{F} = z(\vec{E} + \vec{v} \times \vec{B})$$

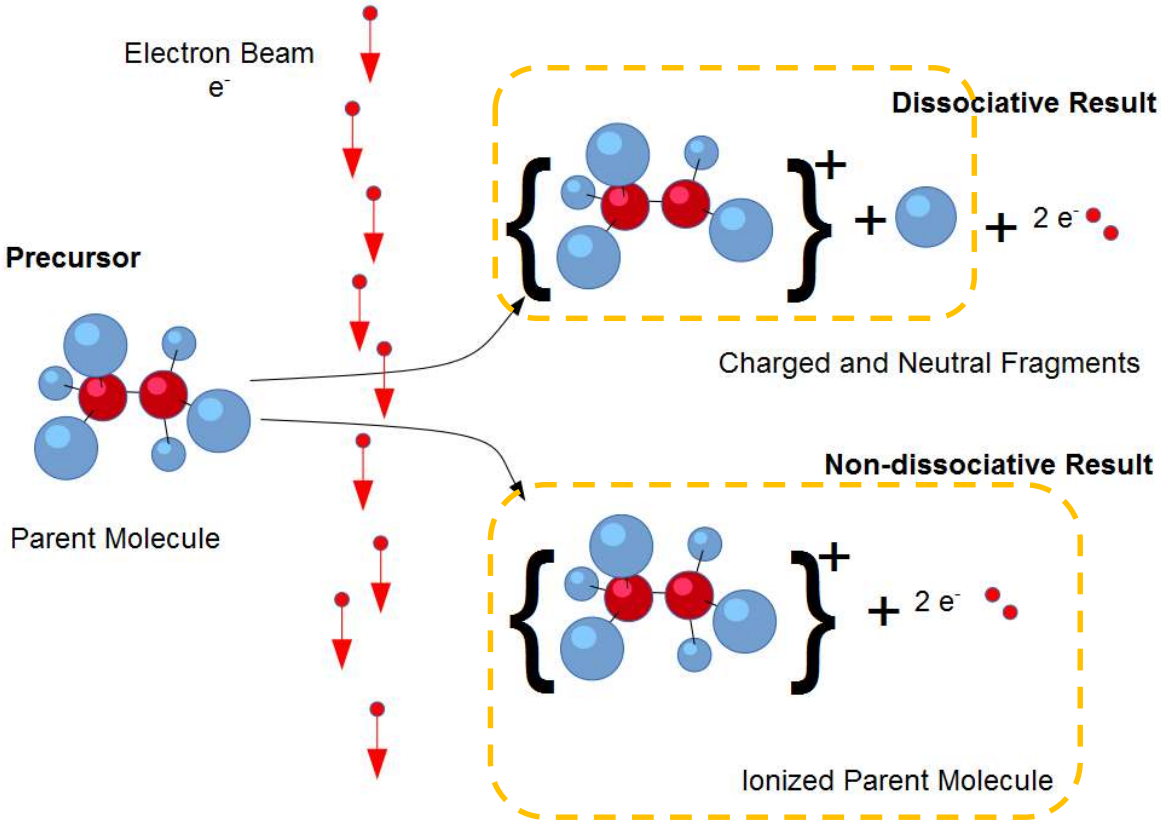
F = Force
z = charge
E = electrical field
v×*B* = vector cross product of *ion velocity* and *magnetic field*

Image: chemwiki.ucdavis.edu

Introduction – Mass Spectra

Example with Electron Ionization (EI)

Gas phase reaction: $M + e^- \rightarrow M^{+\cdot} + 2e^-$

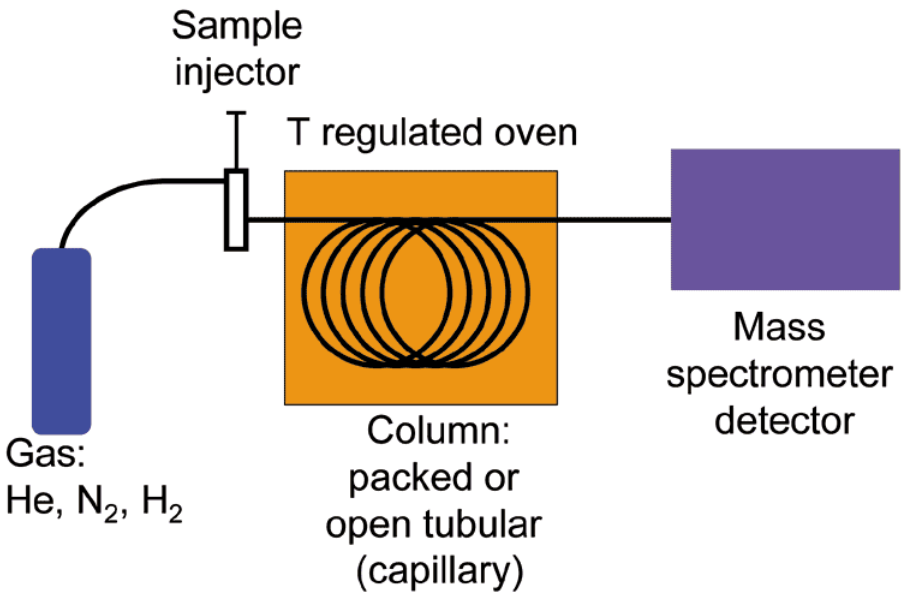


Information about structural moieties (fragments, neutral loss)

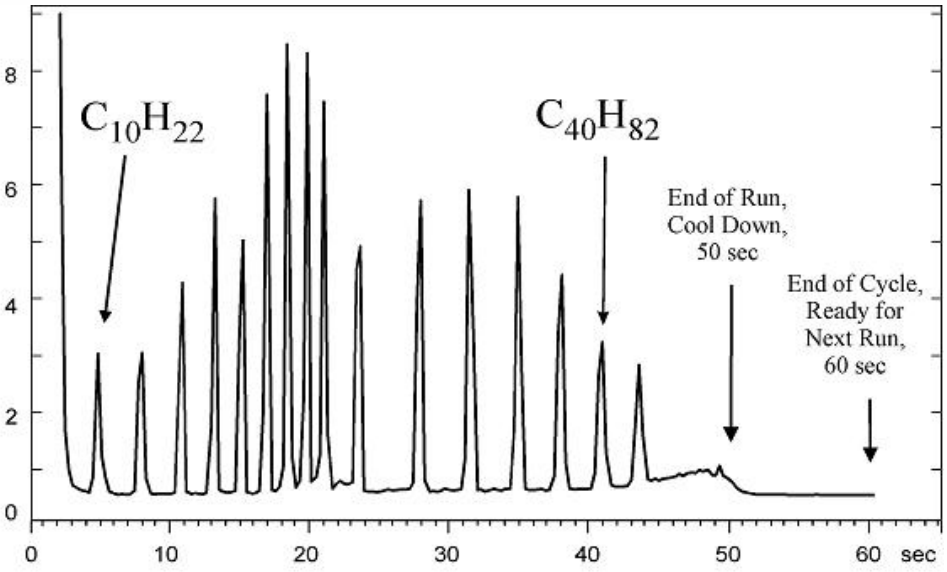
Information about the parent molecule

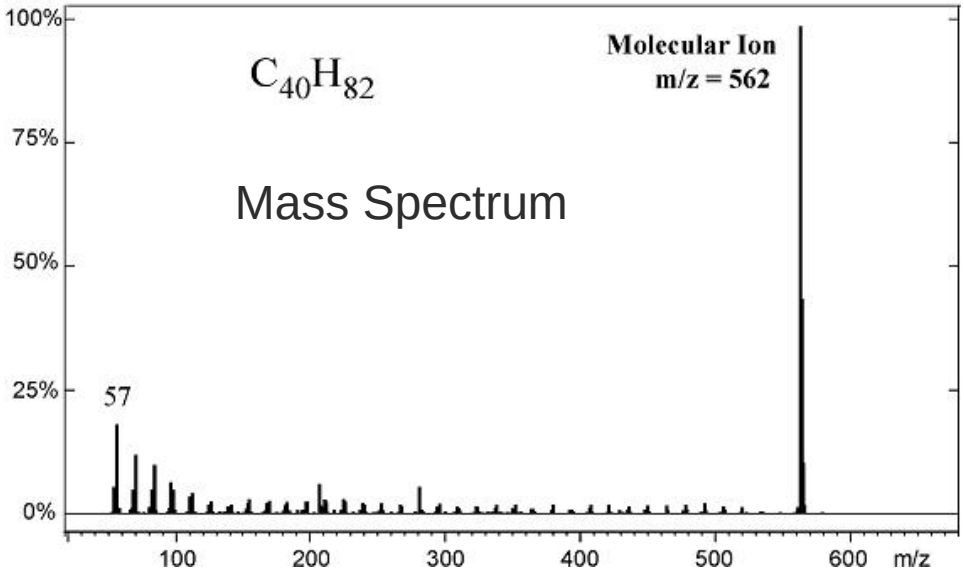
Source: wikipedia.org

GC/MS Schematic



Total Ion Chromatogram





Source: (l) wikipedia.org (K. Murray)
(r) http://www.avivanalytical.com/Fast_GC.aspx

Electron Ionization Mass Spectrometry (EI-MS)

Advantages

- Very reproducible spectra (intensity and fragments)
- Rich in structural information => “molecular fingerprint”
- Extensive libraries are available

Disadvantages

- Signal spread over all fragments (lack of sensitivity)
- Limited compound range => Derivatization often required
- Parent ion often low or missing

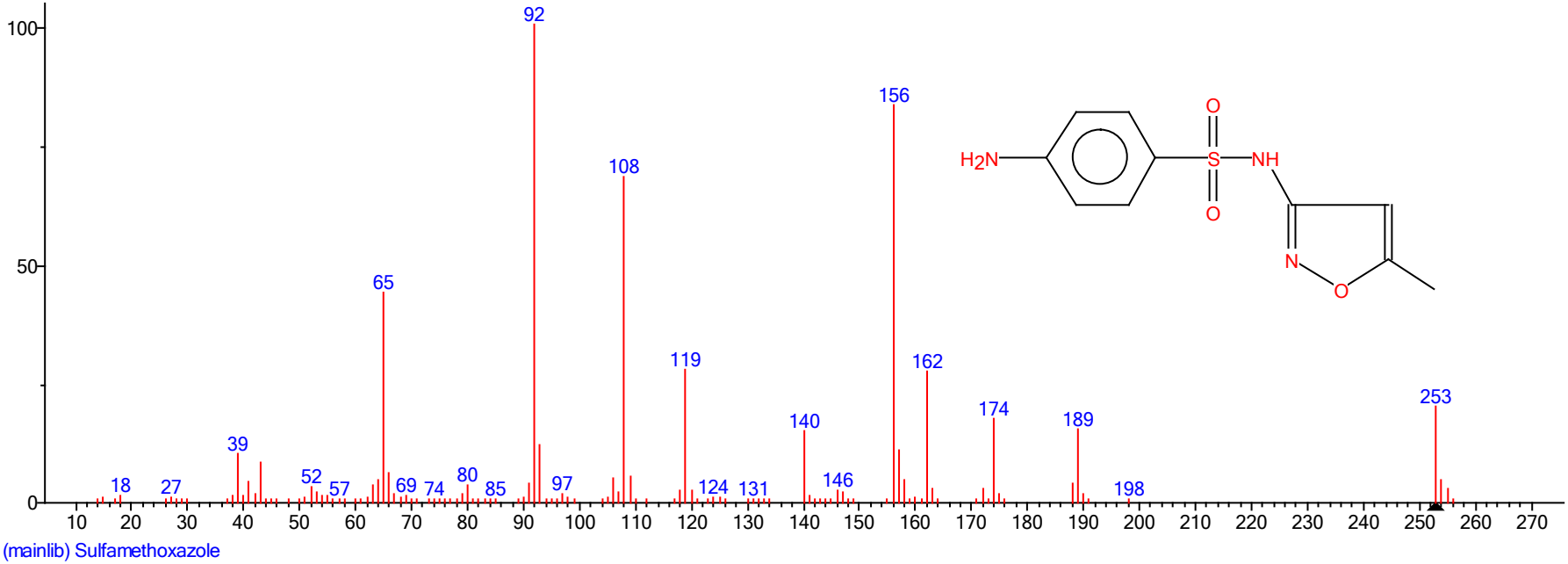


Image: NIST14 Library, MS Search

Spectral Interpretation: EI-MS

"all in one" spectrum

"M" peak – molecular weight and isotopic distribution of elements

Systematic fragmentation

=> Structural information

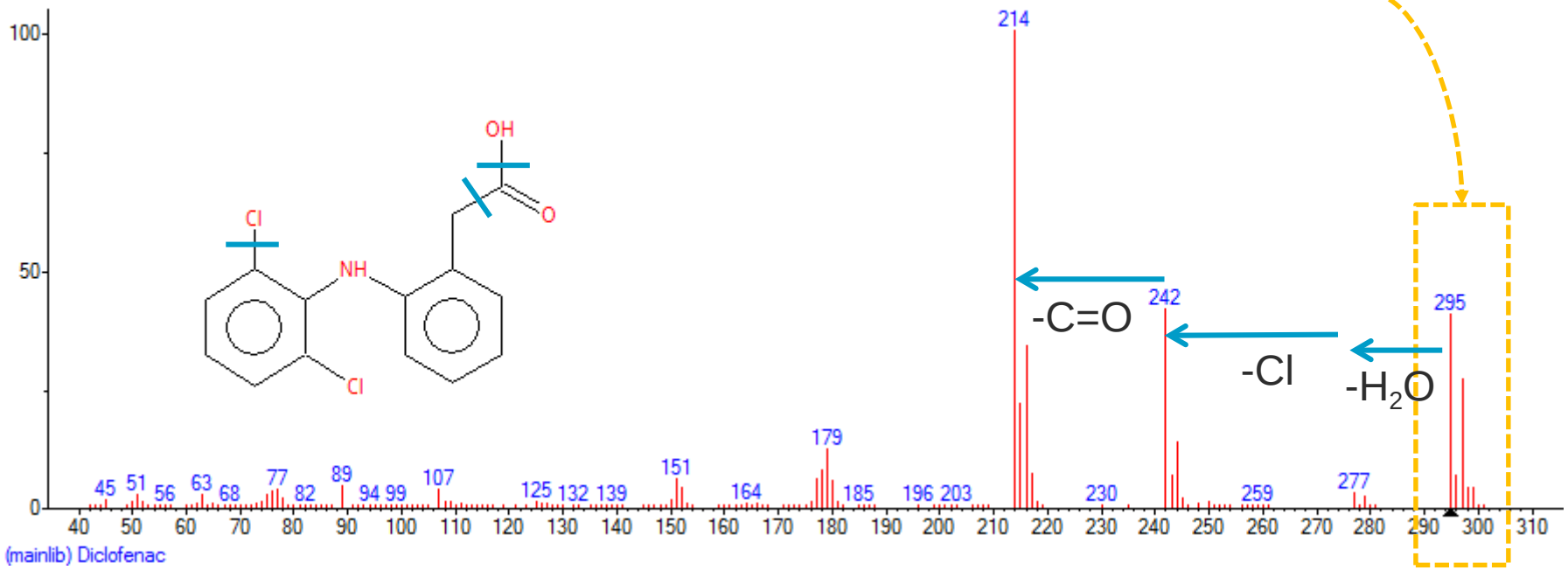
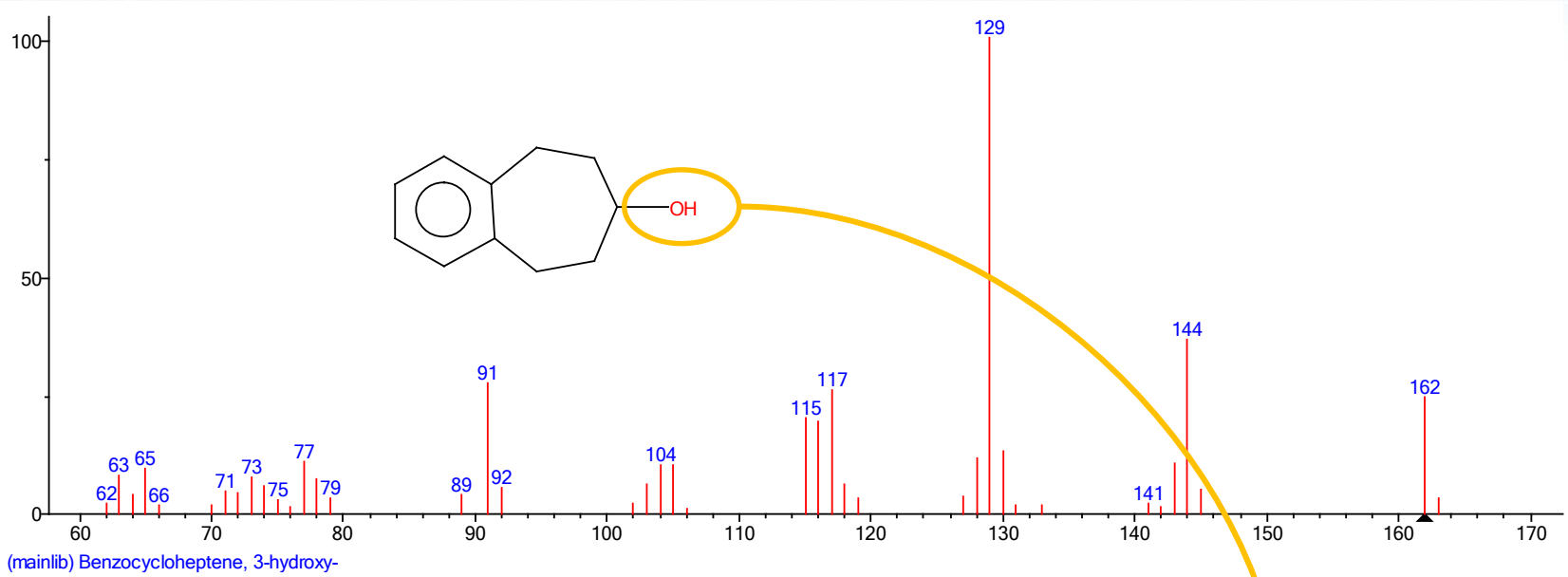
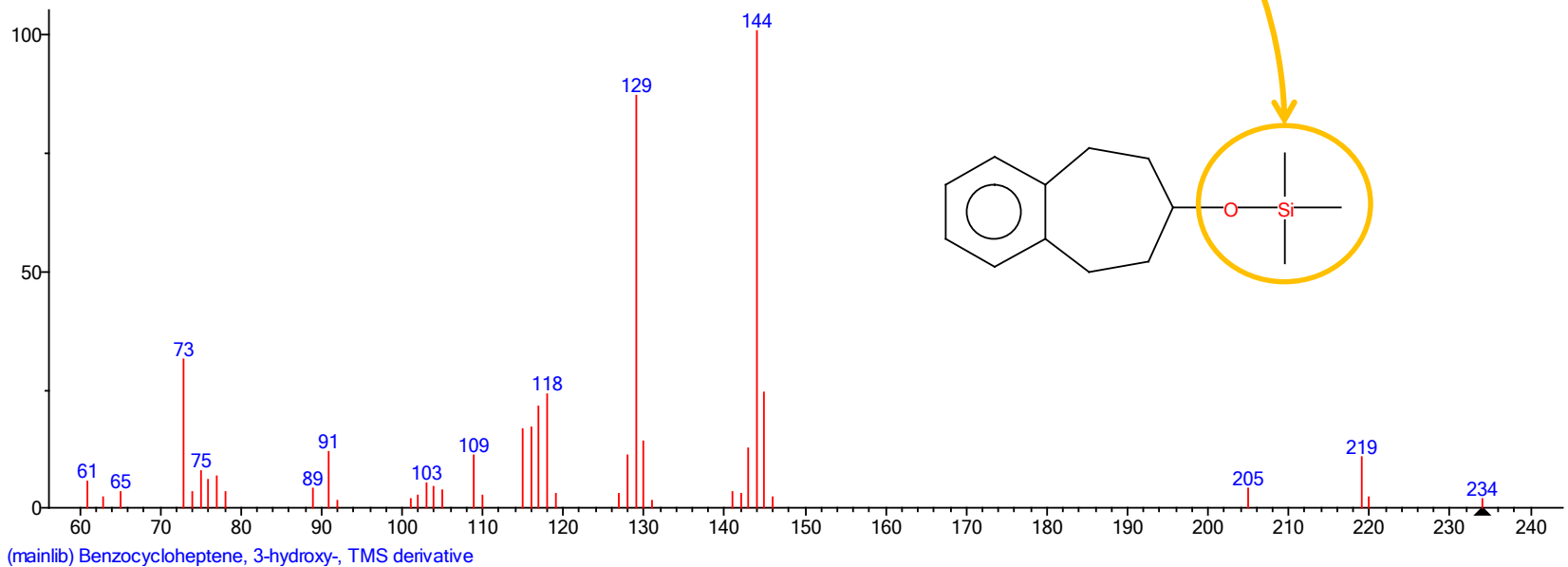


Image: NIST14 Library, MS Search

GC/EI-MS and Derivatization



(mainlib) Benzocycloheptene, 3-hydroxy-



(mainlib) Benzocycloheptene, 3-hydroxy-, TMS derivative

Images: NIST14 Library, MS Search

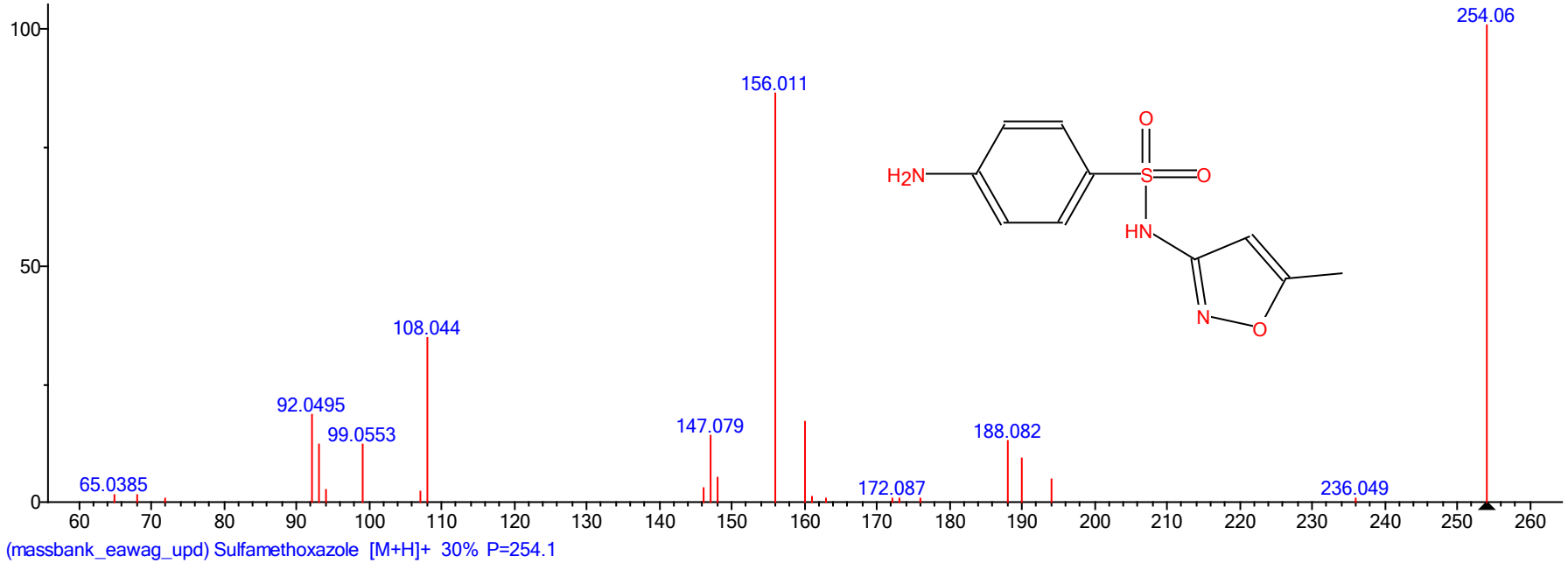
Soft Ionization MS (e.g. electrospray, ESI)

Advantages

- Broader coverage of compounds e.g. polar, thermolabile
- Soft ionization technique => parent mass is often available
- High resolution and accurate mass

Disadvantages

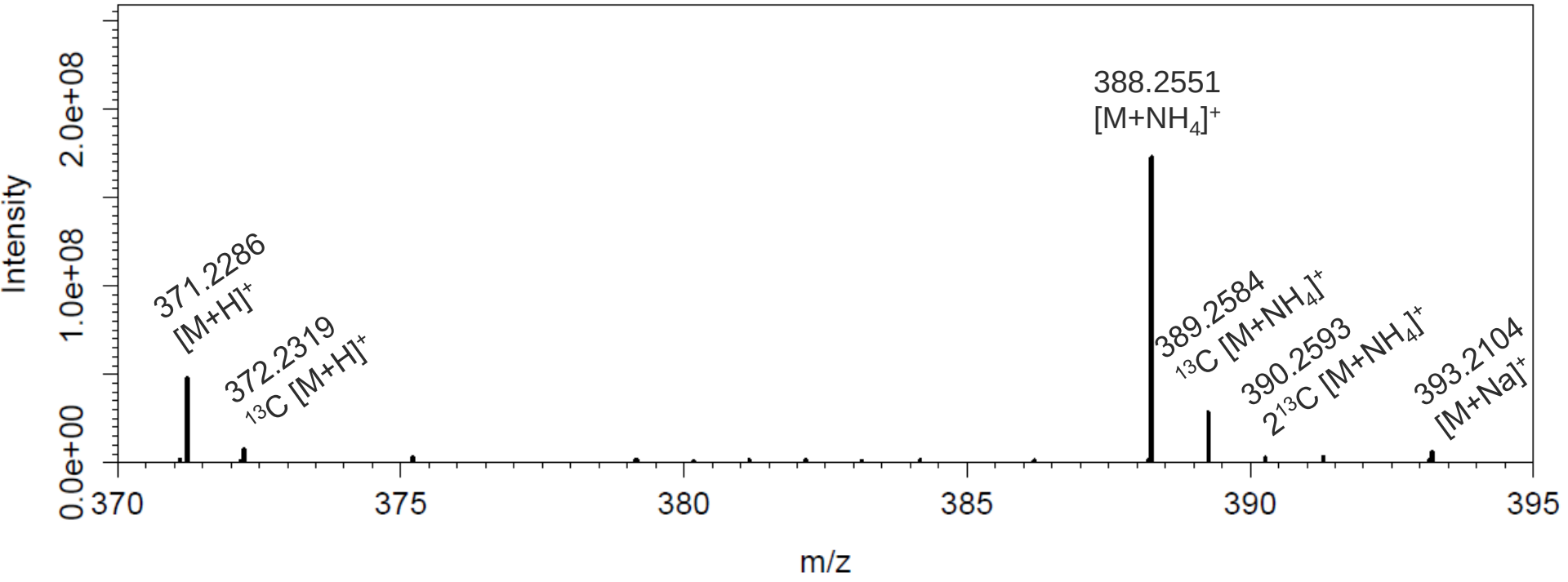
- Presence of adducts and in-source fragments increase complexity
- Poor reproducibility and fewer fragments => spectra less comparable
- Ion suppression



Images: NIST14 Library, MS Search

ESI: Presence of common adducts

$[M+H]^+$, $[M+NH4]^+$, $[M+Na]^+$, $[M+K]^+$ $[M-H]^-$, $[M+Cl]^-$, $[M+FA-H]^-$

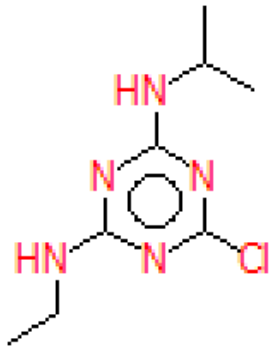


- Can hinder interpretation, but help calculate correct parent mass
- Presence and intensity depend on substance and conditions
- APCI: also see M^+ and M^-
- Can have multiply charged species

Role of Mass Spectral Libraries

… to put a structure to the spectrum (with a score)

#	Lib.	Match	R.Match	Prob. (%)	Name
1	M	999	999	98.3	Atrazine
2	R	928	933	98.3	Atrazine
3	R	915	916	98.3	Atrazine
4	R	901	933	98.3	Atrazine
5	R	870	884	98.3	Atrazine
6	M	670	741	0.98	1,3,5-Triazine-2,4-dia…
7	M	593	633	0.11	1H-Isoindole, 5,6-dichl..



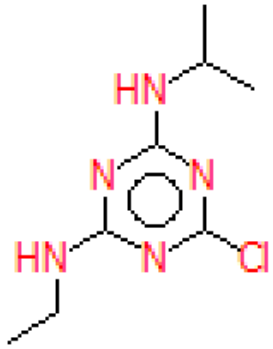
- Match and Reverse Match
  - Compares the *query* (*unknown*) spectrum with a *library entry*
  - Match: a direct match factor for the unknown and library
  - Reverse Match: ignores peaks in *unknown* absent in *library* spectrum
    - 999 = perfect match (e.g. query library spectrum against library)
    - >900: excellent match; 800-900: good match;
    - 700-800: fair match; <600: very poor match
    - 0 = absolutely no peaks in common

Images: NIST14 Library, MS Search

Role of Mass Spectral Libraries

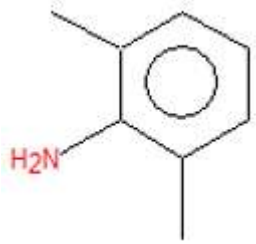
… to put a structure to the spectrum (with a score)

#	Lib.	Match	R.Match	Prob. (%)	Name
1	M	999	999	98.3	Atrazine
2	R	928	933	98.3	Atrazine
3	R	915	916	98.3	Atrazine
4	R	901	933	98.3	Atrazine
5	R	870	884	98.3	Atrazine
6	M	670	741	0.98	1,3,5-Triazine-2,4-dia…
7	M	593	633	0.11	1H-Isoindole, 5,6-dichl..



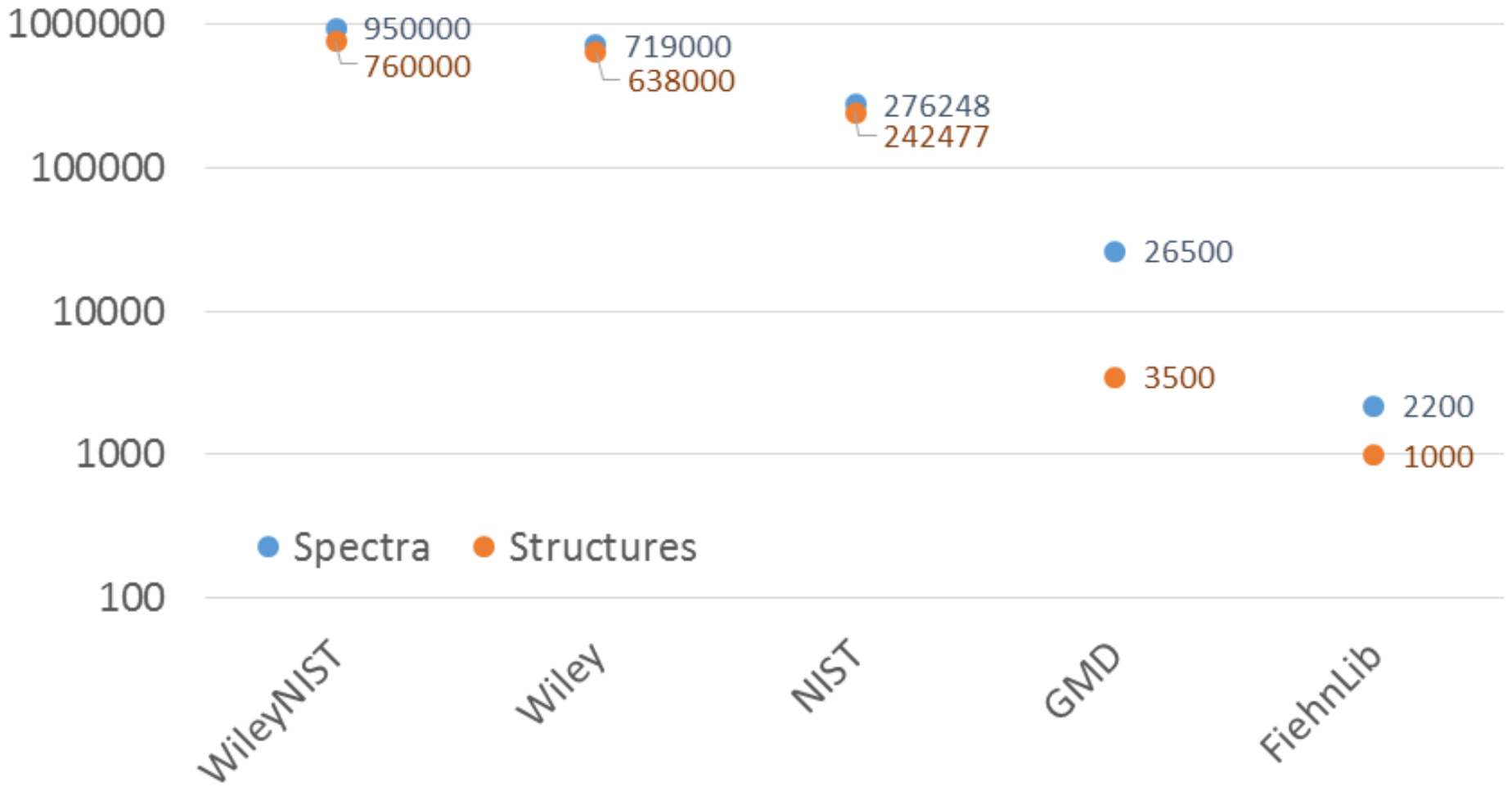
- Probability match
  - % “likelihood” that this spectrum is the correct answer
  - Very high for e.g. atrazine - ~60-70 % are common “cut-offs”
  - BUT: compounds with similar spectra in database can have low values

#	Lib.	Match	R.Match	Prob. (%)	Name
1	M	957	960	25.0	2,6-Xylidine
2	R	945	948	16.7	Benzenamine, 2,4-dimeth
3	R	945	947	16.7	Benzenamine, 2,5-dimeth
4	R	941	942	14.1	Benzenamine, 2,3-dimeth
5	M	940	943	14.1	Benzenamine, 2,3-dimeth
6	R	938	938	25.0	2,6-Xylidine
7	M	937	940	11.9	Benzenamine, 3,5-dimeth



Images: NIST14 Library, MS Search

GC-MS Libraries – Overview



http://www.sisweb.com/software.htm#2

Vinaixa *et al.*, 2016, *TrAC,* 78:23-35, DOI: 10.1016/j.trac.2015.09.005

GC-MS Libraries – Overview

Large commercial collections: NIST14 and Wiley

- NIST14 MS Library (EI-MS; 70 eV)
  - 276,248 spectra of 242,477 compounds
  - ***387,463 retention index values from 82,337 compounds***
- Wiley 10th MS Library (EI-MS; 70 eV)
  - 719,000 spectra of 638,000 compounds
- Combined NIST14 & Wiley 10th
  - >950,000 spectra of >760,000 compounds
  - Both compatible with most instrument manufacturers

http://www.sisweb.com/software.htm#2

Vinaixa *et al.*, 2016, *TrAC,* 78:23-35, DOI: 10.1016/j.trac.2015.09.005

GC-MS Libraries – Overview

Golm Metabolome Database (GMD): http://gmd.mpimp-golm.mpg.de/

ANALYTES

proprietary All

	name	MPIMP ID	isotopomer	formula	molecular mass
Details	NA241006	A241006	ambient		
Details	Mandelic acid, 4-hydroxy- (3TMS)	A177033	ambient		
Details	Pyroglutamic acid (2TMS)	A153002	ambient	C11H23NO3Si2	273.477
Details	Pentachlorophenol; GC-EI-TOF; MS; 1 TMS; BP:93	A189032	ambient		
Details	Ethylenediaminetetraacetic acid (4TMS)	A241009	ambient	C22H48N2O8Si4	580.968
Details	NA	A165007	ambient		
Details	[11321] acetylisatin 4 [17.11]	A186030	ambient		
Details	Glycine, 2-13C- (3TMS)	A133001	^{13}C	C10(13C)H29NO2Si3	

MS and RTI of pure reference substances and frequently observed mass spectral tags
(MST: mass spectrum linked to chromatographic retention) of yet unidentified metabolites

GC-MS Libraries – Overview*

FiehnLib: http://fiehnlab.ucdavis.edu/projects/FiehnLib/index.html

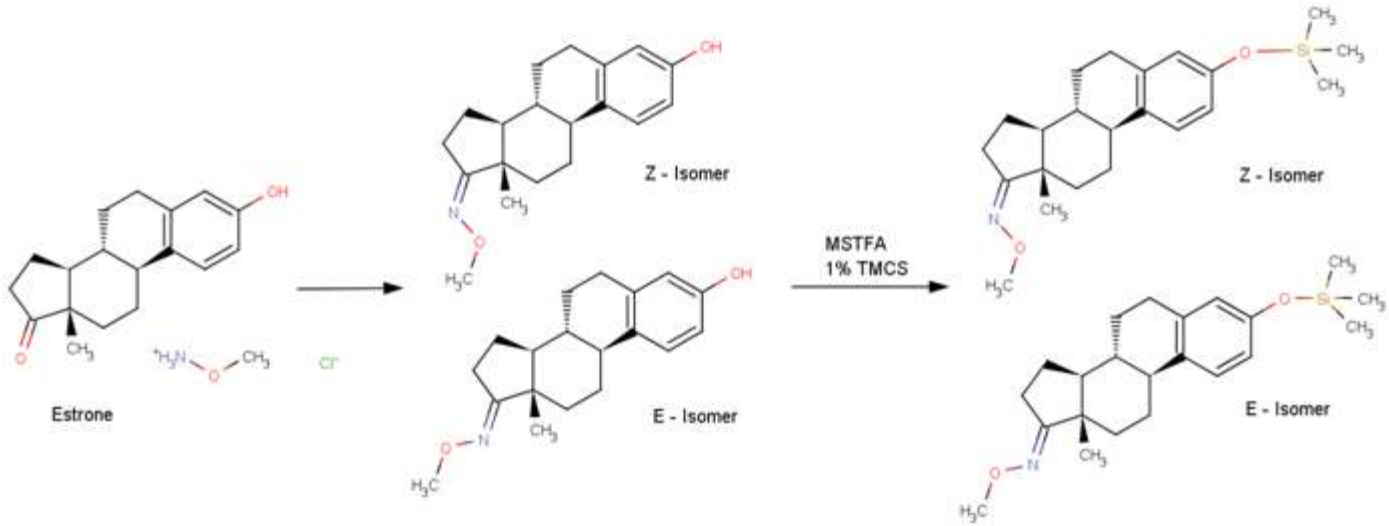
FiehnLib

Project:
FiehnLib - a mass spectral and retention index library for comprehensive metabolic profiling

The current libraries comprise over 1,000 identified metabolites that are currently screened by the Fiehn laboratory. We are continually extending the compound list and welcome compound donations or compound names with PubChem CIDs that are yet missing in our lists.

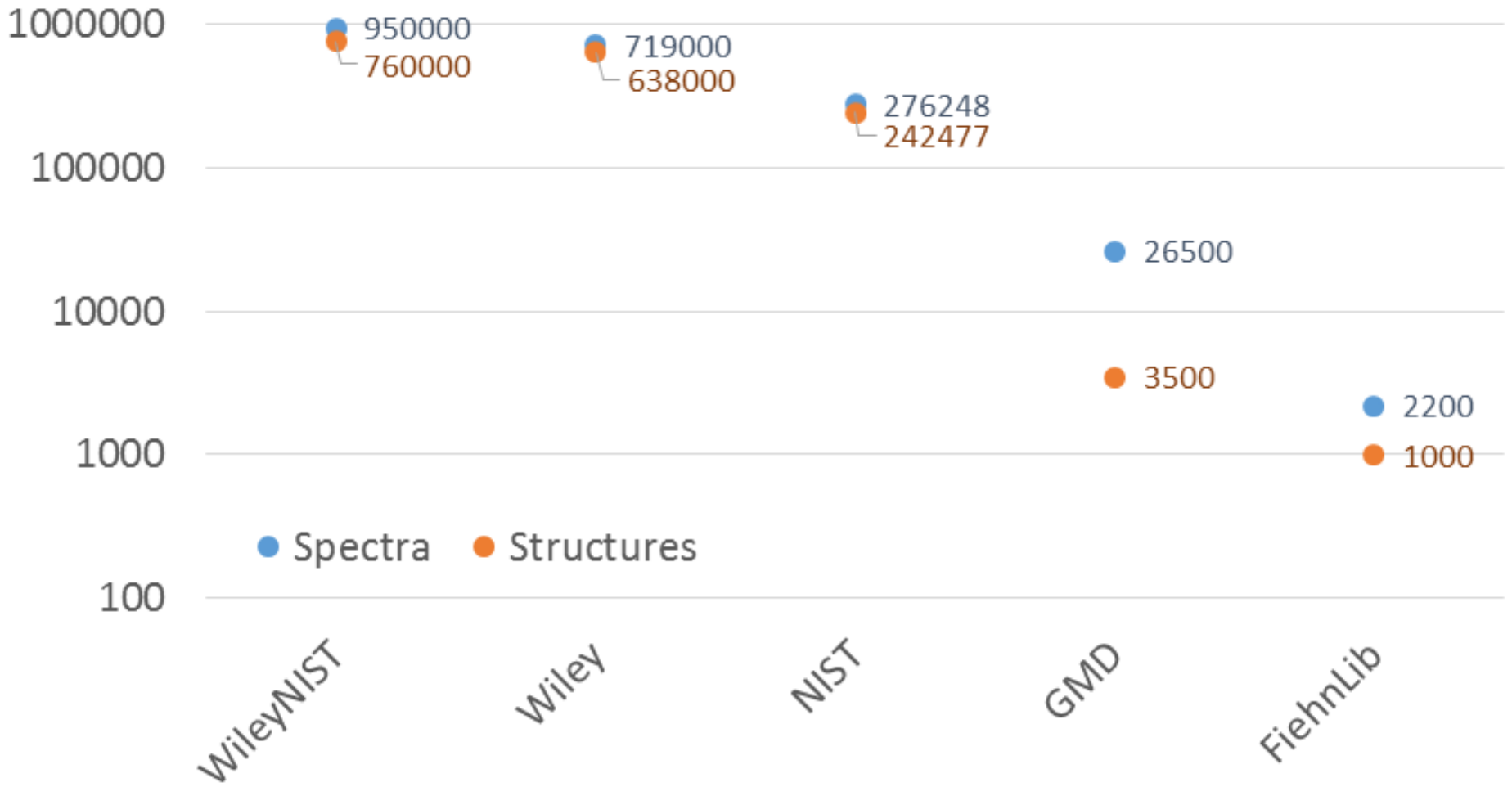
If you are interested in obtaining the Fiehn mass spectral libraries, you can purchase these for GC-quadrupole mass spectrometers from Agilent, and for GC-TOF mass spectrometers from Leco. (see links below)

You can also directly compare mass spectra by compound names, by database identifiers or even against your own GC/MS spectra by querying BinBase.



*HMDB, MassBank will be covered later; other collections are also available
Vinaixa *et al.*, 2016, *TrAC,* 78:23-35, DOI: 10.1016/j.trac.2015.09.005

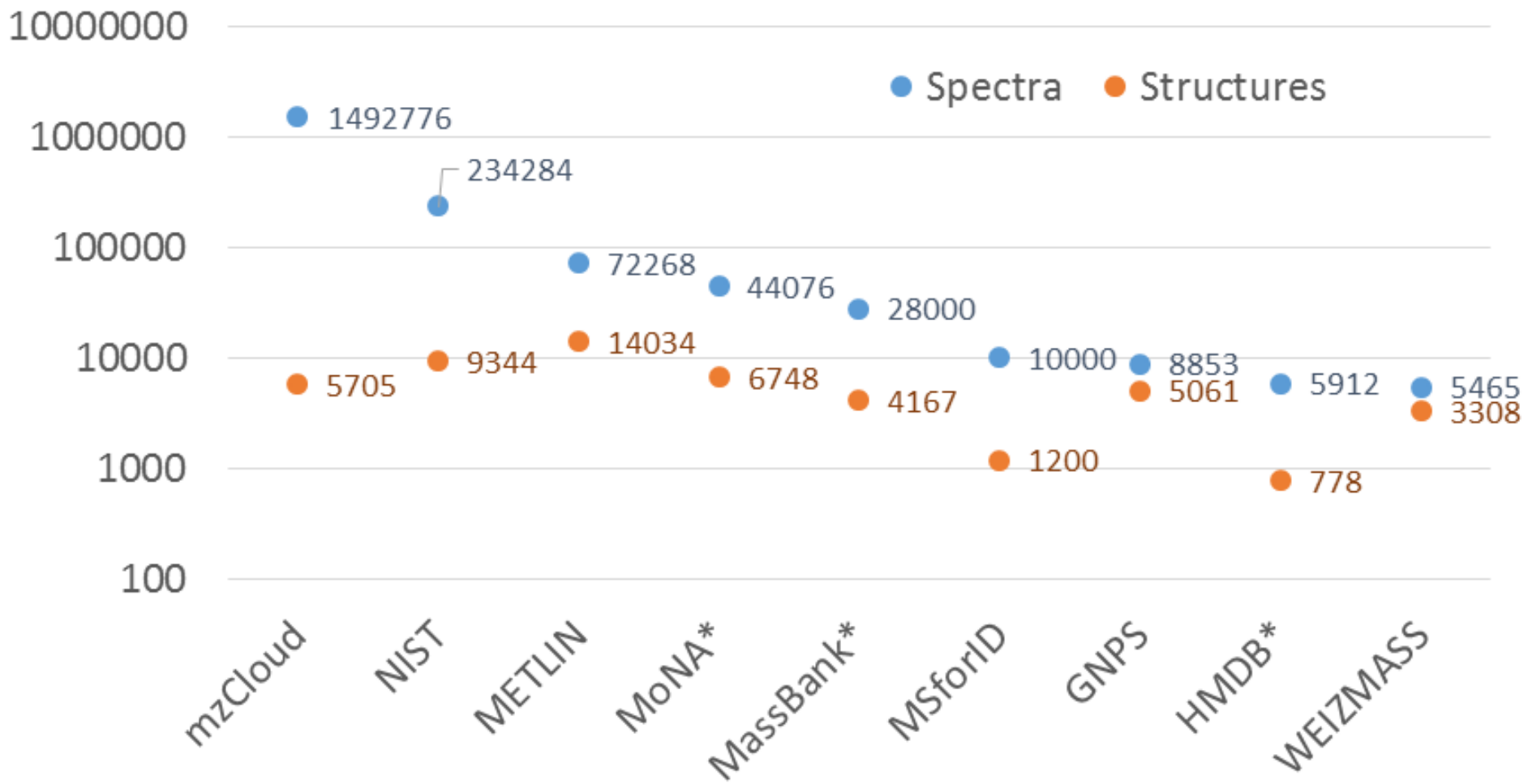
GC-MS Libraries – Overview [recap]



http://www.sisweb.com/software.htm#2

Vinaixa *et al.*, 2016, *TrAC,* 78:23-35, DOI: 10.1016/j.trac.2015.09.005

(LC-)MS/MS Libraries – Overview



*excluding GC-MS and *in silico* spectra

Various sources in addition to:
Vinaixa *et al.*, 2016, *TrAC,* 78:23-35, DOI: 10.1016/j.trac.2015.09.005

(LC-)MS/MS Libraries - Overview

What makes a good library?

Quality? Quantity?

Coverage? Structures?

Relevance? Services?

…it depends on your question!

METLIN: https://metlin.scripps.edu/

Smith *et al*. 2005. DOI: 10.1097/01.ftd.0000179845.53213.39

Statistics

- # Metabolites: 242,032
- # High Resolution MS/MS Spectra: 72,268
- # Metabolites w/ High Resolution MS/MS: 14,034

example | details...

Functionality

- **Single** & **Batch** Precursor Ion *(m/z)* searching
- **Single** & **Multiple** Fragment Ion *(m/z)* searching
- **Neutral Loss** searching
- *De Novo* Fragment Characterization

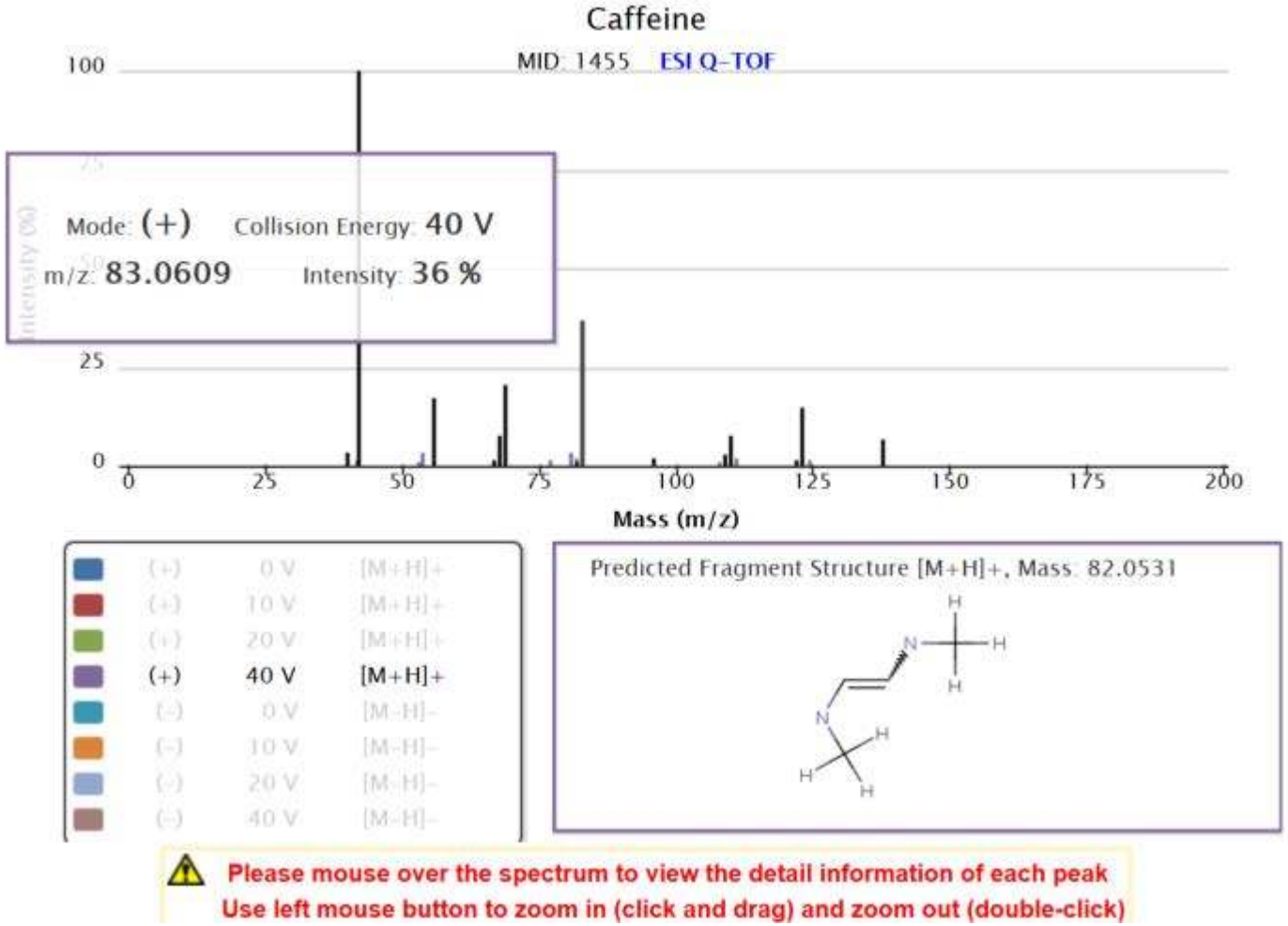
\+ One of largest collections

\+ Consistent format (Agilent 6510 Q-TOF, ESI+/-, 0, 10, 20, 40 eV)

\- Restricted access (“semi-open”); not downloadable

https://metlin.scripps.edu/

Interactive fragment fly-over



GNPS: http://gnps.ucsd.edu/ProteoSAFe/libraries.jsp

Library Name	View	Description	Releases
All GNPS Library Spectra	View	This contains all available spectra available publically for search at GNPS excluding third party libraries.	
GNPS Library	View	The GNPS library contains natural product compounds from user contributions.	Release 8 Release 7 Release 6 Release 5 Release 3 Release 2 Release 1
FDA Library Pt 1	View	Approved drug library from Selleckchem Part 1 run by Sirenas MD.	Release 1
FDA Library Pt 2	View	This set of reference compounds generated by the Dorrestein Lab contains 535 FDA natural product compounds complements part 1.	Release 1
PhytoChemical Library	View	140 compounds from the Prestwick Phytochemical Library generated by the Dorrestein Lab.	Release 1
NIH Clinical Collection 1	View	327 compounds from the NIH Clinical Collection 1 generated by the Dorrestein Lab. Further information about the collection can be found here .	Release 1
NIH Clinical Collection 2	View	164 compounds from the NIH Clinical Collection 2 generated by the Dorrestein Lab. Further information about the collection can be found here .	Release 1
NIH Natural Products Library	View	1256 compounds from the NIH Natural Products Library generated by the Dorrestein Lab. Further information about the collection can be found here .	Release 1
Pharmacologically Active Compounds in the NIH Small Molecule Repository	View	1460 compounds from the Pharmacologically Active Compounds in the NIH Small Molecule Repository generated by the Dorrestein Lab.	Release 1
Faulkner Legacy Library provided by Sirenas MD	View	127 compounds from the Faulkner natural product legacy library.	Release 1
EMBL Metabolomics Core Facility (EMBL MCF)	View	Standards run by EMBL Metabolomics Core Facility (EMBL MCF)	
Dereplicator Identified MS/MS Spectra	View	MS/MS spectra identified in GNPS Public data automatically by dereplicator tool. Searching various compound databases, including marinlit, etc. and best matching MS/MS spectra with significant p-values.	
Massbank Spectral Library (3rd Party)	View	ESI Positive MS/MS spectra from Massbank.	

GNPS: http://gnps.ucsd.edu/ProteoSAFe/libraries.jsp

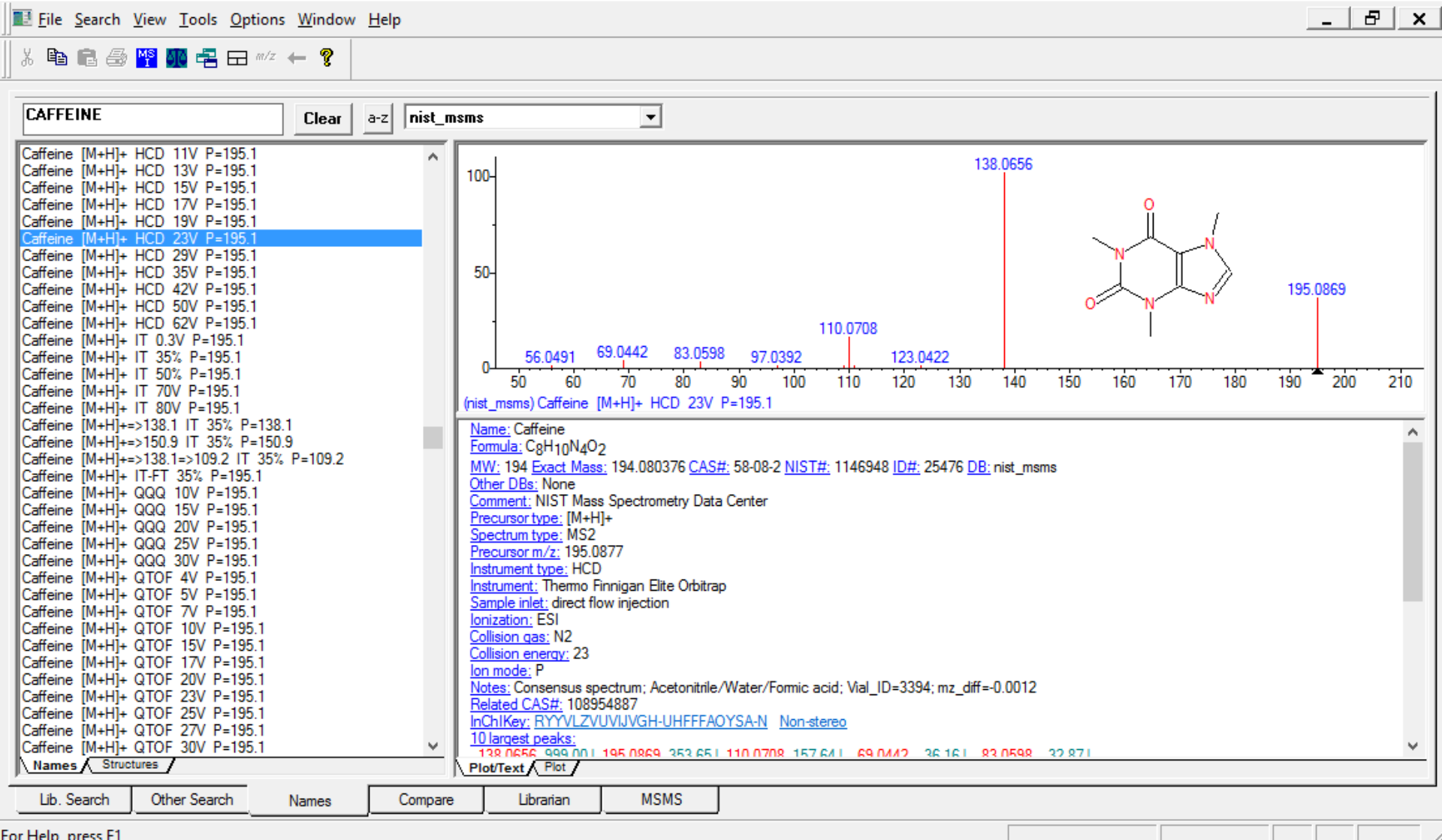
Library Name	View	Description	Releases
All GNPS Library Spectra	View	This contains all available spectra available publically for search at GNPS excluding third party libraries.	
GNPS Library	View	The GNPS library contains natural product compounds from user contributions.	Release 8 Release 7 Release 6 Release 5 Release 3 Release 2 Release 1

+ Fully open, downloadable, individual collections
+ Users can contribute / upload their own
+ MS/MS of adducts, unidentified structures
+ Over 8,853 MS/MS spectra and constantly growing
+ Comes with repository, search and live update functions
+ Data exploration: e.g. http://goo.gl/NmO4tx and http://goo.gl/7sY9Pf
- Very few negative mode spectra
- Limited/incorrect information about compounds/spectra
- No spectral clean up – gold is not really “gold”

Wang, M. *et al*. 2016, *Nat. Biotech*., 34: 828-837. DOI: 10.1038/nbt.3597

NIST14 MS/MS Library

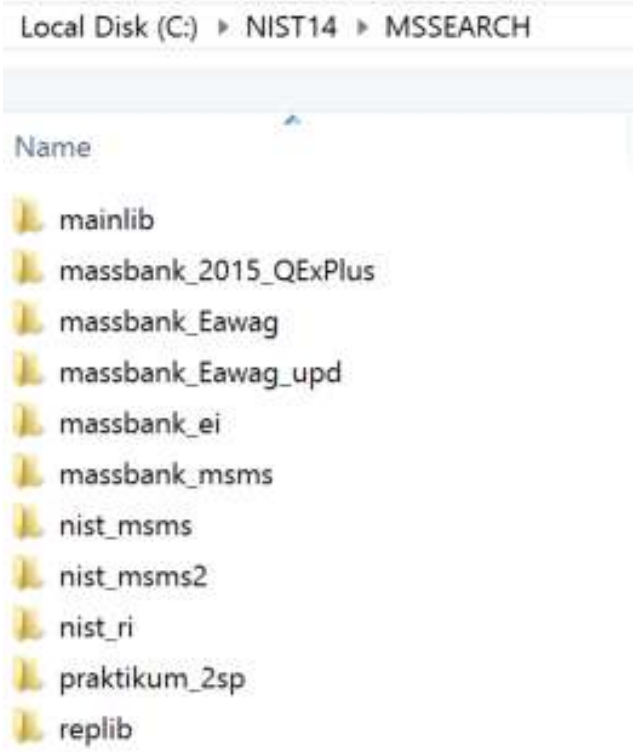
http://www.sisweb.com/software/nist-msms.htm



NIST14 MS/MS Library

http://www.sisweb.com/software/nist-msms.htm

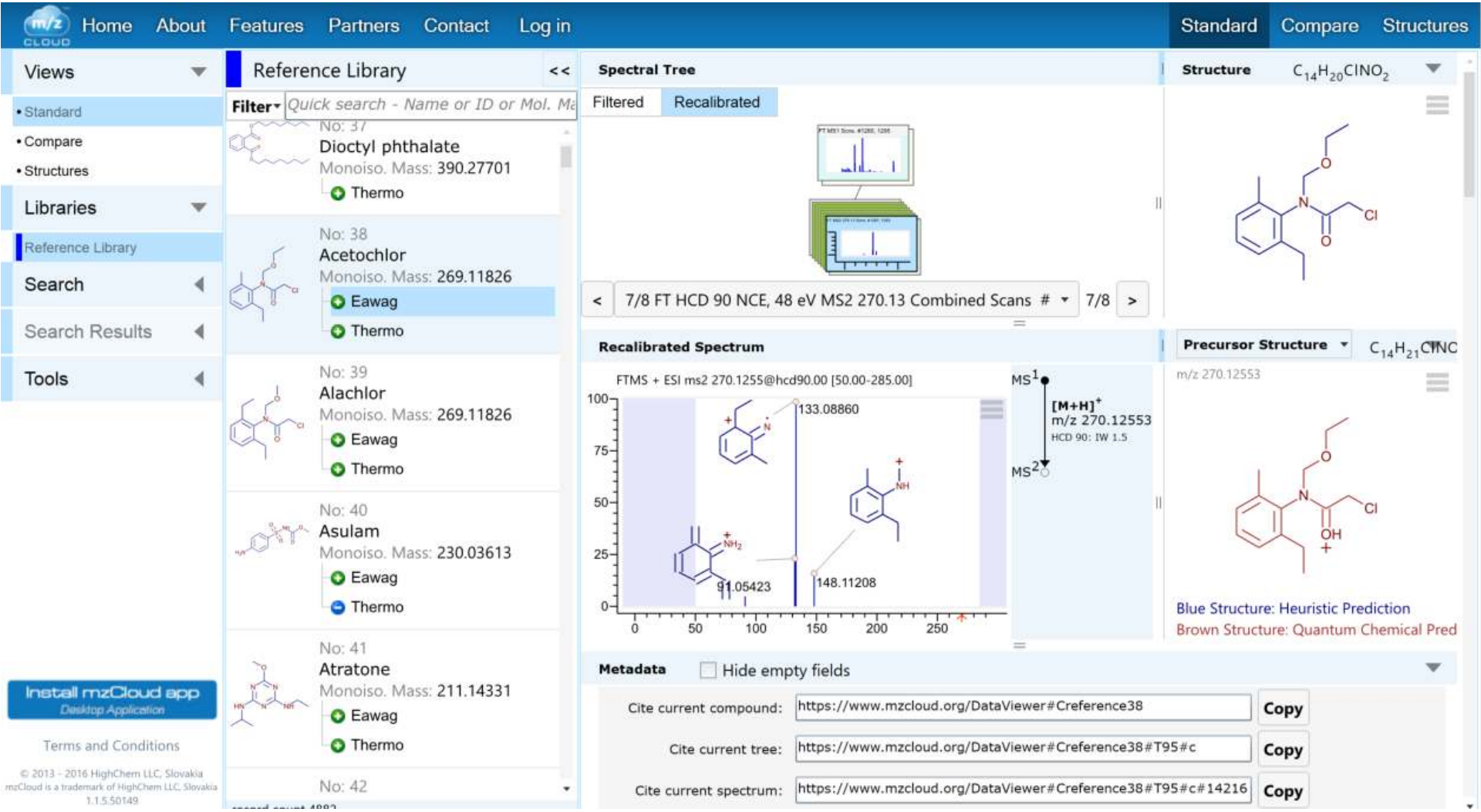
+ Very large collection: 234,284 ESI MS/MS spectra of 9,344 substances
+ Very well curated (Yang et al. 2014, DOI: 10.1021/ac500711m)
+ MS/MS of many adducts
+ Spectra from multiple instruments (low and high resolution)
+ Users can include their own libraries quite easily if in NIST format
+ Integrated into most vendor software/workflows
+- Offline functionality
- Commercial license (but great investment!)
- Lack of external identifiers
- Difficult to integrate into open workflows



NIST-related reading:
Yang *et al*. 2014, DOI: 10.1021/ac500711m
Stein, 2012, DOI: 10.1021/ac301205z

m/z Cloud: https://www.mzcloud.org/

HighChem, Bratislava



m/z Cloud: https://www.mzcloud.org/

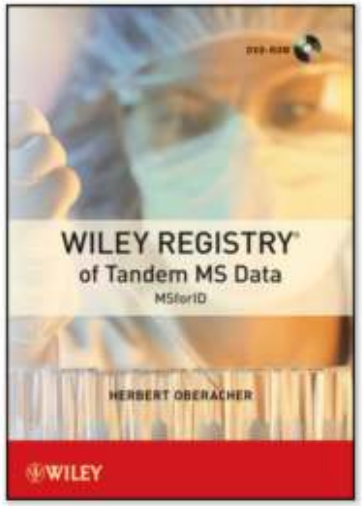
More stats: https://www.mzcloud.org/Stats

+ By far largest number of spectra (but fewer compounds than NIST)
+ Very well curated – raw and processed data both available
+ Spectral trees and MS/MS of many adducts
+ High resolution data (Orbitrap)
+ Annotated spectra – formulas, structures, quantum chemical models
+ Growing and updated constantly – users can contribute
- Semi-open – not downloadable or batchable
- Commercial license for extended functionality
- Extended functionality (at this stage) limited to Thermo users
- Difficult to integrate into open workflows

Wiley: MSforID (Oberacher)

http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118037448.html

Oberacher and Arnhard, 2016, TrAC, DOI: 10.1016/j.trac.2015.12.019



Read an Excerpt

SOFTWARE

Wiley Registry of Tandem Mass Spectral Data, MS for ID

Herbert Oberacher

ISBN: 978-1-118-03744-7

March 2012

+ High resolution, curated collection
+ Specialist substances: drugs, pharmaceuticals, pesticides
- Relatively small collection
- Commercial license

Description

High Mass-Accuracy LC-MSMS Lib...

The Wiley Registry of Tandem Mass Spectral Data: MS for ID contains 10,000 positive and negative mode spectra of over 1200 compounds of interest for forensics, toxicology, and pathology. Areas covered include:

- Illicit drugs
- Pharmaceutical compounds
- Pesticides
- Other small bioorganic molecules

Purchase Options

Software £1,800.00 €2,430.00 BUY

Prices are valid for United Kingdom. Change location to view local pricing and availability.

Permissions

- Request permission to reuse content from this title

Technical Support

- Get technical support for this product

Join An E-mail List

Learn about the latest products, events, offers and content.

Your Email Address

Wishart Lab Collections

http://www.wishartlab.com/web_servers

hmdb
The Human Metabolome Database

http://www.hmdb.ca/

DrugBank
Drug & Drug Target Database

http://www.drugbank.ca/

FooDB

http://foodb.ca/

http://www.t3db.ca/

Wishart Lab Collections

Identification	
Name	Caffeine
Accession Number	DB00201 (APRD00673)
Type	Small Molecule

Spectra

Mass Spec (NIST)	Download (8.61 KB)

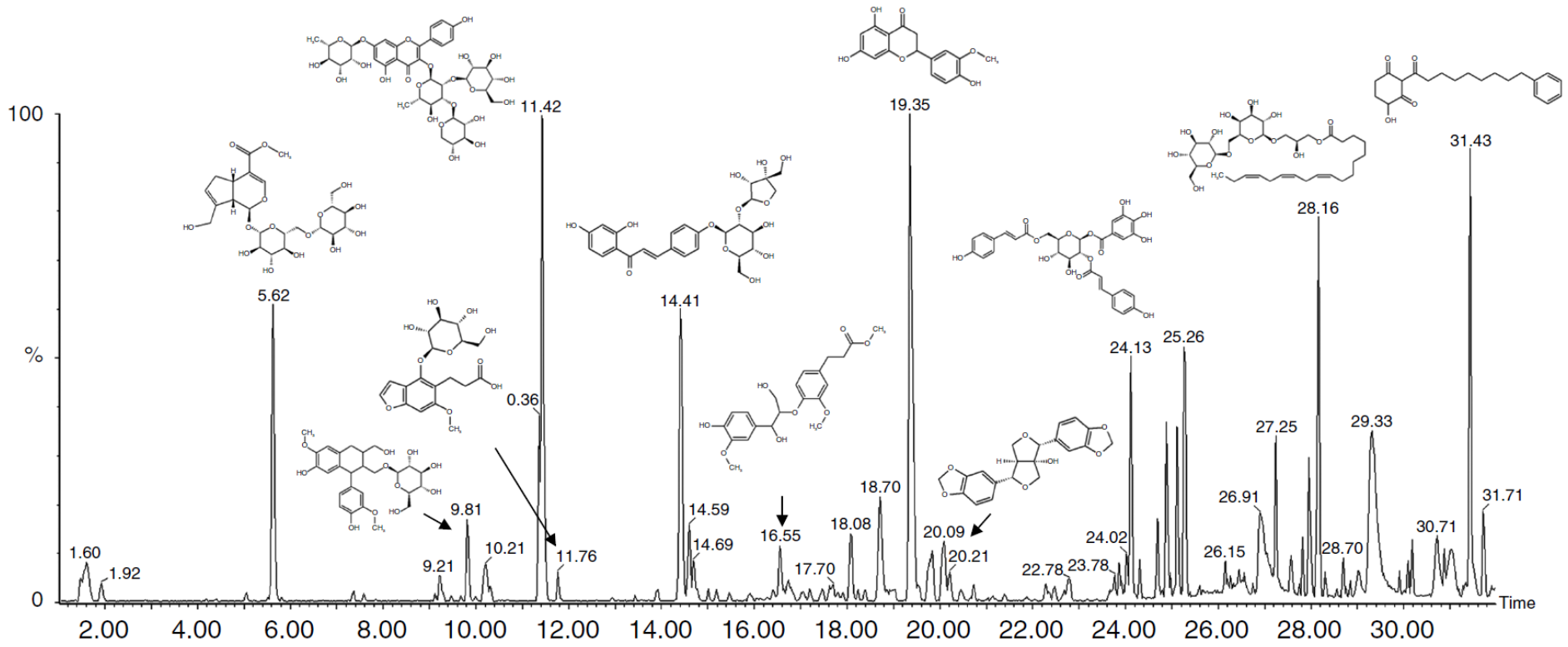
Spectra

Spectrum Type	Description	Splash Key	
GC-MS	GC-MS Spectrum - GC-EI-TOF (Pegasus III TOF-MS system, Leco; GC 6890, Agilent Technologies) (0 TMS)	splash10-0536-3900000000-a9e112713ffae6dabdaa	View in MoNA
GC-MS	GC-MS Spectrum - GC-EI-TOF (Pegasus III TOF-MS system, Leco; GC 6890, Agilent Technologies) (0 TMS)	splash10-0536-2900000000-8cdcd005b2e7622a02a3	View in MoNA
		fddb240696073	View in MoNA
		44cee5d9de78d4	View in MoNA
		2b279a72e34822	View in MoNA
		2b2c709e2e18a9	View in MoNA
		de16ad91c79fe0	View in MoNA
		2ac36c8f25a500	View in MoNA
		0441aa2c1c3ac68	View in MoNA
LC-MS/MS		eda9a9a4094715	View in MoNA
Predicted LC-MS/MS	Predicted LC-MS/MS Spectrum - 40V, Positive	splash10-01p9-3900000000-141118b1a93b4d48b2d4	View in MoNA
Predicted LC-MS/MS	Predicted LC-MS/MS Spectrum - 10V, Negative	splash10-0006-0900000000-c8bd8cccb7dd6c66bb42	View in MoNA
Predicted LC-MS/MS	Predicted LC-MS/MS Spectrum - 20V, Negative	splash10-0006-0900000000-2c8c1760161edd358026	View in MoNA

+ MS or MS/MS data on over 9,500 substances
+ Downloadable collections
+ Spectra from multiple instruments (low and high resolution)
+ Including predicted spectra as well
+ Cross-linked to other resources
+- Mixed collision energies and instrument types
- Fragmented collections – difficult to see what is where

WEIZMASS: Plant metabolite HR-MS/MS

Shahaf *et al*. 2016 *Nat. Comm*. 7:12423. DOI: 10.1038/ncomms12423

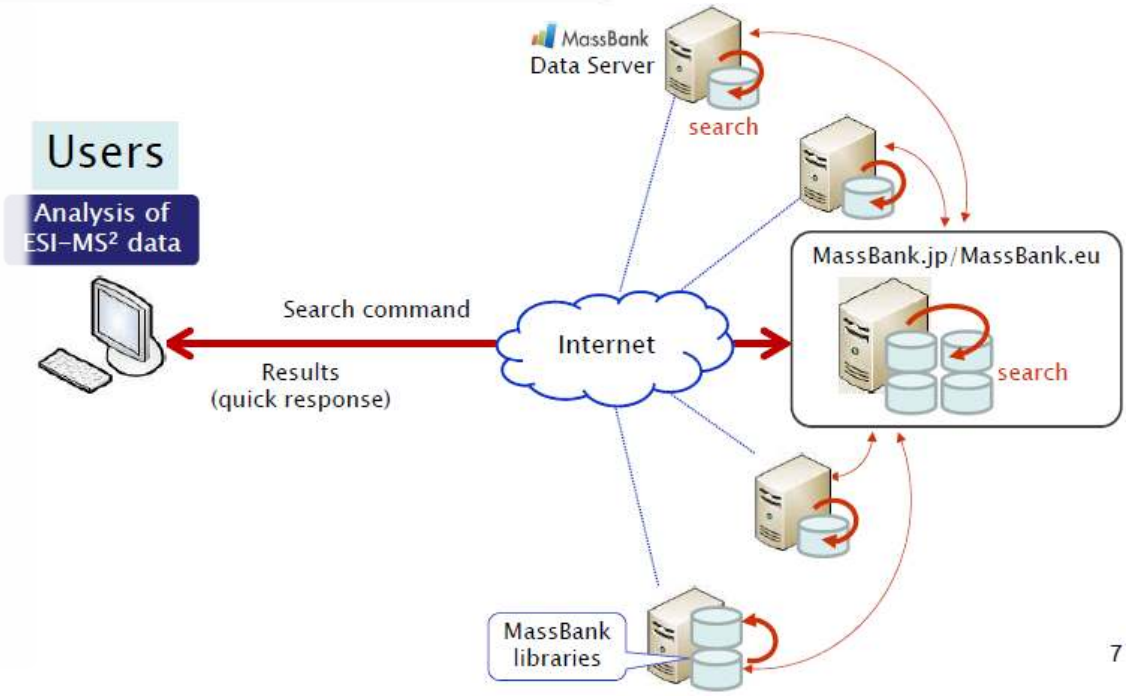


+ Large reference standard collection of 3,308 plant metabolites
+ Consistent MS^E acquisition; spectral annotation
- Available for academic use on request only

MassBank: The original Japanese server

www.massbank.jp, Horai *et al.* 2010, DOI: 10.1002/jms.1777

+ 54,349 spectra
+ >4,100 substances
+ Public repository
+ User contributions
+ Downloadable OpenData
+ Spectra from multiple instruments (EI, ESI, low and high resolution)
- Mixed quality of spectra
- Services out-of-date



MassBank Consortium

Exchanging spectra around the world

European MassBank

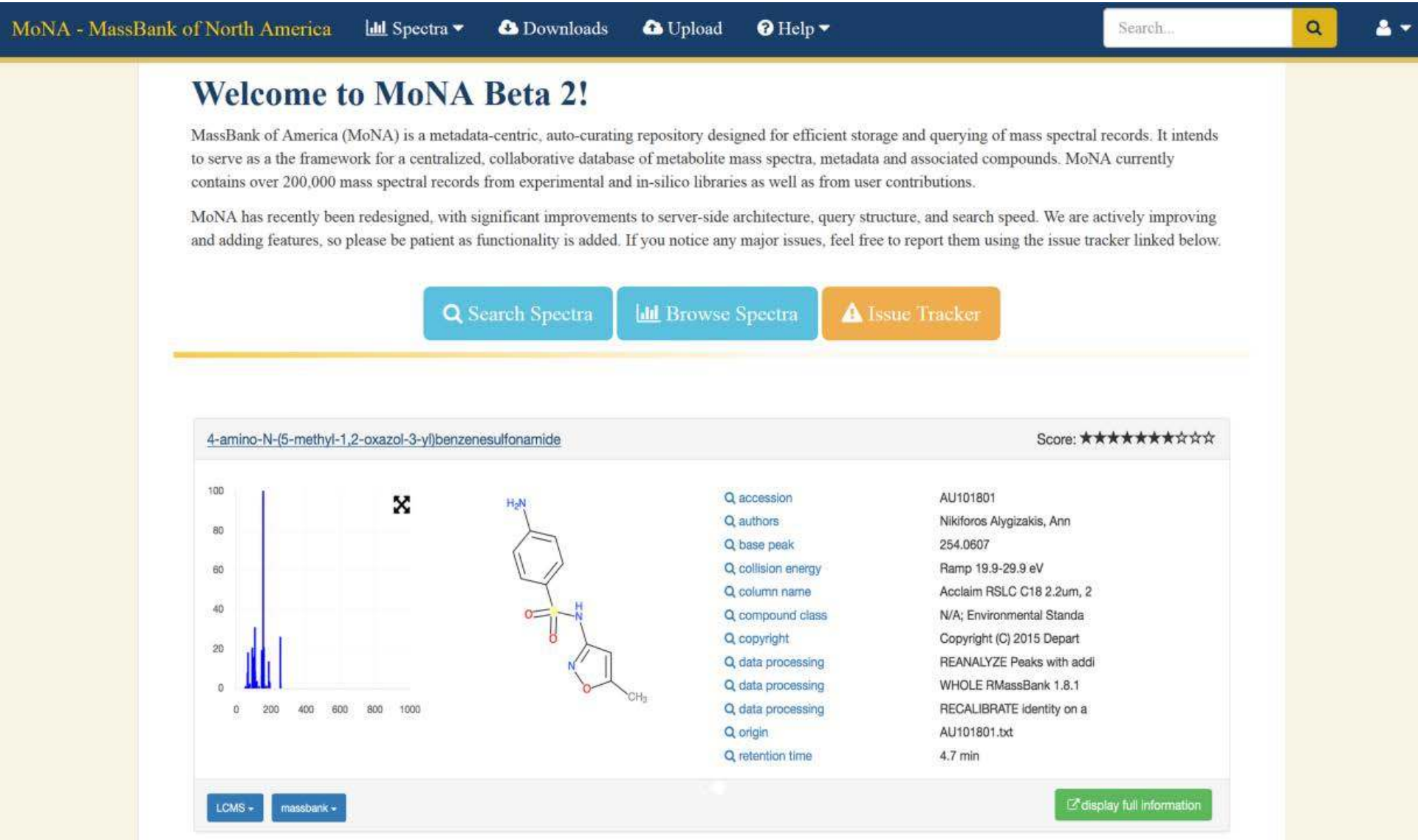
http://massbank.eu/MassBank

- MassBank.EU was founded late 2012, hosted at UFZ, Leipzig, Germany
  - 16,017 MS/MS spectra; 1,232 substances from NORMAN members
  - ***Tentative/unknown/literature*** spectra on massbank.eu (not massbank.jp)

Athens Univ. (1,492)	Boise State Univ. (4)	Chubu Univ. (2,563)
Eawag (10,668)	Eawag Additional Specs (620)	Env Anal Chem, U Tuebingen (116)
European MassBank Server (NORMAN MassBank) (0)	Fac. Eng, Univ. Tokyo (12,379)	Fiocruz (800)
Fukuyama Univ. (340)	GL Sciences Inc. (174)	IPB Halle (528)
JEOL Ltd. (45)	Kazusa (273)	Keio Univ. (10,124)
Kyoto Univ. (184)	Literature Specs (39)	MPI for Chemical Ecology (691)
MSSJ (34)	MetaboLights (58)	Metabolon (149)
NAIST (671)	Nihon Univ. (488)	Osaka MCHRI (20)
Osaka Univ. (449)	PFOS research group (413)	RIKEN (1,718)
Tottori Univ. (16)	UFZ (2,758)	UFZ Additional Specs (107)
UOEH (35)	UPAO (12)	Univ. Connecticut (510)
Univ. Toyama (253)	Washington State Univ. (2,626)	Waters (2,992)

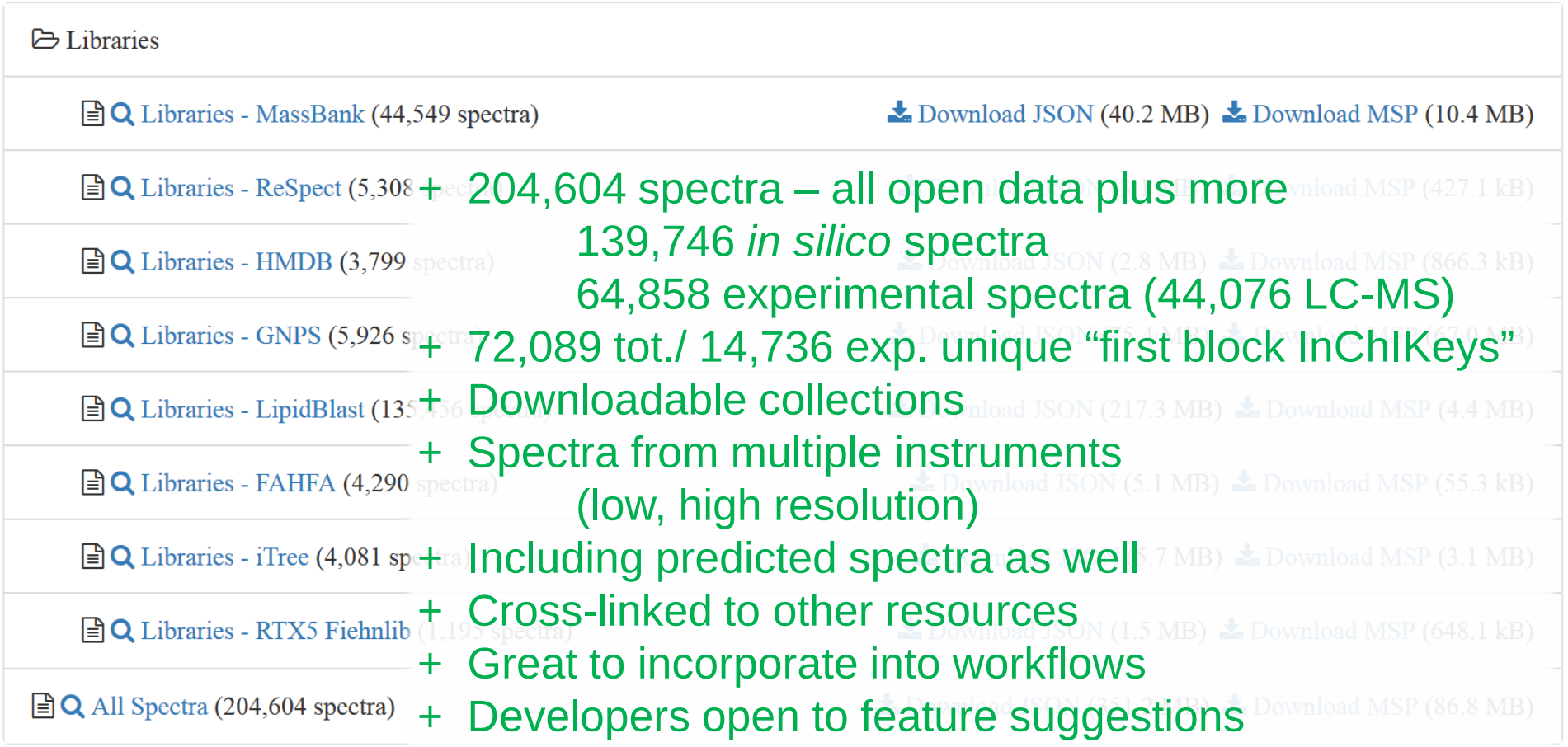
MassBank of North America

http://mona.fiehnlab.ucdavis.edu/



MassBank of North America

Largest collection of (Fully Downloadable) Open Mass Spec Data



+ 204,604 spectra – all open data plus more
  139,746 *in silico* spectra
  64,858 experimental spectra (44,076 LC-MS)
+ 72,089 tot./ 14,736 exp. unique “first block InChIKeys”
+ Downloadable collections
+ Spectra from multiple instruments
  (low, high resolution)
+ Including predicted spectra as well
+ Cross-linked to other resources
+ Great to incorporate into workflows
+ Developers open to feature suggestions
+- In development – not static or fully functional
- Format: great for informaticians, less ideal for users
- Automatic curation/annotations still under development

What do the (environmental) users use?

What do the (environmental) users use?

Schymanski *et al* 2015. DOI: 10.1007/s00216-015-8681-7

Database/Library Name	Total Compounds	Compounds with Spectra
ChemSpider **[35]**	32 million	
DAIOS **[49,50]**	1,404	>1,000[a]
PubChem **[48]**	63,105,228	
STOFF-IDENT **[38]**	8,000[b]	
MassBank **[51,52]**		5,000
mzCloud **[53]**		1,956
NIST MS 2011 **[11,54]**		212,961[c]
NIST MS/MS 2011 **[11,54]**		4,628
Wiley Registry of Mass Spectral Data 7th Edition [12]		289,000
ABSciex Meta Library		2,381
Agilent Broecker, Herre & Pragst toxic/forensics	7,509[c]	~2,500
Agilent Pesticide Library	1,664	~700[c]
Agilent Synthetic Substance Library	23,053	n/a
Agilent METLIN database	64,092	8,040
Bruker Pesticide Screener		700[d]
Thermo Environmental Food Safety (EFS) with RT		454[dp]; 447[p]; 90[dn]; 278[n]
Thermo toxicology		618[p]; 36[n]
Waters database with RT		730[de]
In-house Libraries without spectra (two participants)	2,000; 1,600	
In-house Libraries with spectra (two participants)		526[d]; 63[d]
In-house Libraries with spectra for some substances	2,200[d]	835[ad]
	7,815	1500[ap]; 500[an]
	3,000	350[d]
Surfactant List [3]	394	

Compound DBs

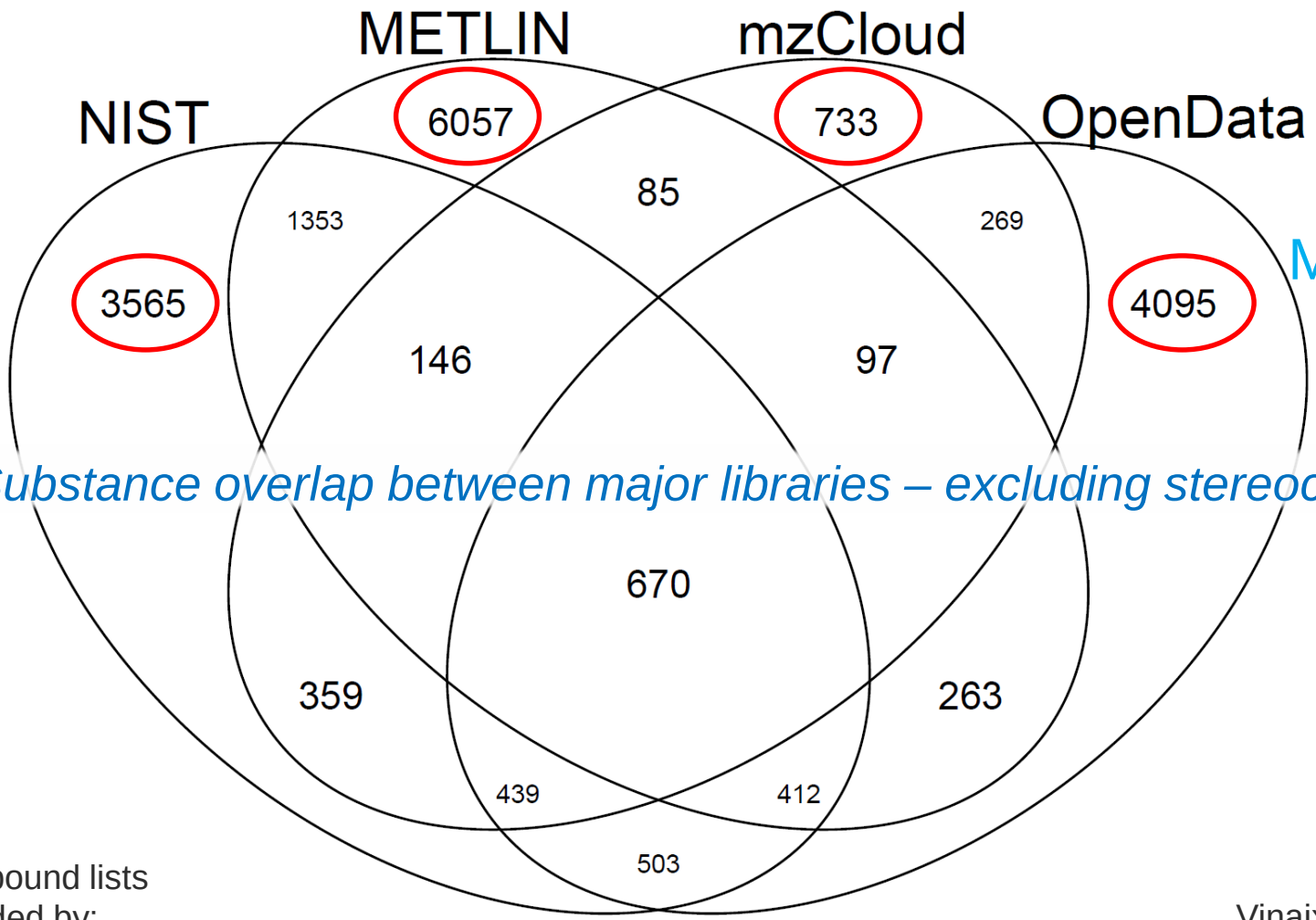
Spectral Libraries

Vendor-Specific Libraries

In-house lists/libraries

Do we need all these MS/MS resources?

…at this stage – YES!



OpenData = HMDB, GNPS, MassBank, ReSpect

Substance overlap between major libraries – excluding stereochemistry

Compound lists provided by:
S. Stein, R. Mistrik, Agilent

Vinaixa *et al.* 2016, *TrAC,*
DOI: 10.1016/j.trac.2015.09.005

Improving Library Coverage

RMassBank

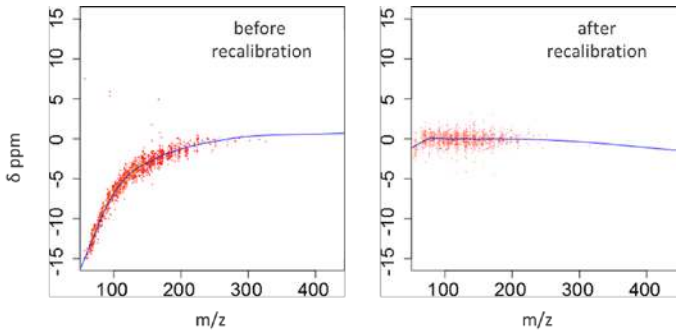
LC-MS/MS raw data

compound list: SMILES, name

online resources: CTS, CACTUS

Automatic MS and MS/MS Recalibration and Clean-up
Remove interfering peaks

Spectral Annotation with
- Experimental Details
- Compound Information



MassBank records

structure files

npmrman

MassBank.eu

https://github.com/MassBank/RMassBank/
http://bioconductor.org/packages/RMassBank/

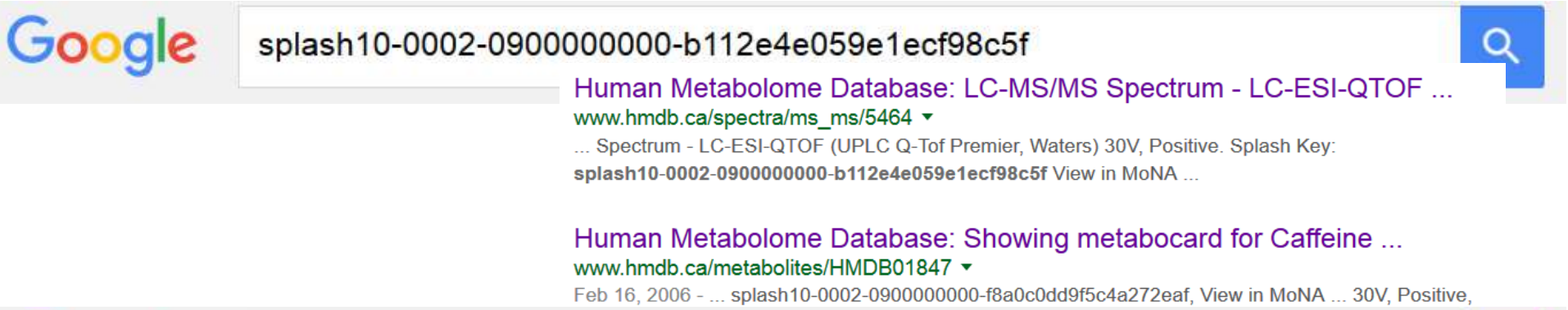
Stravs *et al*. 2013, *J. Mass Spectrom.,* 48, 89–99. DOI: 10.1002/jms.3131

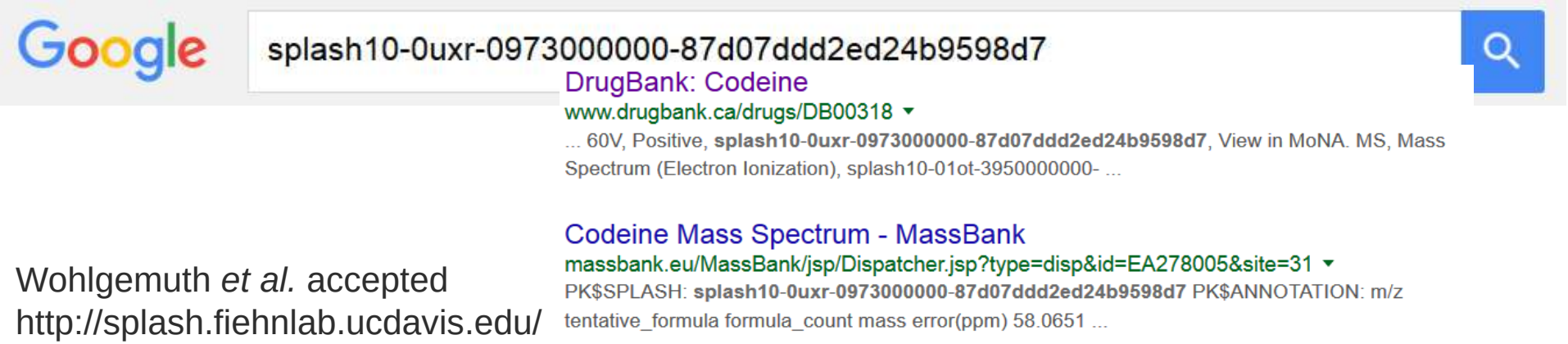
SPLASH – Communicate between libraries

splash10 - 0002 - 0900000000 - b112e4e059e1ecf98c5f
[version] - [top10] - [histogram] - [hash of full spectrum]

http://mona.fiehnlab.ucdavis.edu/#/spectra/splash/splash10-0002-0900000000-b112e4e059e1ecf98c5f

https://www.google.ch/search?q=splash10-0002-0900000000-b112e4e059e1ecf98c5f





Wohlgemuth *et al.* accepted
http://splash.fiehnlab.ucdavis.edu/

SPLASH – Communicate between libraries

http://splash.fiehnlab.ucdavis.edu/

Wohlgemuth *et al.* accepted

MassBank Record: EA278005

```
PK$SPLASH: splash10-0uxr-0973000000-87d07ddd2ed24b9598d7
PK$ANNOTATION: m/z tentative_formula formula_count mass error(ppm)
  58.0651 C3H8N+ 1 58.0651 0.25
  69.0335 C4H5O+ 1 69.0335 -0.45
```

DRUGBANK

Identification	
Name	**Codeine**
Accession Number	**DB00318** (APRD00120, DB09471)

splash10-0uxr-0973000000-87d07ddd2ed24b9598d7 View in MoNA

GNPS Library Spectrum CCMSLIB00000210977

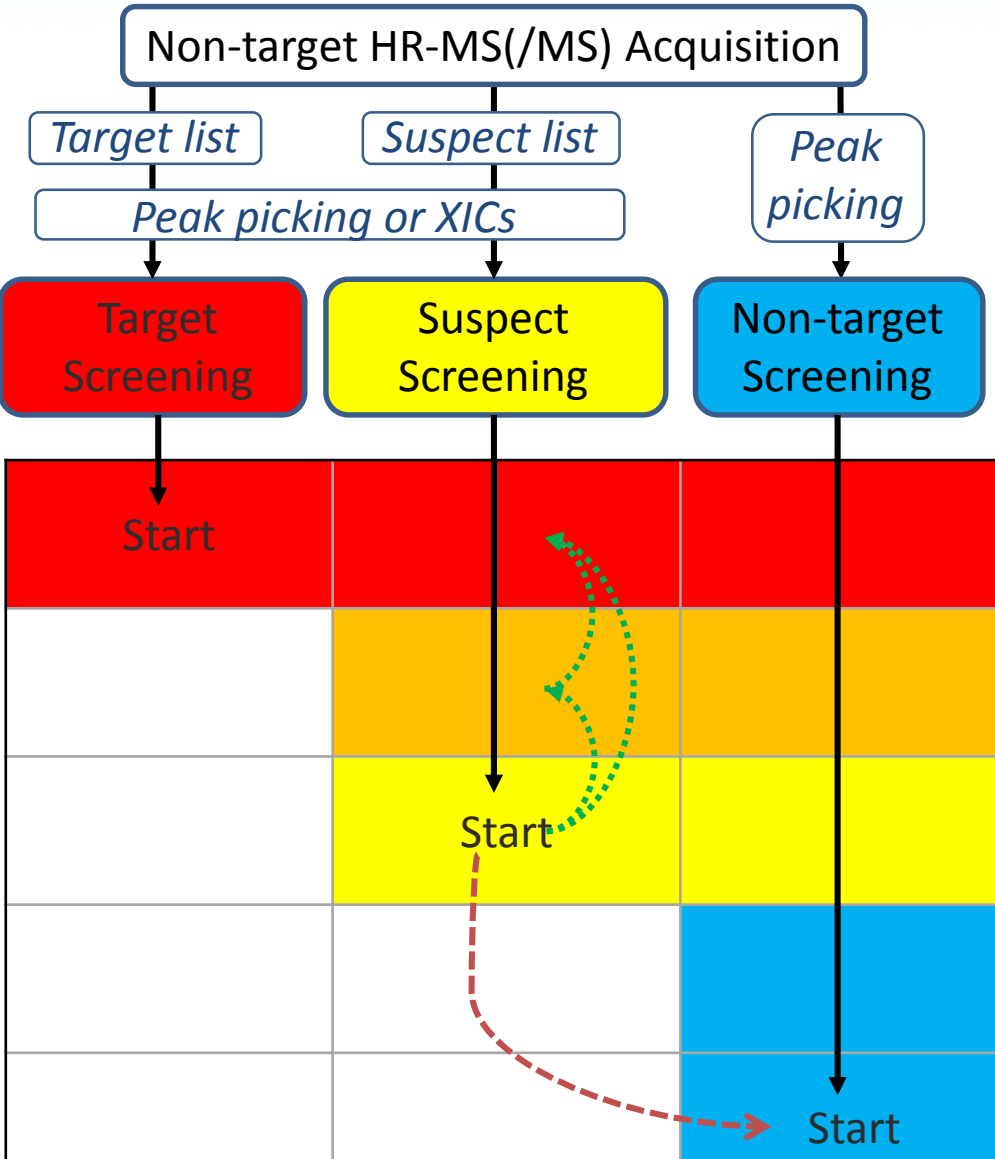
Library Membership	MASSBANK
SPLASH Key	splash10-0uxr-0973000000-87d07ddd2ed24b9598d7

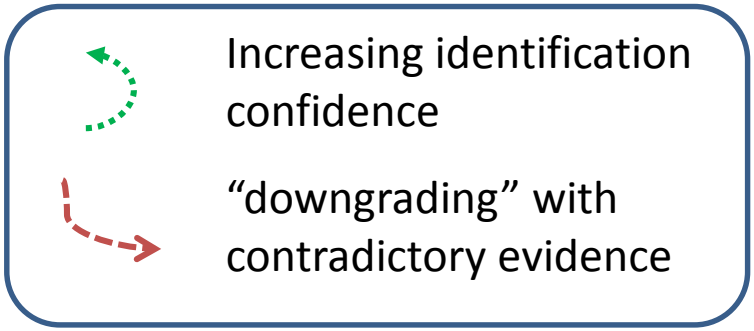
m/z CLOUD

splash10-0udi-0009000000-ff8a91aa5c5d21264a86
splash10-0udi-0009000000-89b20537b39389045b43
splash10-0udi-0009000000-d004653e649f5e6a0cda
splash10-0udi-0009000000-173ad69564777472205e
splash10-0udi-0009000000-a52c1f8d7e070d9019ba

“Probable Structure by Spectral Match”

Schymanski *et al*, 2014, *ES&T*, 48 (4), 2097-2098. DOI: 10.1021/es5002105
Schymanski *et al* 2015. DOI: 10.1007/s00216-015-8681-7

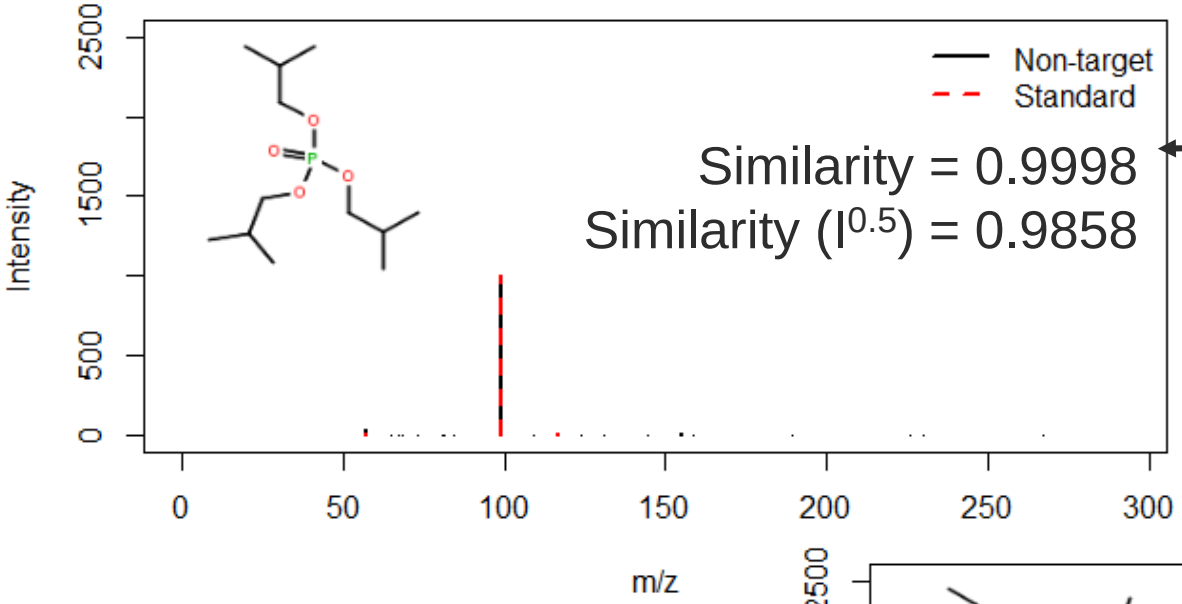




Level 1	***Confirmed Structure*** by reference standard
Level 2	***Probable Structure*** by library/diagnostic evidence
Level 3	***Tentative Candidate(s)*** suspect, substructure, class
Level 4	***Unequivocal Molecular Formula*** insufficient structural evidence
Level 5	***Mass of Interest*** multiple detection, trends, …

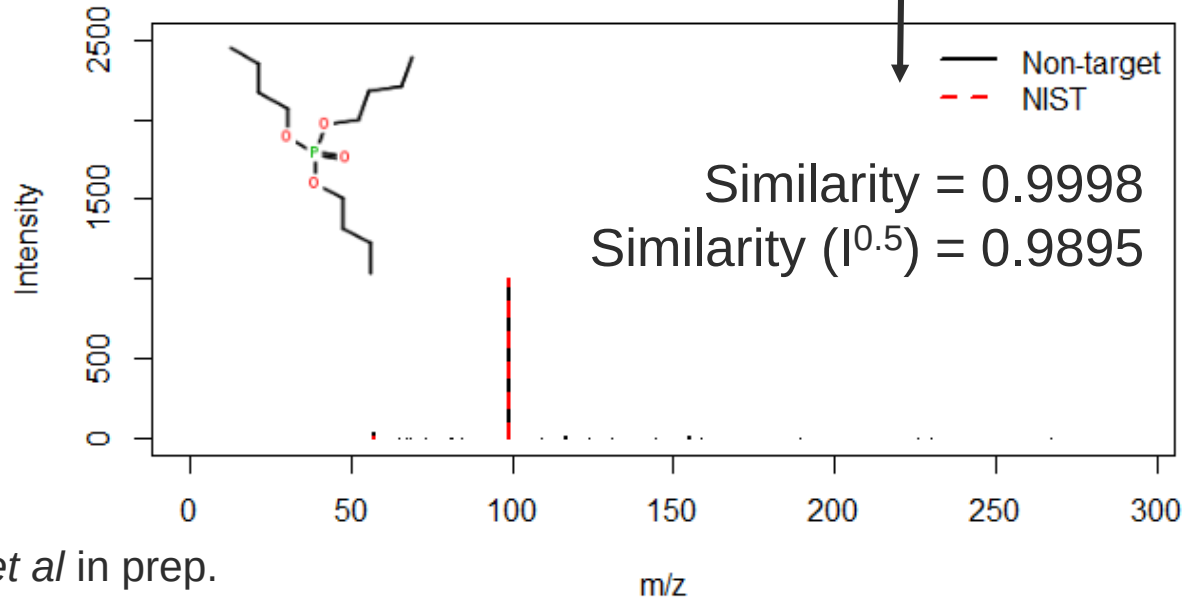
“Probable Structure by Spectral Match”

…interpret this with care!



Non-target and standard measured with identical conditions

Non-target and library spectrum acquired on different set-up



- Be very cautious with single-peak spectra!

Schollée *et al*. in prep.; Schymanski *et al* in prep.

What if my spectrum isn't in the library?

PANIC?

CFM-ID
Competitive Fragmentation Modeling for Metabolite Identification

MetFrag

MAGMa
MAGMa+

MS-FINDER

CSI:FingerId

Images from: http://cfmid.wishartlab.com/#; http://msbi.ipb-halle.de/MetFragBeta/; http://prime.psc.riken.jp/Metabolomics_Software/MS-FINDER/index.html; http://www.csi-fingerid.org/; http://www.emetabolomics.org/magma; https://github.com/savantas/MAGMa-plus

CASMI 2016
Important Dates
Contest Rules
Example Data
Challenge Data
Solutions
Results
Proceedings
About the Team

CASMI 2014

CASMI 2013

CASMI 2012

News

May 6th, 2016
The winners and full results are available.

April 25th, 2016
The solutions are public now.

April 18th, 2016
The contest is closed now, the results are fantastic and will be opened soon!

April 9th, 2016
All teams who submit before

Results in Category 2

Summary of Challenge wins

	Vaniya	Duehrkop	Verdegem	Allen	Brouard
Gold	70	82	44	63	**86**
Silver	26	21	53	71	50
Bronze	35	11	65	40	31
Gold (neg)	33	0	24	26	20
Gold (pos)	37	82	20	37	66

Summary statistics per participant

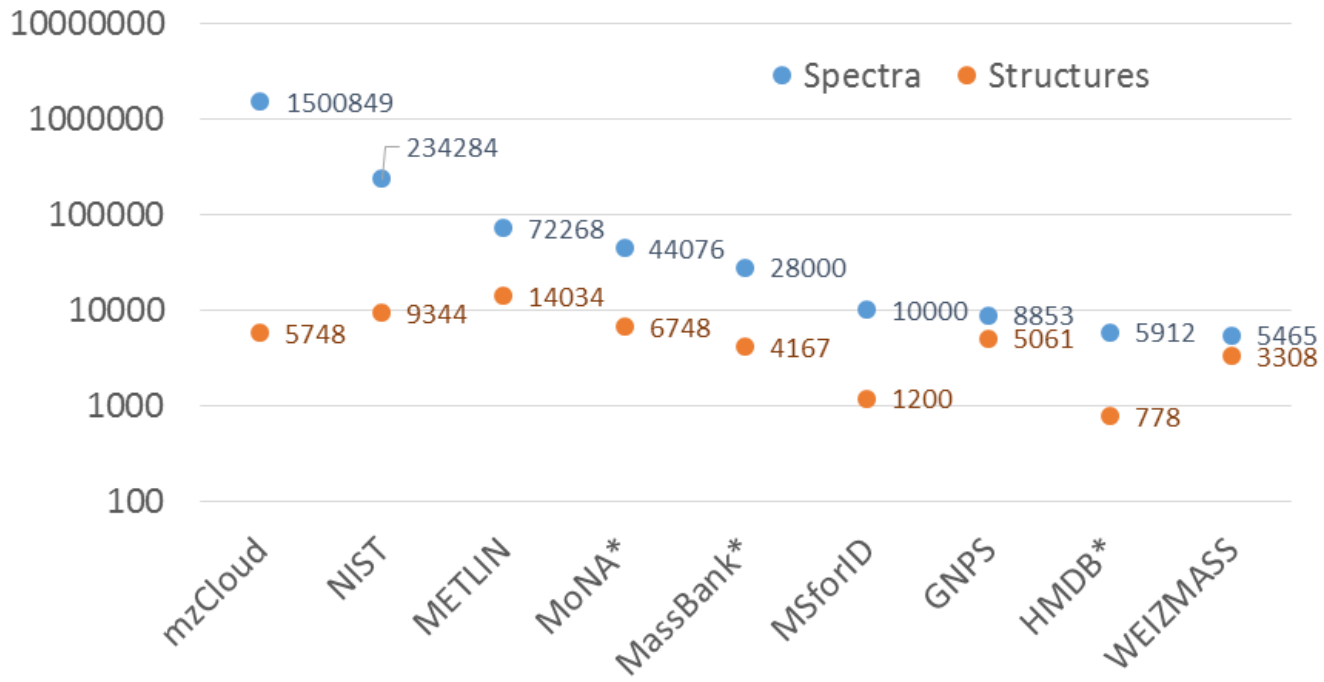
	Mean rank	Median rank	Top	Top3	Top10	Mean RRP	Median RRP
Vaniya	19.75	3.0	46	79	101	0.804	0.922
Duehrkop	25.17	1.0	70	90	100	0.945	1.000
Verdegem	70.79	9.8	24	59	105	0.880	0.972
Allen	47.98	6.0	39	77	123	0.906	0.987
Brouard	127.34	5.2	62	93	118	0.874	0.988

Summary of Rank by Challenge and Participant

For each challenge, the rank of the winner(s) is highlighted in bold. If the submission did not contain the correct candidate this is denoted as "-". If someone did not participate in a challenge, nothing is shown. The tables are sortable if you click into the column header.

Mass Spectral Libraries for Small Molecules

- There are many different mass spectral resources out there!
- Find the one(s) best suited to your actual purpose
- Always take care when interpreting your results – is the answer really clear?
- Consider contributing your spectra to open repositories
  - The more the merrier – BUT – high quality annotation is really important



Acknowledgements

EU Grant 603437

…the mass spec community

…and you all for listening!

Slides:
www.eawag.ch/~schymaem

METABOLOMICS SOCIETY

METABOLOMICS SOCIETY
EARLY- CAREER MEMBERS NETWORK

Next EMN webinar will be in November, 2016

Look out for the Advert