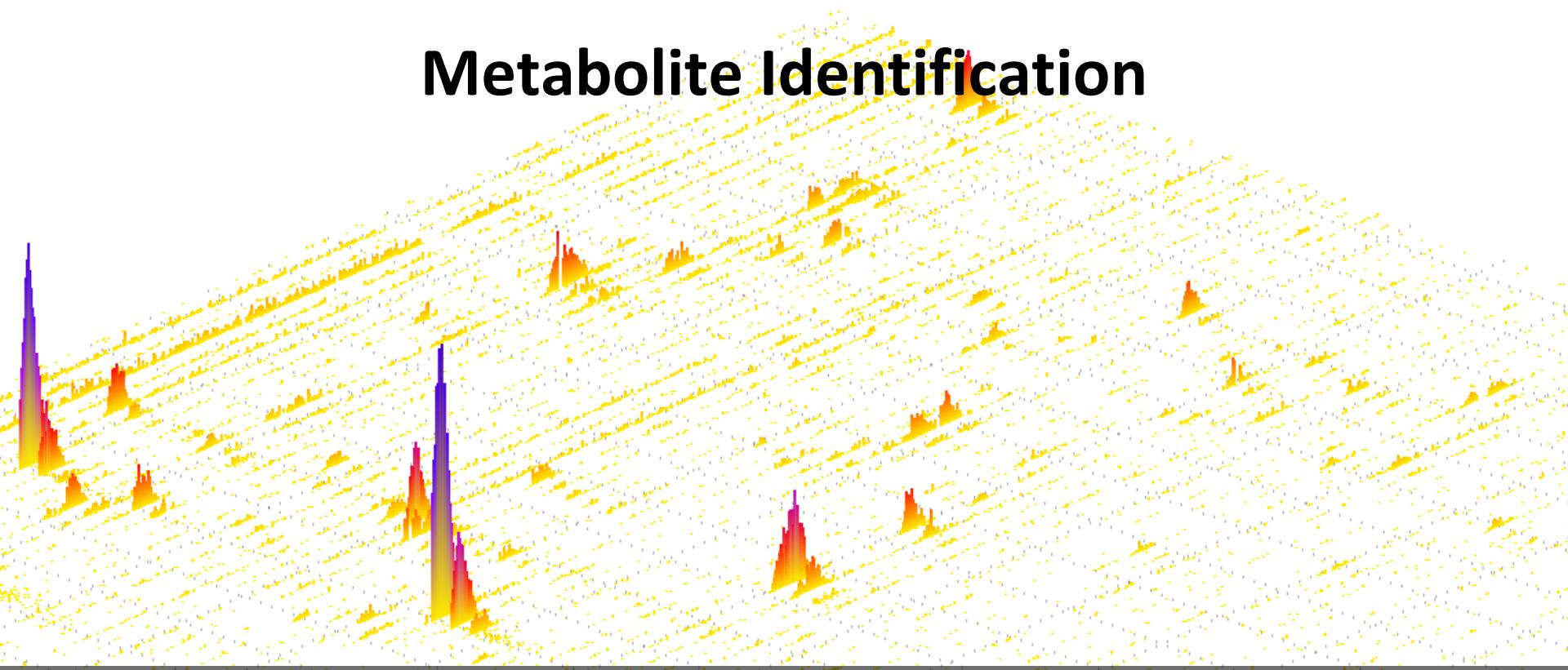


May Institute 2017
*Computation and statistics for mass
spectrometry and proteomics*

Metabolite Identification



MAX-PLANCK-GESELLSCHAFT

Oliver Kohlbacher
University of Tübingen and
MPI for Developmental Biology
KohlbacherLab.org | @okohlbacher

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN



Agenda

Wednesday 5/3/2017	
8:00 AM	Bring your own data or Skyjam
9:00 AM	Lecture: Metabolite identification
10:30 AM	Refreshments
11:00 AM	Hands-on: Metabolite identification through spectral databases and structure databases
12:30 PM	Wrap-up and jeopardy
1:30 PM	Adjourn

METABOLITE DATABASES

- Databases for mass spectra
- Databases for compounds
- Databases for pathways
- Comparison of databases

This work is licensed under a Creative Commons Attribution 4.0 International License.

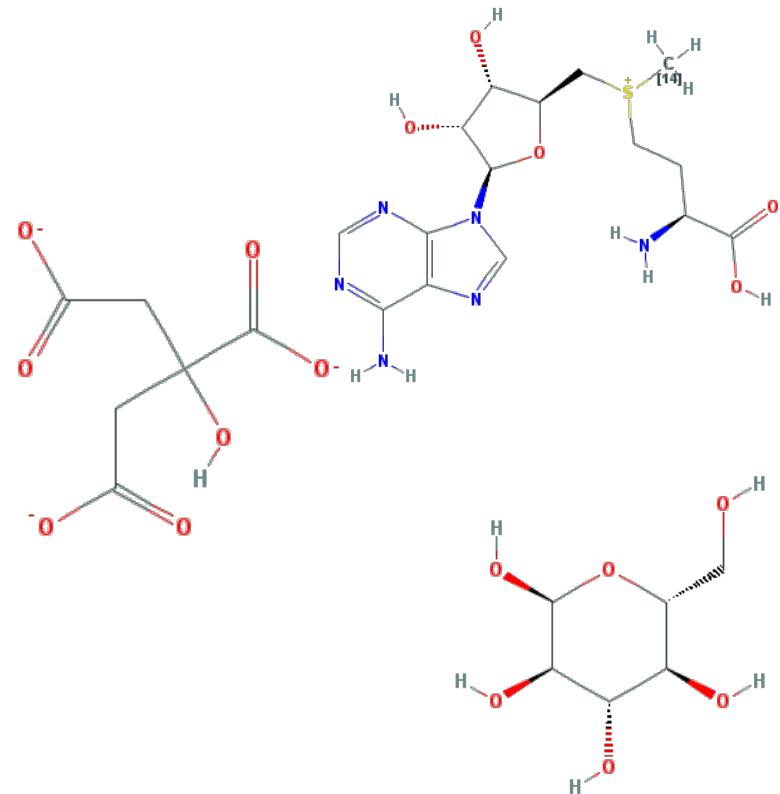


Metabolomics vs. Proteomics

Proteomics

EGGTSVANAER
VADILESNAR
MSIAIMAGVLEAR
TISGQDALPNISDAER
TFVDQEFAQIK
ITNHLVAMIEK
QGVATVLSAPAK

Metabolomics



Databases

- Metabolomics has to rely on different types of databases
 - **Structural databases** – contain structures of metabolites
 - **Pathway databases** – contain metabolic networks
 - **Spectrum databases** – contain information on mass spectra
- In contrast to sequencing data, many of the databases are commercial
- Freely accessible spectrum databases for metabolites are still rare

[Home](#)[Browse](#)[Search](#)[About](#)[Downloads](#)[Contact Us](#)

Human Metabolome Database Version 2.5

Search: [\[Advanced\]](#)



The Human Metabolome Database (HMDB) is a freely available electronic database containing detailed information about small molecule metabolites found in the human body. It is intended to be used for applications in metabolomics, clinical chemistry, biomarker discovery and general education. The database is designed to contain or link three kinds of data: 1) chemical data, 2) clinical data, and 3) molecular biology/biochemistry data. The database (version 2.5) contains over 7900 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant ($> 1 \mu\text{M}$) or relatively rare ($< 1 \text{nM}$). Additionally, approximately 7200 protein (and DNA) sequences are linked to these metabolite entries. Each MetaboCard entry contains more than 110 data fields with 2/3 of the information being devoted to chemical/clinical data and the other 1/3 devoted to enzymatic or biochemical data. Many data fields are hyperlinked to other databases (KEGG, PubChem, MetaCyc, ChEBI, PDR, Swiss Prot, and GenBank) and a variety of structure and pathway viewing applets. The HMDB database supports additional databases, [DrugBank](#), [T3DB](#), [SMPDB](#) and [FooDB](#) are also supported. Information on drugs, [T3DB](#) contains information on 2900 common toxins and environmental pollutants, while [FooDB](#) contains equivalent information on food constituents.

HMDB is supported by [David Wishart](#), Departments of [Computing Science](#)

Database of known human metabolites.
Rich in metadata and annotation, no mass spectra.

Human Metabolome Database

Metabolomics Toolbox MetaboLIM

Home Browse Search About Downloads Contact Us

Human Metabolome Database Version 2.5

Search: glucose [\[Advanced\]](#)



Search Results

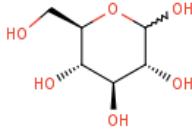
Search for "glucose" returned 546 results ([only search name and synonyms](#))

previous [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) ... [27](#) [28](#) next

Showing 1-20 out of 546 hits

HMDB ID	Name	Formula	Weight
HMDB00122 MetaboCard	D-Glucose ... Cerelose 2001; Clearsweet 95; Clintose L; Corn sugar; D(+)-Glucose; Dextropur; Dextrose; Dextrosol; Glucodin; Glucolin; Glucose; Goldsugar; Grape sugar; Meritose; Roferose ST; Staleydex 111...	C ₆ H ₁₂ O ₆	180.063385
HMDB03369 MetaboCard	CDP-glucose ... 2906-23-2 Cytoplasm (Predicted from LogP) C ₁₅ H ₂₅ N ₃ O ₁₆ P ₂ CDP-glucose is a substrate for Uridine diphosphate glucose pyrophosphatase.	C ₁₅ H ₂₅ N ₃ O ₁₆ P ₂	565.070984
HMDB00060 MetaboCard	Acetoacetic acid Anoxia, Diabetes mellitus type 2, Glucose transporter type 1 deficiency syndrome, Ketosis, Meningitis 1485291 3-KETOBUTYRATE Blood; Cellular Cytoplasm; Cerebrospinal Fluid; Urine 541-50-4 Cytoplasm...	C ₄ H ₆ O ₃	102.031693
HMDB12300 MetaboCard	UDP-4-dehydro-6-deoxy-D-glucose ... UDP-4-dehydro-6-deoxy-D-glucose is synthesized from UDP-glucose through the enzyme UDP-glucose 4,6-dehydratase. HMDB12300 Uridine 5'-[3-(6-deoxy-D-xylo-hexopyranosyl-4-ulose) dihydrogen...	C ₁₅ H ₂₂ N ₂ O ₁₆ P ₂	548.044434
HMDB01586 MetaboCard	Glucose 1-phosphate ... D-Glucose-1-phosphate; D-glucose 1-phosphate; D-glucose-1-P; Glucose 1-phosphate; Glucose 1-phosphoric acid; Glucose monophosphate; Glucose-1-phosphate; α-D-Glucopyranosyl phosphate; α-D-Glucose	C ₆ H ₁₃ O ₉ P	260.029724

Human Metabolome Database

Chemical IUPAC Name	(3R,4R,5S,6S)-6-(hydroxymethyl)oxane-2,3,4,5-tetrol
Chemical Formula	C ₆ H ₁₂ O ₆
Chemical Structure	
Chemical Taxonomy	<p>Kingdom</p> <ul style="list-style-type: none">Organic <p>Super Class</p> <ul style="list-style-type: none">Carbohydrates and Carbohydrate conjugates <p>Class</p> <ul style="list-style-type: none">Carbohydrates <p>Sub Class</p> <ul style="list-style-type: none">Monosaccharides <p>Family</p> <ul style="list-style-type: none">Mammalian Metabolite <p>Species</p> <ul style="list-style-type: none">hemiacetalprimary alcoholsecondary alcohol1,2-diolheterocyclic compound <p>Biofunction</p> <p>—</p> <p>Application</p> <p>—</p> <p>Source</p> <ul style="list-style-type: none">Endogenous
Average Molecular Weight	180.156
Monoisotopic Molecular Weight	180.063385
Isomeric SMILES	OC[C@H]1OC(O)[C@H](O)[C@@H](O)[C@@H]1O



Human Metabolome Database

Metabolomics Toolbox MetaboLIMS

Home Browse Search About Downloads Contact Us

Human Metabolome Database Version 2.5

Search: [\[Advanced\]](#)



Spectra Search

MS Search MS/MS Search GC/MS Search NMR Search

MS Search	
	<input type="button" value="Find Metabolites"/> [Help]
Database	<input checked="" type="checkbox"/> HMDB <input type="checkbox"/> FoodDB <input type="checkbox"/> DrugBank
Molecular Species	<input checked="" type="radio"/> Positive Mode <input type="radio"/> Negative Mode <input type="radio"/> Neutral Molecule
MW (Da) (May enter multiple MW's, one on each line) Positive Mode example Negative Mode example Neutral Molecule example	181.07063
MW Tolerance (\pm)	0.1 (Da)
	<input type="button" value="Find Metabolites"/> [Help]

MS Search Result

First Prev Next Last | 150 Rows Displayed | Export XLS

Human Metabolome Database

209 results found, displaying 1 to 150

Search
Clear

HMDB ID	Common Name	Chemical Formula	Adduct MW (Da) [Matching HMDB MW]	MW Difference (Da) [QueryMass - AdductMass]	Adduct
HMDB06088	Scyllitol	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB03345	Alpha-D-Glucose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB01151	Allose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00660	D-Fructose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00516	Beta-D-Glucose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB12326	L-Gulose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00346	3-Deoxyarabinohexonic acid	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00211	Myoinositol	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00169	D-Mannose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00143	D-Galactose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB00122	D-Glucose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB03449	Beta-D-Galactose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB01266	L-Sorbose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB03418	D-Tagatose	C6H12O6	181.070663 [180.063385]	3.1E-5	M+H [1+]
HMDB02825	Theobromine	C7H8N4O2	181.072006 [180.064728]	0.001373	M+H [1+]
HMDB01889	Theophylline	C7H8N4O2	181.072006 [180.064728]	0.001373	M+H [1+]
HMDB01860	Paraxanthine	C7H8N4O2	181.072006 [180.064728]	0.001373	M+H [1+]
HMDB12883	Adrenochrome o-semiquinone	C9H10NO3	181.073349 [180.066071]	0.002716	M+H [1+]
HMDB03269	Nicotinuric acid	C8H8N2O3	181.060776 [180.053497]	0.009857	M+H [1+]
HMDB11751	3-Methoxybenzenepropanoic acid	C10H12O3	181.085922 [180.078644]	0.015289	M+H [1+]
HMDB12915	Coniferyl alcohol	C10H12O3	181.085922 [180.078644]	0.015289	M+H [1+]
HMDB01087	5-Methylthioribose	C6H12O4S	181.052902 [180.045624]	0.017731	M+H [1+]
HMDB00707	4-Hydroxyphenylpyruvic acid	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB06915	2-Hydroxy-3-(4-hydroxyphenyl)propanoic acid	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB01964	Caffeic acid	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB11663	3-Hydroxyphenylpyruvic acid	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB01879	Aspirin	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB02130	Monomethyl phthalate	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB03501	3,4-Dihydroxy-trans-cinnamate	C9H8O4	181.049530 [180.042252]	0.021103	M+H [1+]
HMDB04076	5-Hydroxykynurenamine	C9H12N2O2	181.097153 [180.089874]	0.02652	M+H [1+]
HMDB13319	Tyrosinamide	C9H12N2O2	181.097153 [180.089874]	0.02652	M+H [1+]
HMDB00700	Hydroxypyropionic acid	C3H6O3	181.070663 [90.031693]	3.1E-5	2M+H [1+]
HMDB00694	L-2-Hydroxyglutaric acid	C5H8O5	181.070663 [148.037170]	3.1E-5	M+CH3OH+H [1+]
HMDB00606	D-2-Hydroxyglutaric acid	C5H8O5	181.070663 [148.037170]	3.1E-5	M+CH3OH+H [1+]
HMDB01051	Glyceraldehyde	C3H6O3	181.070663 [90.031693]	3.1E-5	2M+H [1+]
HMDB00428	3-Hydroxyglutaric acid	C5H8O5	181.070663 [148.037170]	3.1E-5	M+CH3OH+H [1+]
HMDB00426	Citramalic acid	C5H8O5	181.070663 [148.037170]	3.1E-5	M+CH3OH+H [1+]
HMDB01900	Ribonolactone	C5H8O5	181.070663 [148.037170]	3.1E-5	M+CH3OH+H [1+]
HMDB00190	L-Lactic acid	C3H6O3	181.070663 [90.031693]	3.1E-5	2M+H [1+]
HMDB11676	D-Xylo-1,5-lactone	C5H8O5	181.070663 [148.037170]	3.1E-5	M+CH3OH+H [1+]
HMDB01882	Dihydroxyacetone	C3H6O3	181.070663 [90.031693]	3.1E-5	2M+H [1+]
HMDB01311	D-Lactic acid	C3H6O3	181.070663 [90.031693]	3.1E-5	2M+H [1+]
HMDB00120	Cysteine sulphoxide	C9H12N2O2	181.060606 [148.038051]	0.001023	M+ACN+H [1+]

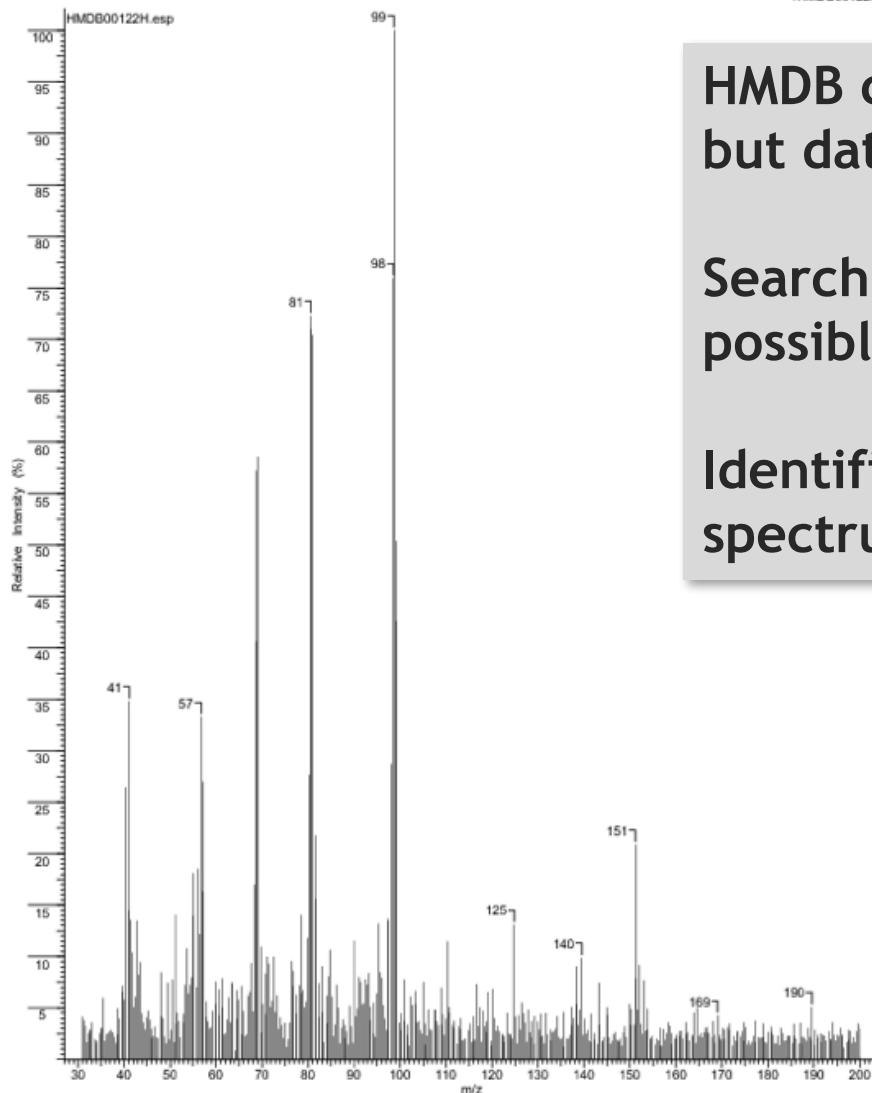
Human Metabolome Database

Material Safety Data Sheet (MSDS)	<ul style="list-style-type: none">• Show PDF/HTML
MOL File	Show
SDF File	Show
PDB File	Show
2D Structure	View 2D Structure
3D Structure	View 3D Structure
Experimental PDB ID	1A47
Experimental PDB File	Show
Experimental PDB Structure	View 3D Structure
Experimental ^1H NMR Spectrum	Download Spectrum Download FID (Varian) Show Experimental Conditions
Experimental ^{13}C NMR Spectrum	Download Spectrum Download FID (Bruker) Show Experimental Conditions
Experimental ^{13}C HSQC Spectrum	Download Spectrum Download FID (Bruker) Show Experimental Conditions
Predicted ^1H NMR Spectrum	Show Image Show Peaklist
Predicted ^{13}C NMR Spectrum	Show Image Show Peaklist
Mass Spectrum	Low Energy Download File Show Experimental Conditions
	Medium Energy Download File Show Experimental Conditions
	High Energy Download File Show Experimental Conditions

Human Metabolome Database

HMDB00122 (High Energy)

7 Aug 2008
/HMDB00122H



HMDB contains mass spectra of metabolites, but data is not downloadable in bulk.

Search against the spectra database is only possible through the web interface.

Identification by accurate mass or MS/MS spectrum comparison.

Human Metabolome Database

Spectra Search

MS Search MS/MS Search GC/MS Search NMR Search

NOTES:

To query the database using spectral pattern matching, upload the MS/MS data file for the metabolite OR paste its content in the textarea box below.

**** Fields are mandatory**

MS/MS Search

Search By

[\[Help\]](#)

[\[Help\]](#)

******MW of Parent Ion

(Da)

******MW Tolerance (\pm)

(Da)

Instrument Type

[\[Help\]](#)

******Fragment Ion Tolerance (\pm)

(Da)

CID Energy Level

[\[Help\]](#)

Ionization Mode

[\[Help\]](#)

MS/MS Data File

No file chosen

OR

Content of MS/MS Data File

[Aconitic Acid \(Low Energy, Positive Ion Mode\) example](#)

[Xanthine \(Medium Energy, Positive Ion Mode\) example](#)

[Hippuric Acid \(Medium Energy, Negative Ion Mode\) example](#)

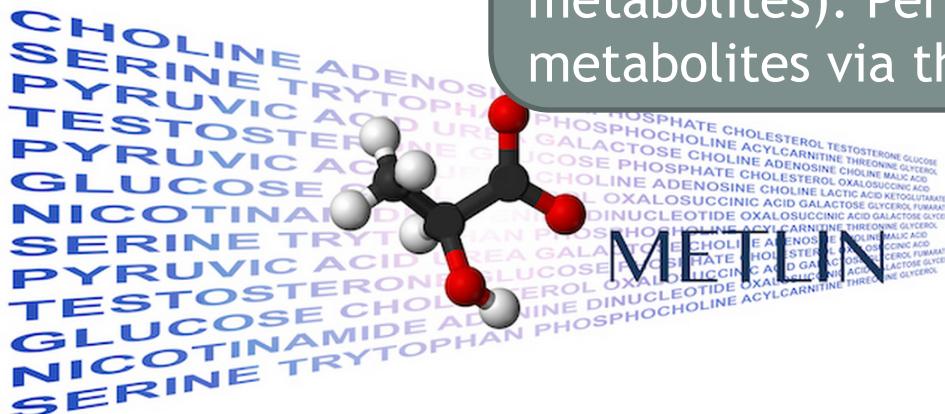
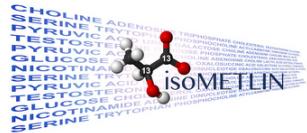
[Citric Acid \(Low Energy, Negative Ion Mode\) example](#)

41.400 9.926
85.030 100.000
111.118 24.412
128.847 30.000
172.851 11.912

NOTES:

1. m/z (Da) and relative intensities, RInt (%), delimited by a space (" "),
2. m/z and RInt MUST contain a decimal.
3. m/z MUST be less than m/z of the parent ion minus 10 Da.

[\[Help\]](#)



Database containing a large number of metabolites (240,000+) and spectra for those (12,000 metabolites). Permits search of metabolites via their mass spectra.

Statistics

- # Metabolites: 240,516
- # High Resolution MS/MS Spectra: 61,872
- # Metabolites w/ High Resolution MS/MS: 12,057

Functionality

- Single & Batch**
Precursor Ion (m/z) searching
- Single & Multiple**
Fragment Ion (m/z) searching

MassBank

The screenshot shows the MassBank Statistics page. On the left is a sidebar with links: Database Service, Statistics (which is selected), Publications, Download, Manuals, About MassBank, Contact, Consortium Members, Site Map, and Use Restrictions. The main content area has a title 'Statistics' with a bar chart icon. Below it is a sub-section titled 'Statistics' with a bar chart icon. A message says 'Last updated Mar 5, 2014 | Total Number of Spectra : 40,889 new'. A table lists research groups and their analysis equipment:

Research Groups (Contact Name)	Prefix of ID	Analysis Equipment (Analysis Method)	Number of Spectra	Number of Compounds
01. IAB, Keio U (Dr. Tomoyoshi Soga)	KOX	LC-ESI-QTOF (MS2)	2,839	672
	KO	LC-ESI-QQ (MS2)	4,265	
		LC-ESI-IT (MS2,MS3,MS4)	515	
		GC-EI-TOF (MS)	241	
02. PSC, RIKEN (Dr. Masanori)				
03. Nihon Waters K (Dr. Katsutos)				
04. Grad Sch Pharm & Res Inst Prod Dev (Dr. Naoshige)				
05. College Life Hea (Dr. Takashi Maoka)				

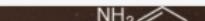
A callout bubble in the bottom right corner contains the text: 'Database containing mass spectra of a large number of metabolites and metadata for these compounds. Permits search of metabolites via their mass spectra.'

GNPS

GNPS: Global Natural Products Social Molecular Networking

User: Pass: Sign in
Don't have an account? [Register](#)

MassIVE Datasets | Documentation | Forum | Contact



[Back to main page](#)

[Back to status page](#)

Collapse all

[Download](#)

GNPS Library Release 8

◀

Hits 1 ~ 30 out of 1607

▶ Go to

Go

Select column

Rapidly growing resource of mass spectra with rich interaction and annotation options.

MoNA – Mass Bank of North America

Search...



Welcome to MoNA!

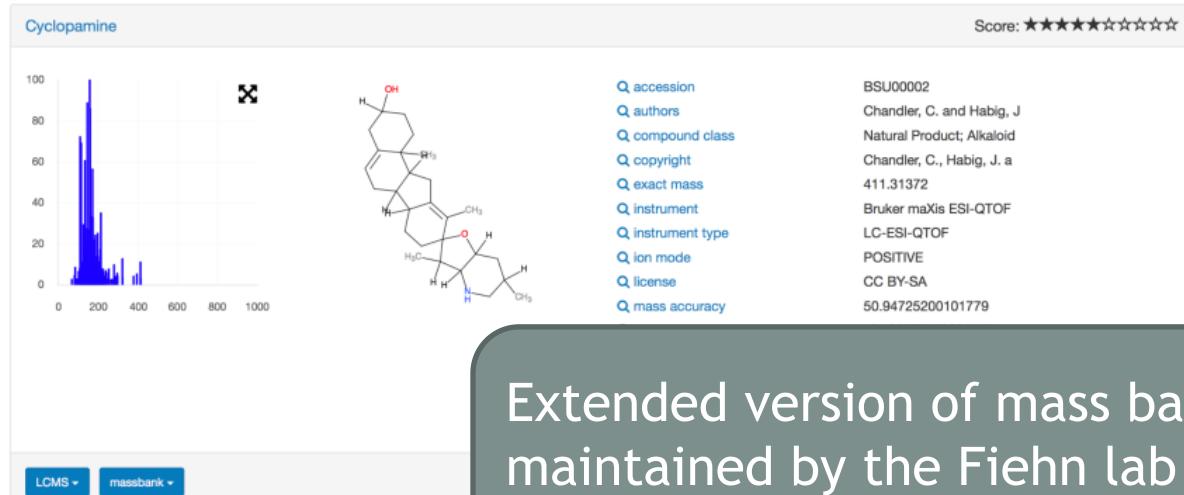
MassBank of North America (MoNA) is a metadata-centric, auto-curating repository designed for efficient storage and querying of mass spectral records. It intends to serve as a framework for a centralized, collaborative database of metabolite mass spectra, metadata and associated compounds. MoNA currently contains over 200,000 mass spectral records from experimental and in-silico libraries as well as from user contributions.

MoNA has recently been redesigned, with significant improvements to server-side architecture, query structure, and search speed. We are actively improving and adding features, so please be patient as functionality is added. If you notice any major issues, feel free to report them using the issue tracker linked below.

Search Spectra

Browse Spectra

Issue Tracker



Extended version of mass bank
maintained by the Fiehn lab at UC
Davis.

mzCloud

mzcloud.org

Views Standard Compare Structures

Reference Library

Results for 'caffeine'

No: 339

Caffeine

Monoiso. Mass: 194,08038

+ Thermo

+ Thermo

+ UC Davis

+ Eawag

+ Thermo

- Thermo

No: 770

Isocaffeine

Monoiso. Mass: 194,08038

+ Thermo

Spectral Tree

Raw Filtered Recalibrated

MS1 Scns. #34, 93

MS2 195.09 Scns. #35, 94

MS3 196.07 Scns. #42, 101

1/2 MS1 Combined Scans #34, 93

Recalibrated Spectrum

195,0877 $[M + H]^+$

217,0696 $[M + Na]^+$

$[2M + Na]^+$
411,1500

Precursor structure

To see precursor structure please select tandem spectrum in Spectral Tree

mzCloud (currently) contains only a very limited number of metabolites, but very high quality spectra and is growing quickly.



KEGG - Table of Contents

KEGG2 PATHWAY GENES LIGAND KO SSDB EXP

1. KEGG Databases

KEGG (Kyoto Encyclopedia of Genes and Genomes) contains structural information of metabolites as well as pathway information.

Category	Database		Search & Compute	DBGET Search
Pathway information	KEGG PATHWAY Database	XML	Search objects in KEGG pathways Color objects in KEGG pathways	PATHWAY
Genomic information	KEGG GENES Database	KO	Search orthologs or gene clusters in SSDB Search similar GENES sequences Search similar GENOME sequences	KO GENES GENOME
Chemical information	KEGG LIGAND Database	RC	Search similar compound structures Search similar glycan structures Predict reactions and assign EC numbers Generate possible reaction paths	COMPOUND GLYCAN REACTION RPAIR ENZYME LIGAND



HumanCyc

Encyclopedia of
Human
Genes and
Metabolism

BioCyc Home Search

[Database Search](#)
[Advanced Database Search](#)
[Help](#)

News

Nov 09 [BioCyc 8.6 released](#)
Sep 17 [BioCyc 8.5 released](#)
Sep 17 [Online Licensing](#)

Services

[Software/Data Download](#)
[User Support](#)
[Subscribe to Mailing List](#)
[EcoCyc T-shirts](#)

Information

[Introduction to BioCyc](#)
[Guided Tour](#)
[Pathway Tools Software](#)

HumanCyc: Encyclopedia of Genes and Metabolism

- [Query the HumanCyc database](#)

Authors

Pedro Romero, Markus Krummenacker and [Peter D. Karp, SRI International](#).

Project Overview

HumanCyc is a bioinformatics database that describes the human metabolic pathways and the human genome. By presenting metabolic pathways as an organizing framework for the human genome, HumanCyc provides the user with an extended dimension for functional analysis of

HumanCyc contains a very comprehensive set of metabolic pathways.

 NCBI

PubChem

National Library of Medicine 

HOME | SEARCH | SITE MAP | PubMed | Entrez | Structure | GenBank | Text Search

PubChem Text Search

PubChem Substance

PubChem contains the chemical structures of small molecules and information on their biological activities.

PubChem Substance: Search PubChem/Substance database using text terms including name, synonym, keyword, external ID, formula, SID, etc.

PubChem Compound: Search PubChem/Compound using text terms including name, synonym, keyword, external ID, CID, formula, etc.

PubChem BioAssay: Search PubChem/BioActivity database using text terms such as cell name, protocol keyword, etc.

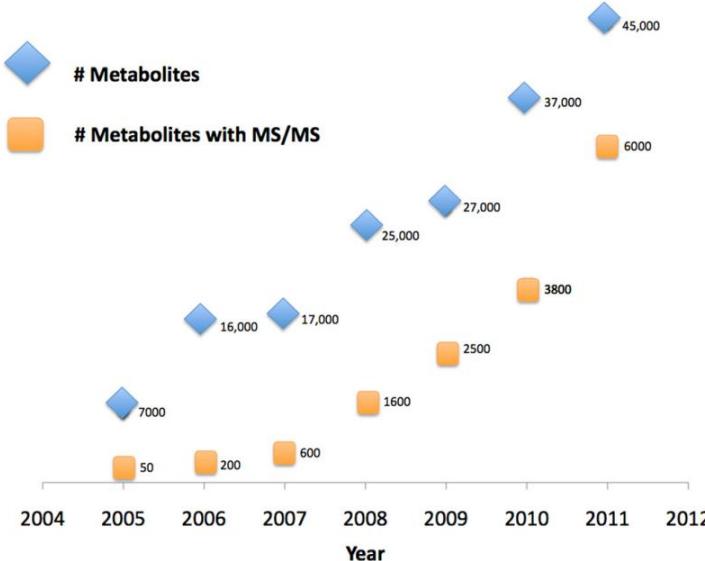
PubChem Structure Search: Search PubChem/Compound using chemical structure. Structure may be specified using SMILES, MOL file, molecular

PubChem contains structures and names of around 30 mio. compounds, including most structurally characterized metabolites.

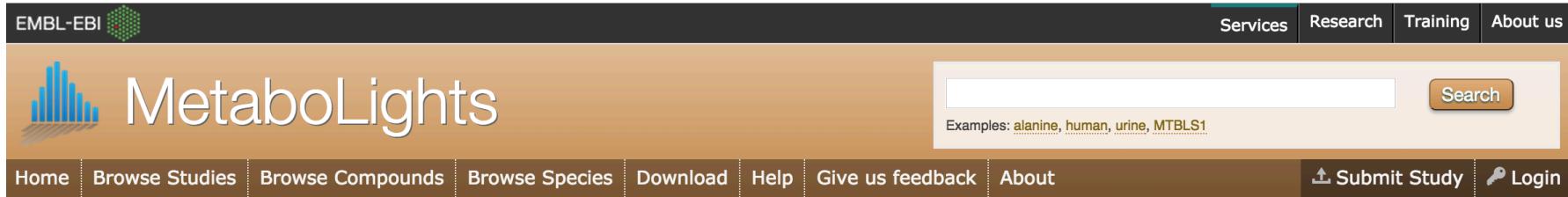
Database Comparison

DB	# cpds	# MS ² spectra
HMDB (2.5)	41,818	10,869
METLIN (2015)	240,000	62,204
MassBank (2015)	?	40,889
NIST (commercial, 2011)	4,694	~100,000
mzCloud (2015)	2,091	142,690
PubChem (2015, cpds only!)	63,156,000	-

METLIN Statistics



Raw Data Repositories



The screenshot shows the MetaboLights homepage. At the top, there's a dark header with the EMBL-EBI logo and navigation links for Services, Research, Training, and About us. Below the header is a search bar with a placeholder 'Search' and a 'Search' button. To the left of the search bar is a blue bar with the MetaboLights logo and the word 'MetaboLights'. Below the search bar is a brown navigation bar with links for Home, Browse Studies, Browse Compounds, Browse Species, Download, Help, Give us feedback, and About. To the right of the navigation bar are links for Submit Study and Login.

MTBLS1 FROM UNIVERSITY OF CAMBRIDGE
NMR based metabolomics of Human Type 2 Diabetes urine sam

MetaboLights stores original experimental data (raw data) and interpreted data from metabolomics studies.

MetaboLights

MetaboLights is a database for Metabolomics experiments and derived information. The database is cross-species, cross-technique and covers metabolite structures and their reference spectra as well as their biological roles, locations and concentrations, and experimental data from

Download

 **Pre-packaged ISACreator download.** To make it easy for new users, please download and just unzip our pre-packaged ISACreator with plugin and configurations.

 **Experiments.** All public MetaboLights experiments can be downloaded from our public [ftp archive](#). Please find zip archives under the "studies" folder. Each public study can be found in the corresponding MTBLS-id

Submit a new study

Use this option if your study has not been submitted before

Update an existing study

Use this option if you like to update a previously submitted study

Raw Data Repositories



MetaboLights

Examples: alanine, human, urine, MTBLS1

Search

Home Browse Studies Browse Compounds Browse Species Download Help Give us feedback About

Submit Study Login

MetaboLights > Study

MTBLS17: Utilization of Metabolomics to Identify Serum Biomarkers for Hepatocellular Carcinoma in Patients with Liver Cirrhosis

[Share Study](#) | [View all files](#)

Submitted: 11-Jun-2013 , Release date: 30-Jun-2013

Habtom Ressom

Characterizing the metabolic changes pertaining to hepatocellular carcinoma (HCC) in patients with liver cirrhosis is believed to contribute towards early detection, treatment, and understanding of the molecular mechanisms of HCC. we compare metabolite levels in sera of 78 HCC cases with 184 cirrhotic controls by using ultra performance liquid chromatography coupled with a hybrid quadrupole time-of-flight mass spectrometry (UPLC-QTOF MS). Several candidate metabolic biomarkers for early detection of HCC cases in high risk population of cirrhotic patients are identified using mass spectrum.

Study Design Description Protocols Samples Assay 1  Assay 2  Study Files

Organism(s):
Homo sapiens (Human)

Study Design Description:

- Hepatocellular carcinoma
- liver cirrhosis
- Liquid Chromatography Mass Spectrometry

Datasets are identified via the MetaboLights ID (here:MTBLS17). Metadata for the study describes the samples and method employed (here: HCC study using LC-MS).

Raw Data Repositories



Show 10 entries

Filter:

Sample Name	Protocol REF	Post Extraction	Derivatization	Extract Name	Protocol REF	Chromatography Instrument	Column model	Column type
Exp1_CRR_1a_NEG	Extraction		none	Chromatography	ACQUITY UPLC Systems with 2D Technology	50 × 2.1 mm ACQUITY 1.7-µm C18 column (Waters)	reverse phase	
Exp1_CRR_1b_NEG	Extraction		none	Chromatography	ACQUITY UPLC Systems with 2D Technology	50 × 2.1 mm ACQUITY 1.7-µm C18 column (Waters)	reverse phase	
Exp1_CRR_3a_NEG	Extraction		none	Chromatography	ACQUITY UPLC Systems with 2D Technology	50 × 2.1 mm ACQUITY 1.7-µm C18 column (Waters)	reverse phase	
				Chromatography	ACQUITY UPLC Systems with 2D Technology	50 × 2.1 mm ACQUITY 1.7-µm C18 column (Waters)	reverse phase	

For every study there is also information available on each of the individual samples/runs (here: instrument, separation, sample description).

Raw Data Repositories

Study Design Description

Protocols

Samples

Assay 1 

Assay 2 

Study Files

 Download whole study |  Download metadata |  View all files

List of study files

Type part of a filename and press enter to select. Prefix with ! to deselect.

Select	File
<input type="checkbox"/>	Exp2_CRR_22a_POS.CDF
<input type="checkbox"/>	Exp1_CRR_79b_POS.CDF
<input type="checkbox"/>	Exp2_CRR_2b_POS.CDF
<input type="checkbox"/>	Exp1_CRR_31b_NEG.CDF
<input type="checkbox"/>	Exp2_CRR_12b_POS.CDF
<input type="checkbox"/>	Exp1_HCC_32b_NEG.CDF
<input type="checkbox"/>	Exp1_CRR_82b_NEG.CDF
<input type="checkbox"/>	Exp2_HCC_13a_POS.CDF
<input type="checkbox"/>	Exp2_CRR_32b_POS.CDF
<input type="checkbox"/>	Exp1_CRR_114b_NEG.CDF
<input type="checkbox"/>	Exp1_CRR_61b_POS.CDF
<input type="checkbox"/>	Peaklist_EXP1_NEG.xlsx
<input type="checkbox"/>	Exp1_CRR_78a_POS.CDF
<input type="checkbox"/>	Exp1_CRR_43b_NEG.CDF
<input type="checkbox"/>	Exp2_CRR_37b_NEG.CDF

Usually the raw data for each run/sample (as acquired on the instrument) can be downloaded for (re-)analysis.

Raw Data Repositories

EMBL-EBI

MetaboLights

Home Browse Studies Browse Compounds Browse Species Download

MetaboLights > Compound search

FILTER YOUR RESULTS 291 RESULTS FOUND

Page : 1 Showing results 1 to 10 1 2 3 4 5 ... 30 >

Technology

- NMR spectroscopy
- mass spectrometry
- not reported

Organism

- Homo sapiens (Human)
- Alpinia species
- Arabidopsis thaliana
- Arabidopsis thaliana (thale cress)
- Caenorhabditis elegans
- Carthamus oxyacantha
- Cordyceps sinensis
- Daphnia magna
- Escherichia coli
- Ficus mucuso
- Homo sapiens

dUMP (MTBLC17622)

A pyrimidine 2'-deoxyribonucleoside 5'-monophosphate having uracil as the nucleobase.

Identified in MTBLS20

5-methylcytosine (MTBLC27551)

A pyrimidine that is a derivative of cytosine, having a methyl group at the 5-position.

Identified in MTBLS20

Data can also be browsed by (identified) compound to see what metabolites have been identified by which method (here: MS) or in a certain organism (here: human).

METABOLITE ID VIA SPECTRAL MATCHING AND ACCURATE MASS

- Distribution of metabolite masses
- Matching MS/MS spectra against databases
- Accurate mass search

This work is licensed under a Creative Commons Attribution 4.0 International License.



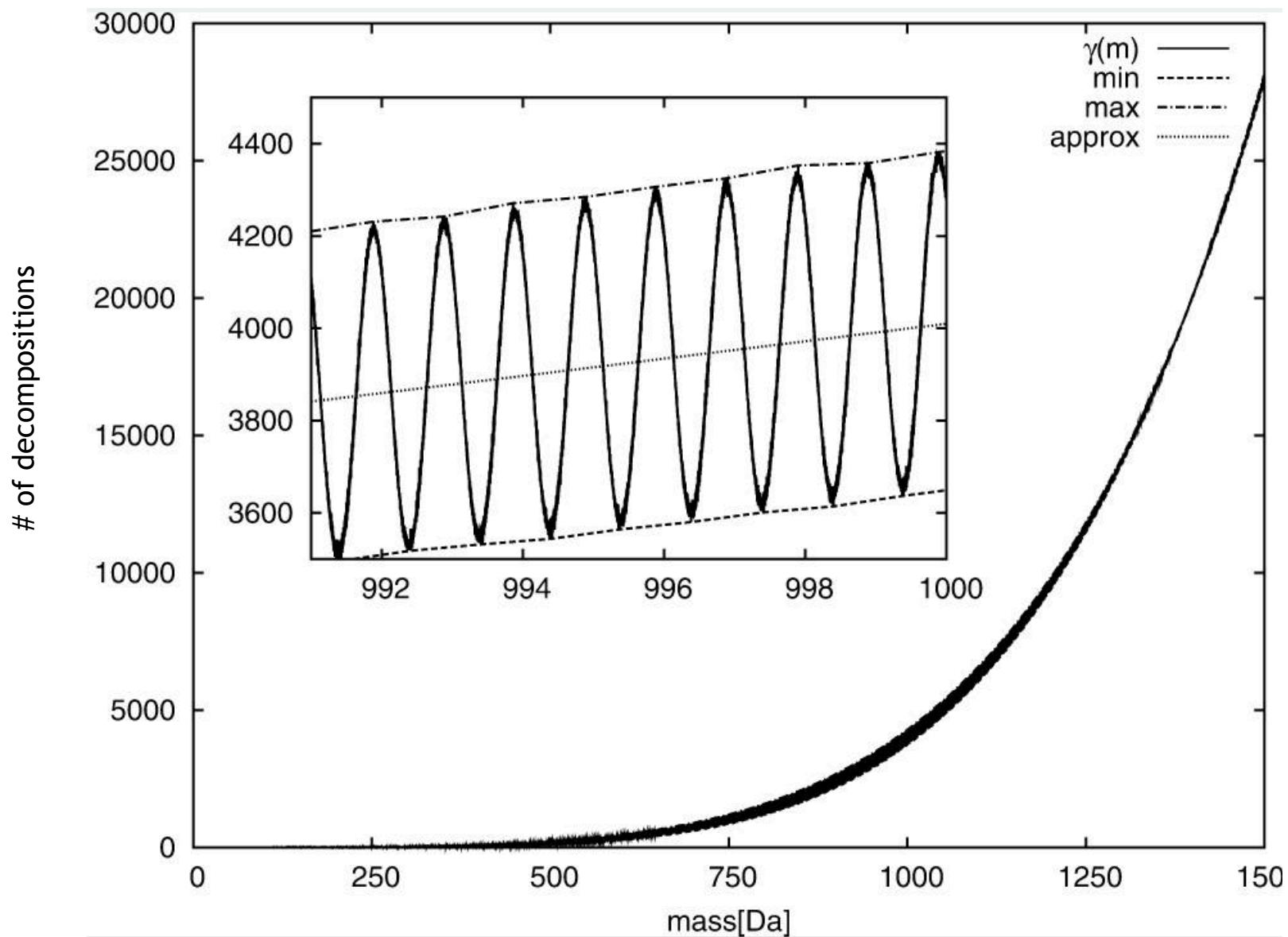
Identification Algorithms

- Identification by **accurate mass**
 - Given a sufficient mass resolution, the mass alone contains valuable information on the metabolite
 - Usually, these masses are not unique
 - Problem: structural isomers are isobaric!
- Identification through **spectrum comparison**
 - **Library search**
 - Compare the experimental spectrum against other experimental spectra
 - Requires library of experimental spectra (difficult)
 - **Fragment tree approaches**
 - Try to predict fragmentation patterns
 - Compare spectrum against theoretical fragmentation pattern

Mass Distribution

Mass Range	# of PubChem Hits	Mass Range	# of PubChem Hits
50-51 Da	71	100.0-100.1 Da	367
100-101 Da	2,180	100.1-100.2 Da	1,628
150-151 Da	9,163	100.2-100.3 Da	60
200-201 Da	23,867	100.3-100.4 Da	10
250-251 Da	48,909	100.4-100.5 Da	21
300-301 Da	78,577	100.5-100.6 Da	28
350-351 Da	112,566	100.6-100.7 Da	3
400-401 Da	130,737	100.7-100.8 Da	3
Limited number of different isotope masses gives rise to ‘lumpy’ distribution across the mass range.			

Mass Distribution



Isomers

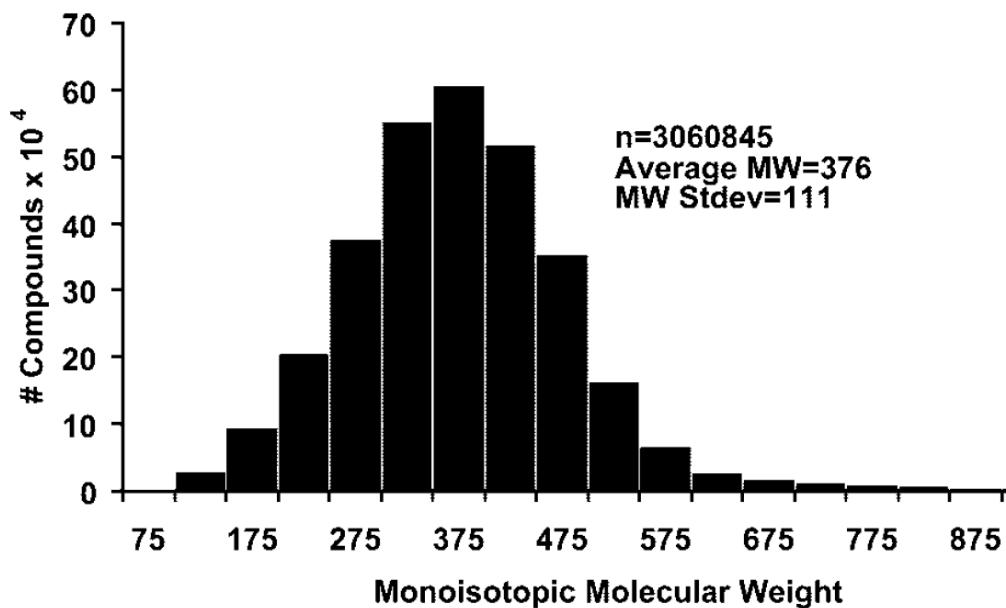
Example: Structural isomers with mass 117.0790 Da, formula: C₅H₁₁NO₂

METLIN ID	MASS	Appm	NAME	MS/MS	STRUCTURE
287	117.0790	0	Betaine Formula:C5H11NO2 CAS:107-43-7	View	
6508	117.0790	0	N-Methyl-a-aminobutyric acid Formula:C5H11NO2 CAS:	View	
6762	117.0790	0	Isoamyl nitrite Formula:C5H11NO2 CAS:	NO	
6902	117.0790	0	5-Aminopentanoic acid Formula:C5H11NO2 CAS:	View	
35	117.0790	0	L-Valine Formula:C5H11NO2 CAS:72-18-4	View	

63884	117.0790	0	4-Methylaminobutyrate Formula:C5H11NO2 CAS:1119-48-8	NO	
35941	117.0790	0	4S-aminopentanoic acid Formula:C5H11NO2 CAS:	NO	
35942	117.0790	0	4-amino-pentanoic acid Formula:C5H11NO2 CAS:	NO	
35949	117.0790	0	2S-amino-pentanoic acid Formula:C5H11NO2 CAS:	NO	
35940	117.0790	0	4R-aminopentanoic acid Formula:C5H11NO2 CAS:	NO	

Metabolite Masses

- Metabolites have a rather limited mass range (compared to peptides)
- They cluster mostly in a mass range usually not relevant to proteomics (proteomics: 500+ Da)
- Most metabolites have between 50-100 atoms and a mass between 300-500 Da



METLIN MS/MS Search

- METLIN provides search options based on
 - Accurate mass
 - Return all metabolites matching a specific mass
 - Tandem spectra
 - Return all metabolites with a matching tandem spectrum
- As with HMDB, spectra cannot be downloaded in bulk, image display only



Scripps Center For Metabolomics

METLIN: Metabolite and Tandem MS Database

[MS HOME](#)[Overview](#)[Search](#)[Software/Services](#)[Metabolomics Science](#)[Publications](#)

MS/MS Spectrum Match

[Simple](#) | [Advanced](#) | [Batch](#) | [Fragment](#) | [Multiple Fragment](#) | [Neutral Loss](#) | [MS/MS Spectrum Match](#) | [Unknowns](#)

m/z, intensity

Peaks:
(MAX: 30 peaks)

EXAMPLE DATA[POSITIVE](#)[NEGATIVE](#)

Mode:

 Positive Negative

Collision Energy (eV):

20eV

Tolerance MS/MS (Da):

0.01

Tolerance Precursor (ppm)

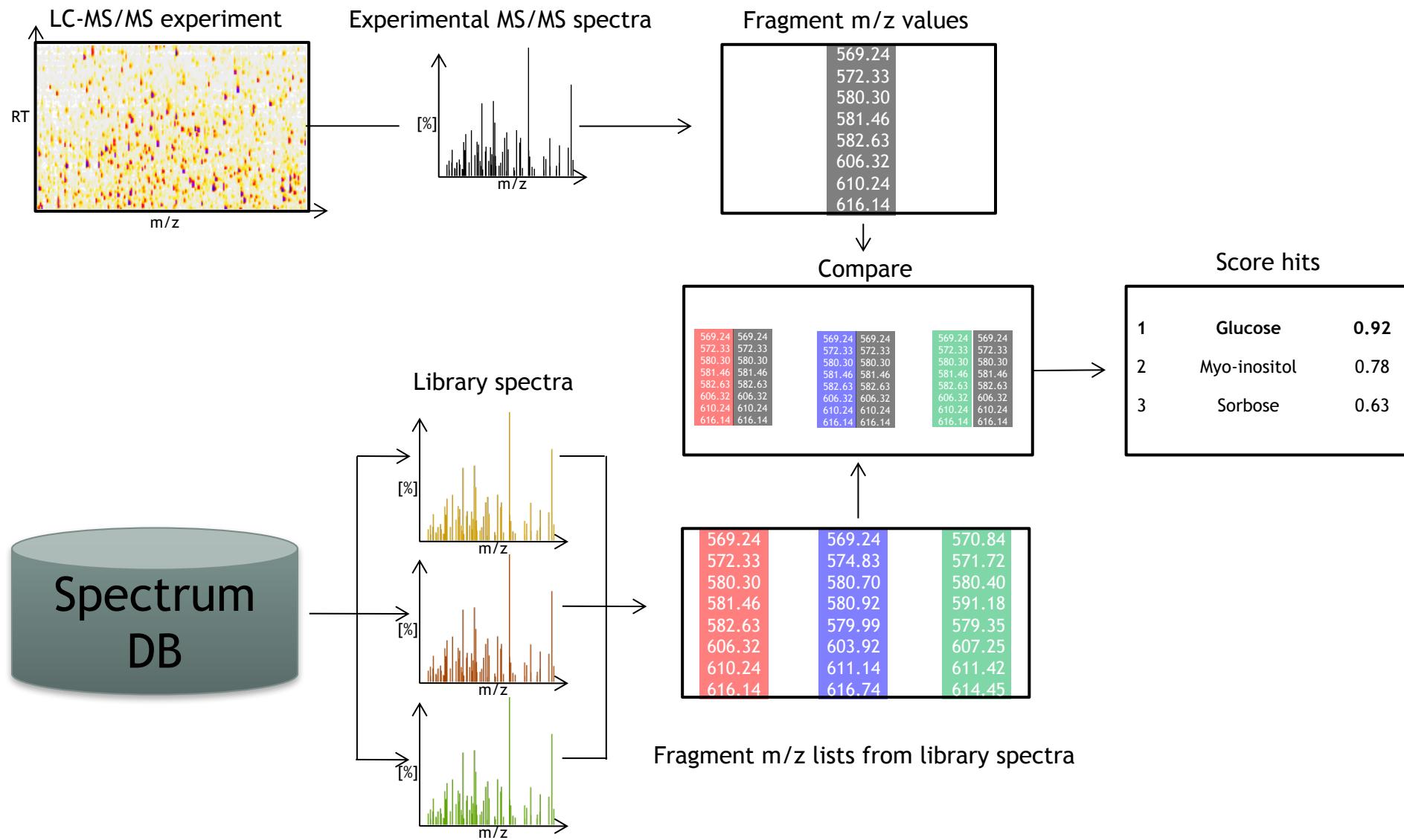
20

Precursor m/z

195.0877

[Find Metabolites](#)[Reset](#)

Spectral Library Search



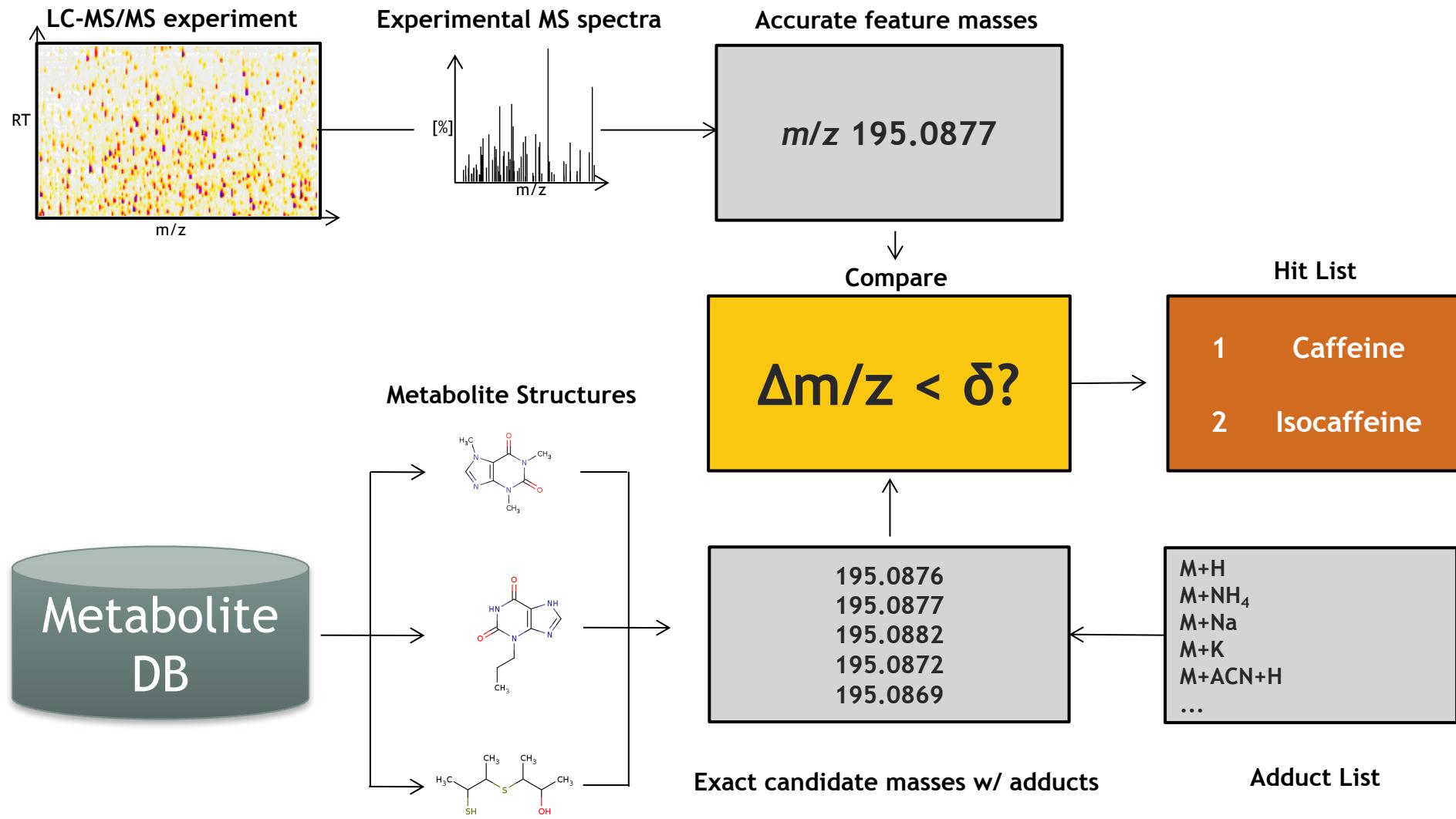
Spectral Library Search

- Searching spectra against a library is done with similarity functions similar to those used in peptide ID (e.g., correlation, shared peak count)
- Success of the spectrum depends on similar experimental conditions:
 - Instrument type
 - Instrument resolution
 - Fragmentation method and parameters (e.g., collision energy)
- Acquiring a library of spectra requires acquisition of spectra, which in turn requires the physical ownership of reference metabolites
- Not all metabolites are readily commercially available

Spectral Library Search

- Key issues
 - Metabolite spectra often less information-rich than peptide spectra
 - Few readily available (downloadable!) databases of metabolite spectra
 - Coverage of spectra thus much worse than in proteomics – we usually can identify only a negligible fraction of the metabolome
 - Fragmentation differs drastically between instruments, depends on collision energy

Accurate Mass Search



Accurate Mass Search

- If spectra are not available for a compound, accurate mass search can be used
- The metabolite m/z differs from the metabolite monoisotopic mass:
 - Ionization adds or removes proton(s)/electron(s)
 - Metabolites often appear as adducts
 - Addition of counter ions (sodium, ammonium)
 - Solvent adducts (acetonitrile, methanol)
- Searching each for each compound with multiple possible adducts increases search space
- Search critically depends on instrument mass accuracy (recalibration helps)

Accurate Mass Search w/ METLIN



Scripps Center For Metabolomics
METLIN: Metabolite and Tandem MS Database

MS HOME

Overview

Search

XCMSOnline

Software/Services

Metabolomics Science

Publications

METLIN: Metabolite Search

Simple

[Simple \(Saved Searches\)](#) | [Advanced](#) | [Batch](#) | [Fragment](#) | [Neutral Loss](#) | [MS/MS Spectrum Match](#) | [Unknowns](#)

Mass:
Tolerance (\pm):
Charge:

195.0877
30 ppm

Neutral
Positive
Negative

M+H
M+NH4
M+Na
M+H-2H2O
M+H-H2O
M+K
M+ACN+H
M+ACN+Na
M+2Na-H
M+2H
M+3H
M+H+Na
M+2H+Na
M+2Na
M+2Na+H
M+Li
M+CH3OH+H

Remove peptides from search:

[Find Metabolites](#) [Reset](#)

• To select multiple Adducts:
Windows - Hit Ctrl + Adducts
Mac OS X - Hit Command + Adducts
Select: [all](#) | [none](#)

METLIN Login

You can use your XCMS Online login.

Simple mass spectrometry data processing.

[Register >](#)

Current Users:

E-mail:
(e.g. researcher@scripps.edu)

Password:

[Sign In](#)

[Forgot your password?](#)

v. c1.1 beta

Accurate Mass Search w/ METLIN

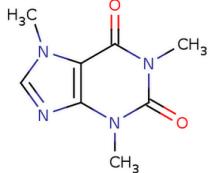
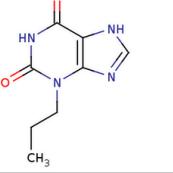
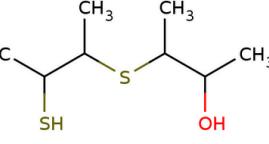


METLIN Metabolites

Mass 195.0877 with 30 ppm mass accuracy

[Change Query](#)

Total: 18 Metabolites

METLIN ID	MASS	Δppm	NAME	MS/MS	STRUCTURE
1455	[M+H] ⁺ m/z 195.0877 M 194.0804	0	Caffeine <i>Formula: C8H10N4O2</i> <i>CAS: 58-08-2</i>	View	
85437	[M+H] ⁺ m/z 195.0877 M 194.0804	0	Enprofylline <i>Formula: C8H10N4O2</i> <i>CAS: 41078-02-8</i>	NO	
92244	[M+H] ⁺ m/z 195.0872 M 194.0799	2	3-[(2-Mercapto-1-methylpropyl)thio]-2-butanol <i>Formula: C8H18OS2</i> <i>CAS: 54957-02-7</i>	NO	

Accurate Mass Search w/ METLIN

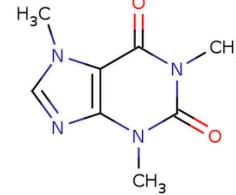


METLIN Metabolites

Mass 195.0877 with 1 ppm mass accuracy

[Change Query](#)

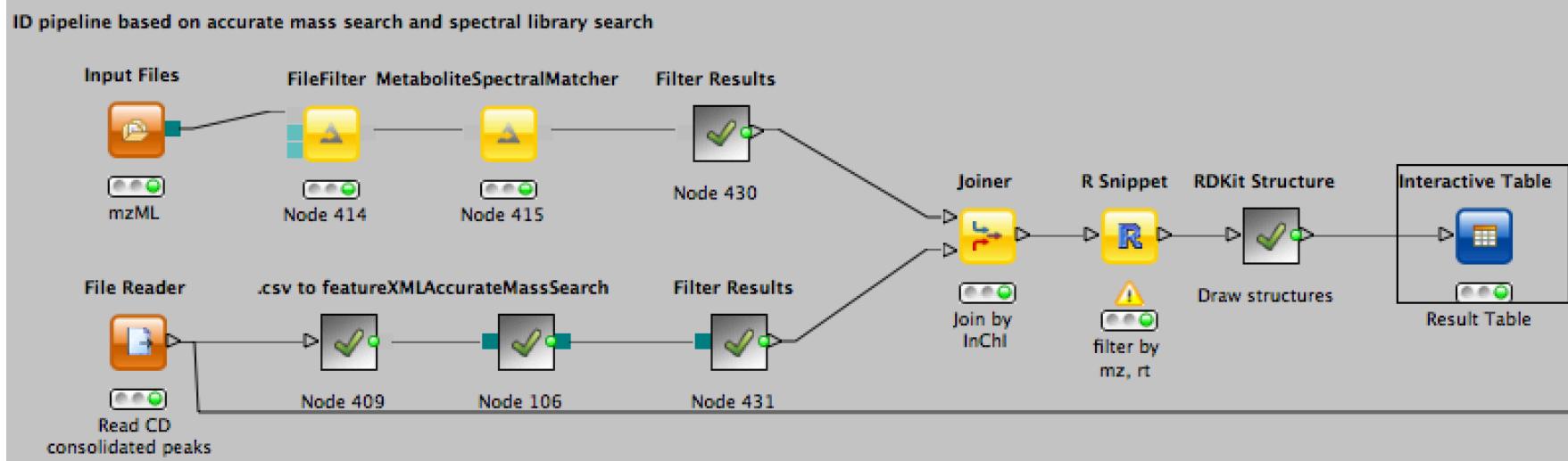
Total: 2 Metabolites

METLIN ID	MASS	Δppm	NAME	MS/MS	STRUCTURE
1455	[M+H] ⁺ <u>m/z</u> 195.0877 M 194.0804	0	Caffeine <i>Formula: C₈H₁₀N₄O₂</i> <i>CAS: 58-08-2</i>	View	
85437	[M+H] ⁺ <u>m/z</u> 195.0877 M 194.0804	0	Enprofylline <i>Formula: C₈H₁₀N₄O₂</i> <i>CAS: 41078-02-8</i>	NO	

Accurate Mass Search - Caveats

- Limiting factors are
 - **Instrument accuracy**: big mass error implies huge search space
 - **Database completeness**: we still search against a (metabolite structure) database – compounds need to be known to be identifiable
 - Isobaric ambiguities: isobaric metabolites cannot be distinguished
- Consequence: identified metabolites are validated again experimentally (standard compounds, reference spectra) whenever possible

Metabolite ID



Multiple ID strategies

- Accurate mass
- Retention time database
- Retention time prediction
- Spectral matching

KNIME provides

- Online access to structure databases
- Structure visualization
- Cheminformatics
 - Metabolization
 - Substructure search

Table View - 0:427 - Interactive Table(Result Table) (74 x 11)					
Row ID	mas...	retenti...	description.ams	identifier	RDKit Mol
Row0	184.061	504.25	4-Pyridoxic acid	HMDB000...	
Row1	245.095	752.3	Biotin	HMDB000...	
Row10	170.082	412.65	Pyridoxine	HMDB002...	
Row11	377.146	732.5	Riboflavin	HMDB002...	

FRAGMENTATION TREES

- Fragmentation tree concepts
- Algorithmic approaches
- Software packages

This work is licensed under a Creative Commons Attribution 4.0 International License.



Fragmentation Trees

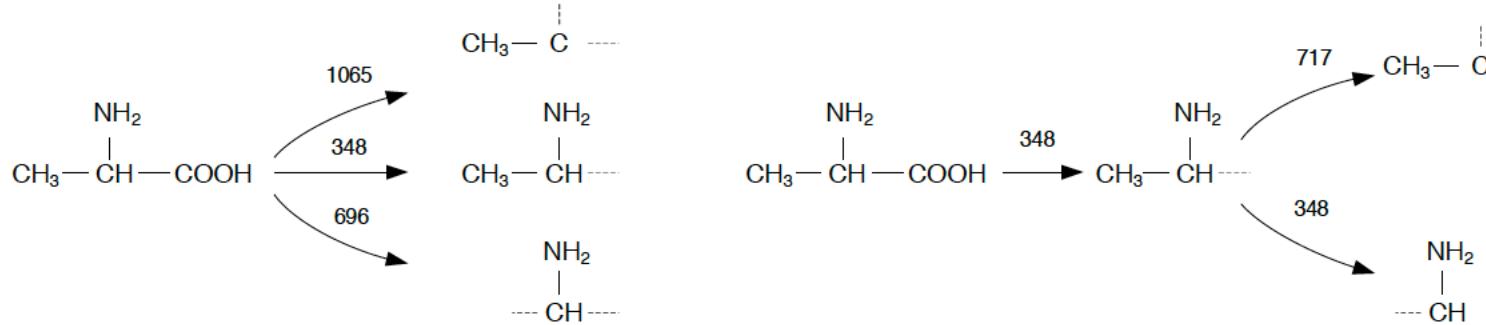
- Fragmentation trees are directed acyclic graphs relating the molecular structure of a metabolite to potential fragment masses
- **Idea**
 - Fragmentation means breaking bonds in the metabolite
 - By systematically and repeatedly breaking all bonds in a structure and its fragments, one can construct the set of all possible fragments
 - Fragmentation occurs preferentially along low-energy dissociation pathways
- **Problems**
 - Not all fragments will be observed
 - Gas phase fragmentation chemistry is very complex
 - Rearrangement reactions lead to different structures
 - Neutral losses are not covered by this
 - Cycles in molecular graphs pose problems

Fragmentation Trees

- There are different ways for constructing fragmentation trees
 - Decomposition of molecular structures
 - Decomposition of molecular formulas
- The first method requires – and leverages – information from available molecular structures
- The second approach relies solely on sufficiently accurate masses
- We will first discuss the approach based on structures as suggested by Heinonen (2006) and then discuss the algorithmic strategies of the second approach by Böcker & Rasche (2008)

Structural Fragmentation Trees

- Heinonen et al. suggest to start with a molecular graph $G = (V, E)$ where V represent atoms and E bonds between atoms and decompose this graph by breaking bonds, i.e. removing edges from G



- Any connected subgraph of $G = (V, E)$ is a possible fragment that can be formed from G
- More formally:

The **fragmentation tree** $G_F = (F, E_F, c)$ of G is an **directed acyclic graph** where

- F is the set of nodes corresponding to all fragments that can be formed of G (i.e., all possible connected subsets $E' \subseteq E$)
- E_F is the set of directed edges from each fragment in F to its subfragments (i.e., fragments that are themselves connected subsets of a fragment or the original molecule)
- $c: E_F \rightarrow \mathbb{R}$ associates each edge with a cost for forming this fragment

Structural Fragmentation Trees

- Source molecules can be fragmented in a single step (yielding MS^2 spectra) or iteratively until no further fragments can be found (yielding multiple fragmentations as in MS^n spectra)
- The cost function models the likelihood for forming a particular fragment
- Cost function can be modeled by bond energies (more stable bonds are more ‘expensive’, less likely to break)
- The best explanation for a tandem spectrum (or for fragment masses observed across multiple fragmentation experiments) should thus be the **most likely fragmentation pattern**

Structural Fragmentation Trees

- **Identification**
 - Given a spectrum $S = (s_1, \dots, s_k)$ with distinct peak masses s_i
 - Find the lightest connected subtree G_F^* of G_F of the fragmentation tree that explains all masses s_i in the spectrum
- This can be repeated for a set of possible structures and results can be scored
- The lightest subtree needs to break the fewest/weakest bonds and still explains all fragments
- Masses are represented by internal nodes as well as leaves of G_F^*
- It has to be connected, because the ions represented by leaves cannot be formed if the fragments represented by internal nodes are not formed

Structural Fragmentation Trees

- Heinonen *et al.* suggest that lightest fragmentation trees can be found using mixed integer linear programming (MILP) – for details see their paper
- They tested the method on a small set of metabolite spectra
- For the majority of metabolites, their method could explain 80-90% of the fragments
- Their method was not used to perform a thorough identification benchmark, though

Compositional Fragmentation Trees

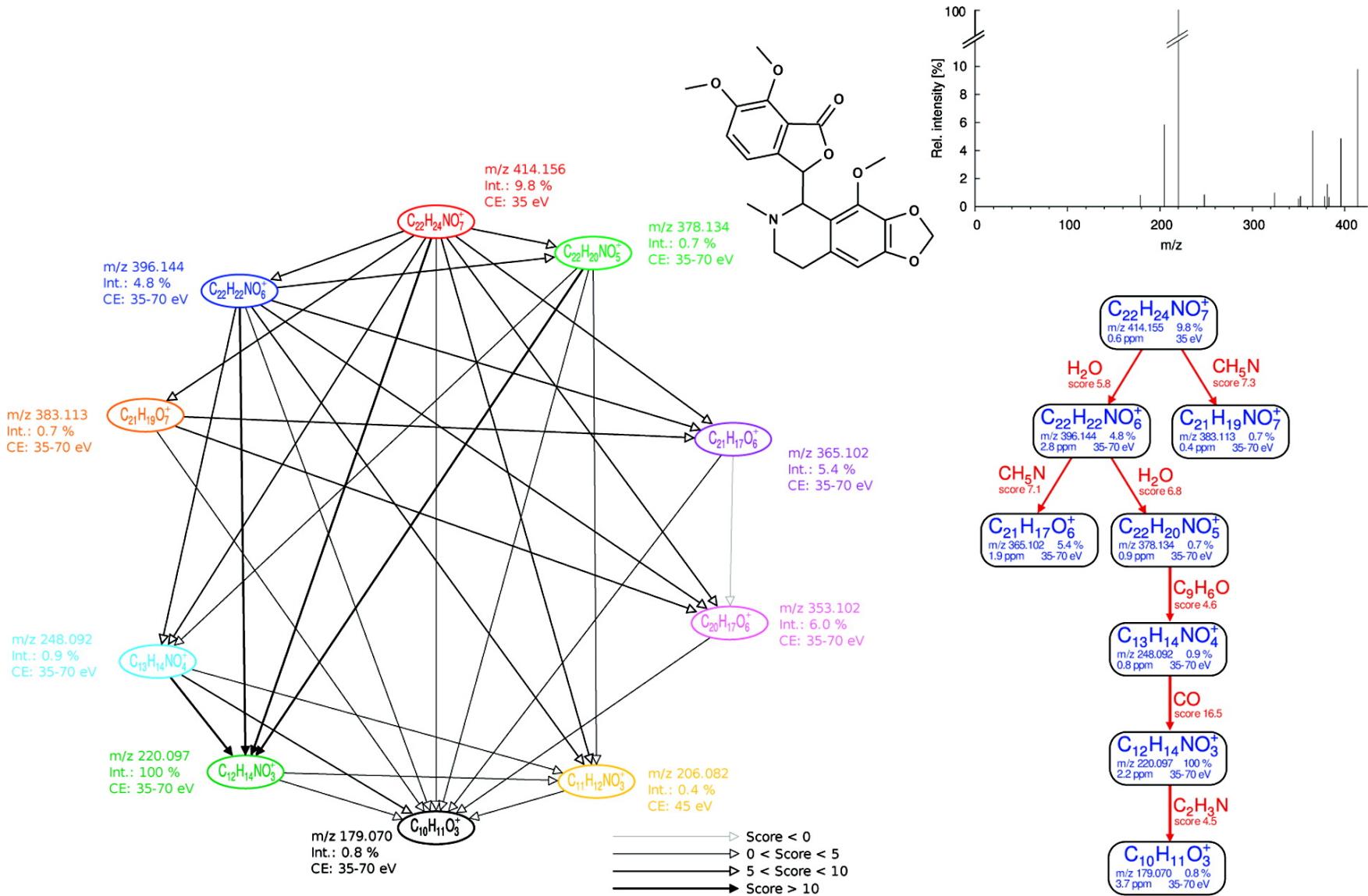
- Heinonen's structural fragmentation trees are the metabolomics equivalent of protein database search: a set of structures is fragmented and theoretical fragmentation trees can then be compared to an experimental spectrum
- **Compositional fragmentation trees** can be used in a way similar to *de novo* identification in proteomics
- The idea is based on the concept of **mass decomposition**
 - A fragment/molecule mass has to correspond to the molecular mass of an existing chemical structure
 - It thus has to correspond to the sum of integer multiples of atomic masses
 - Given a mass measurement accuracy δ , each measured fragment mass m_i has to satisfy

$$m_i - \delta \leq \sum_i a_i c_i \leq m_i + \delta$$

for monoisotopic atomic masses a_i of all elements involved and $c_i \in \mathbb{N}_0$

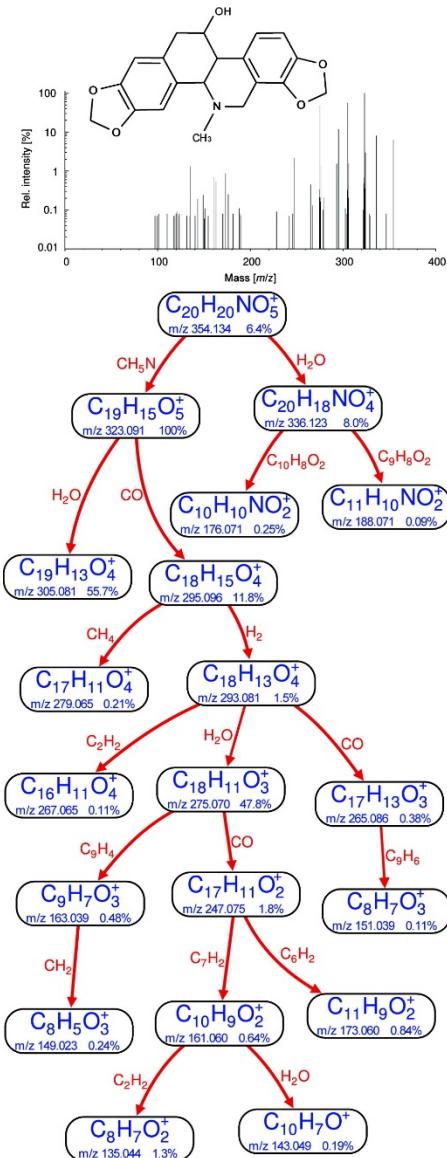
- Coefficients c_i define the composition of the fragment/molecule, i.e. its chemical formula (number of atoms of each element)

Compositional Fragmentation Trees

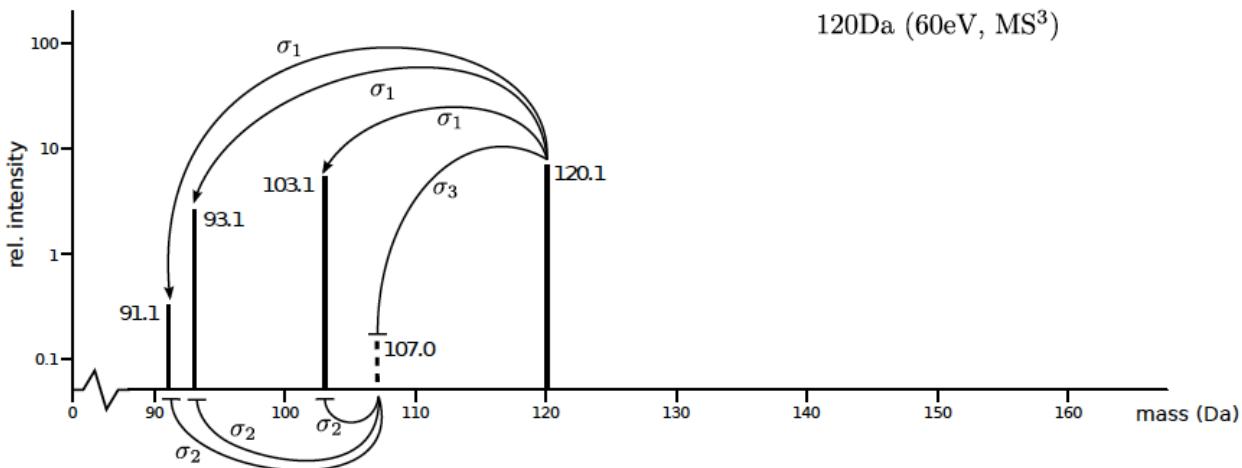
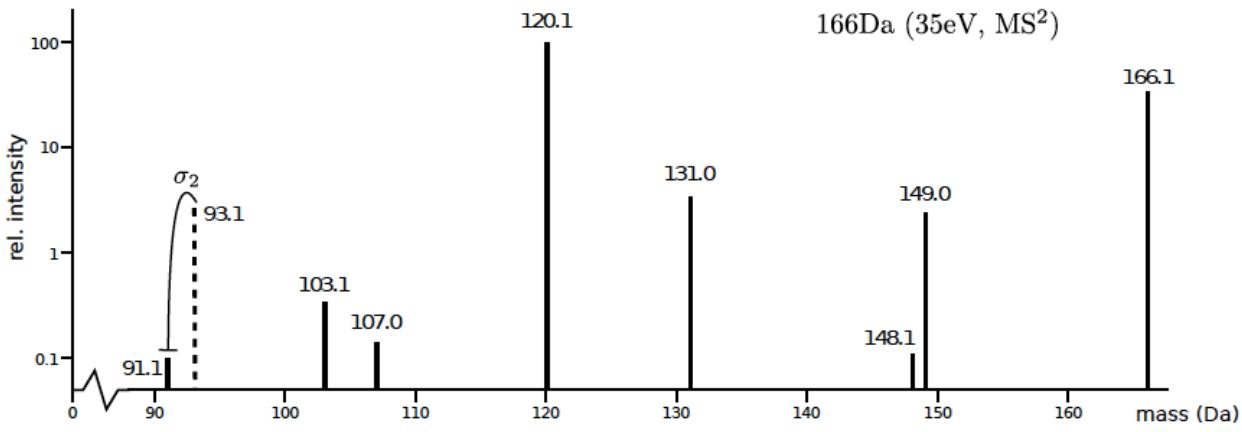
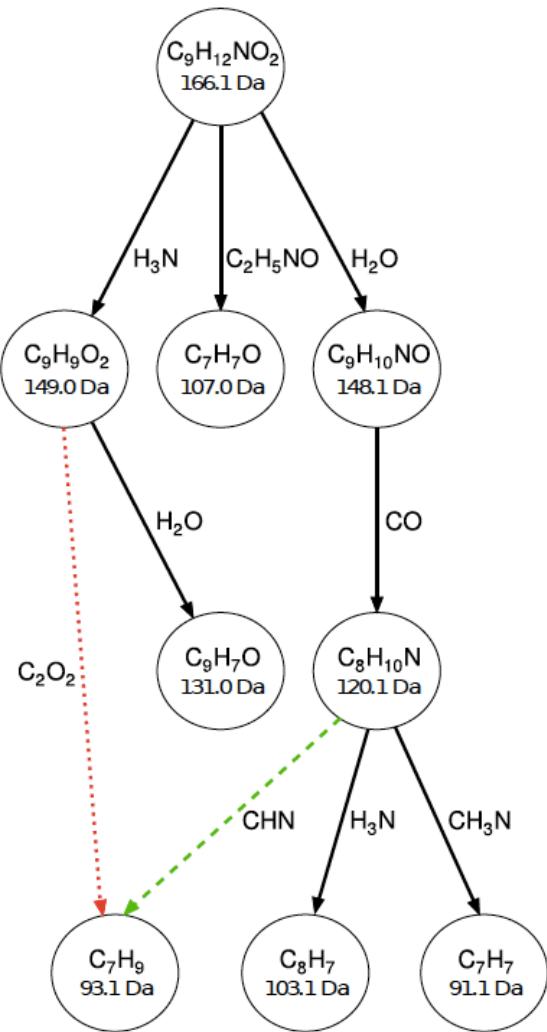


Compositional Fragmentation Trees

- Nodes represent compositions and thus many nodes may represent the same peak
- Multiple nodes can thus have the same color
- The graph is transitive: $(u, v), (v, w) \in E$ also implies $(u, w) \in E$
- Information from multiple spectra (MS^n or MS^2 at varying collision energies) can be freely combined

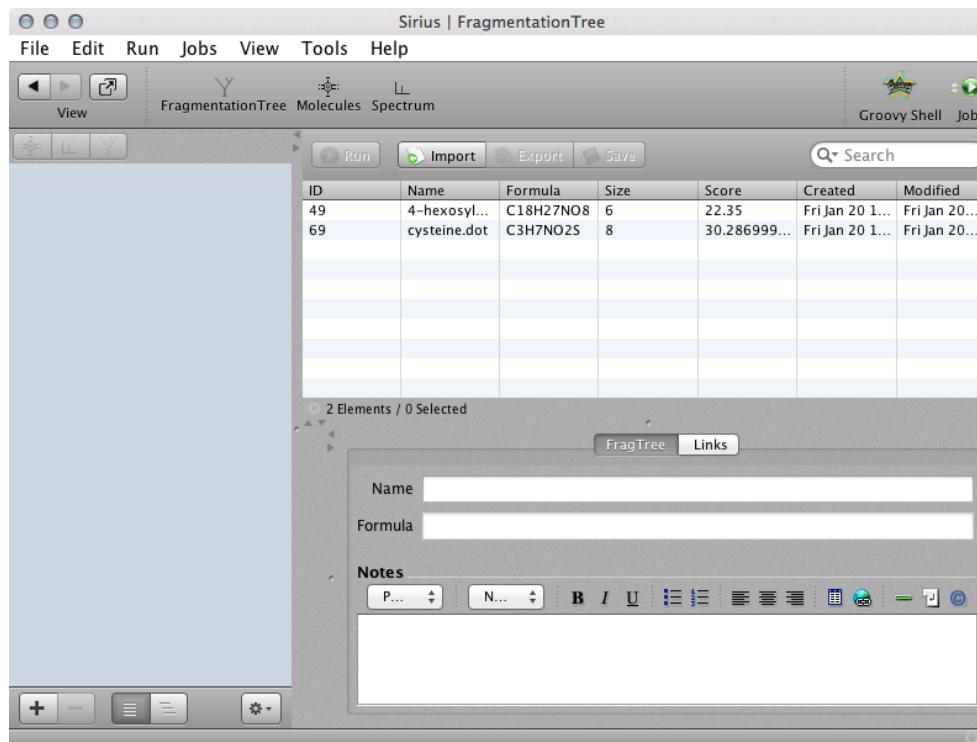


Compositional Fragmentation Trees

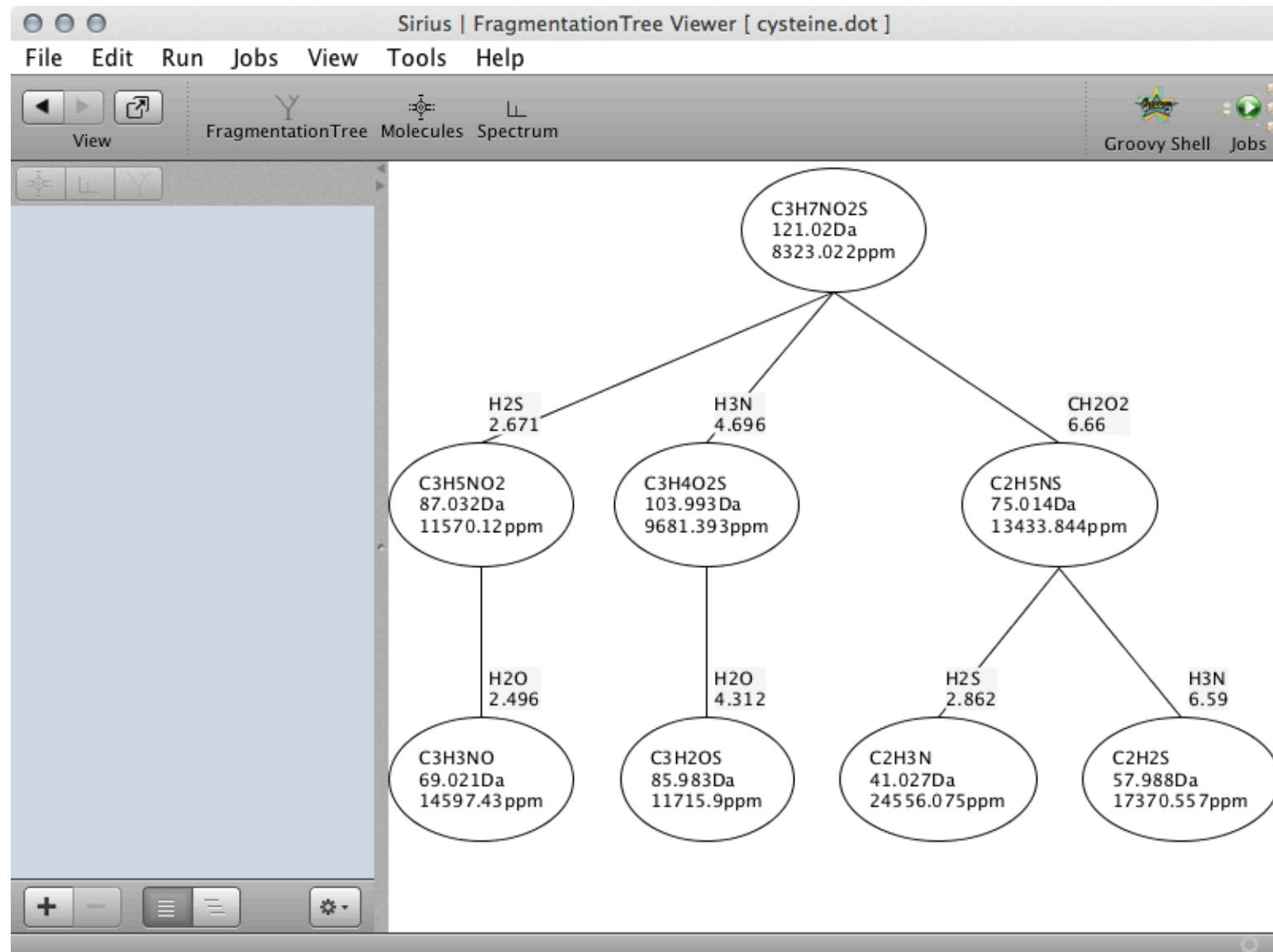


Sirius²

- Sirius² is a software solution implementing fragmentation tree approaches for metabolite identification developed by Sebastian Böcker's group in Jena
- It permits the calculation of theoretical fragmentation trees from sets of spectra and a scoring of fragmentation trees against spectra and structures
- The scoring is based a modified Bayesian formulation for fragmentation trees



Sirius²



Welcome to MoNA!

MassBank of North America (MoNA) is a metadata-centric, auto-curating repository designed for efficient storage and querying of mass spectral records. It intends to serve as a framework for a centralized, collaborative database of metabolite mass spectra, metadata and associated compounds. MoNA currently contains over 200,000 mass spectral records from experimental and in-silico libraries as well as from user contributions.

MoNA has recently been redesigned, with significant improvements to server-side architecture, query structure, and search speed. We are actively improving and adding features, so please be patient as functionality is added. If you notice any major issues, feel free to report them using the issue tracker linked below.

[Search Spectra](#)[Browse Spectra](#)[Issue Tracker](#)

Cyclopamine Score: ★★★★★☆☆☆☆☆

The mass spectrum shows a base peak at m/z 204 and several other significant peaks at lower m/z values, such as 182, 196, and 220.

The chemical structure is a tricyclic alkaloid with a complex arrangement of rings, hydroxyl groups, and methyl groups.

Q accession	BSU00002
Q authors	Chandler, C. and Habig, J
Q compound class	Natural Product; Alkaloid
Q copyright	Chandler, C., Habig, J. a
Q exact mass	411.31372
Q instrument	Bruker maXis ESI-QTOF
Q instrument type	LC-ESI-QTOF
Q ion mode	POSITIVE
Q license	CC BY-SA
Q mass accuracy	50.94725200101779
Q mass error	-21.005552000019634
Q ms level	MS2
Q origin	BSU00002.txt
Q precursor m/z	412.3
Q precursor type	[M+H] ⁺

[LCMS ▾](#) [massbank ▾](#) [display full information](#)

References

- **Databases**
 - **HMDB**
<http://hmdb.ca>
 - **METLIN**
<http://metlin.scripps.edu>
 - **mzCloud**
<http://mzcloud.org>
 - **MassBank**
<http://massbank.org>
 - **MetaboLights**
<http://www.ebi.ac.uk/metabolights>
- **Papers**
 - Hill DW, Kertesz TM, Fontaine D, Friedman R, Grant DF. Mass spectral metabonomics beyond elemental formula: chemical database querying by matching experimental with computational fragmentation spectra. *Anal Chem.* 2008, 80:5574-82.
 - Heinenen M, Rantanen A, Mielikäinen, Pitkänen E, Kokkonen J, Rousu J, Ab initio prediction of molecular fragments from tandem mass spectrometry data, Proc. German Conference on Bioinformatics (GCB 2006), LNI P-83, 2006, Gesellschaft für Informatik
 - Scheubert K, Hufsky F, Rasche F, Böcker S. Computing Fragmentation Trees from Metabolite Multiple Mass Spectrometry Data, RECOMB 2011 (2011), LNBI 6577, pp. 377-391, Springer, Heidelberg, 2011.
 - Böcker S, Letzel MC, Liptak Z, Pervukhin A. Decomposing Metabolomic Isotope Patterns, WABI 2006, LNBI 4175, pp. 12-23 ,2006, Springer, Heidelberg
 - Böcker S and Rasche F. Towards de novo identification of metabolites by analyzing tandem mass spectra. *Bioinformatics*, 2008, 24:i49-i55.
 - Florian Rasche, Aleš Svatoš, Ravi Kumar Maddula, Christoph Böttcher, and Sebastian Böcker. Computing Fragmentation Trees from Tandem Mass Spectrometry Data. *Analytical Chemistry* (2011) 83 (4): 1243–1251