Metabolomics Databases

Xiuxia Du, Stephen Barnes

Outline

- Comprehensive metabolomics databases
- Compound databases
- Spectral databases
- Metabolic pathway databases
- Drug databases
- Disease & physiology databases
- Raw data databases

Outline

- Comprehensive metabolomics databases
- Compound databases
- Spectral databases
- Metabolic pathway databases
- Drug databases
- Disease & physiology databases
- Raw data databases

Comprehensive databases

- HMDB

HMDB

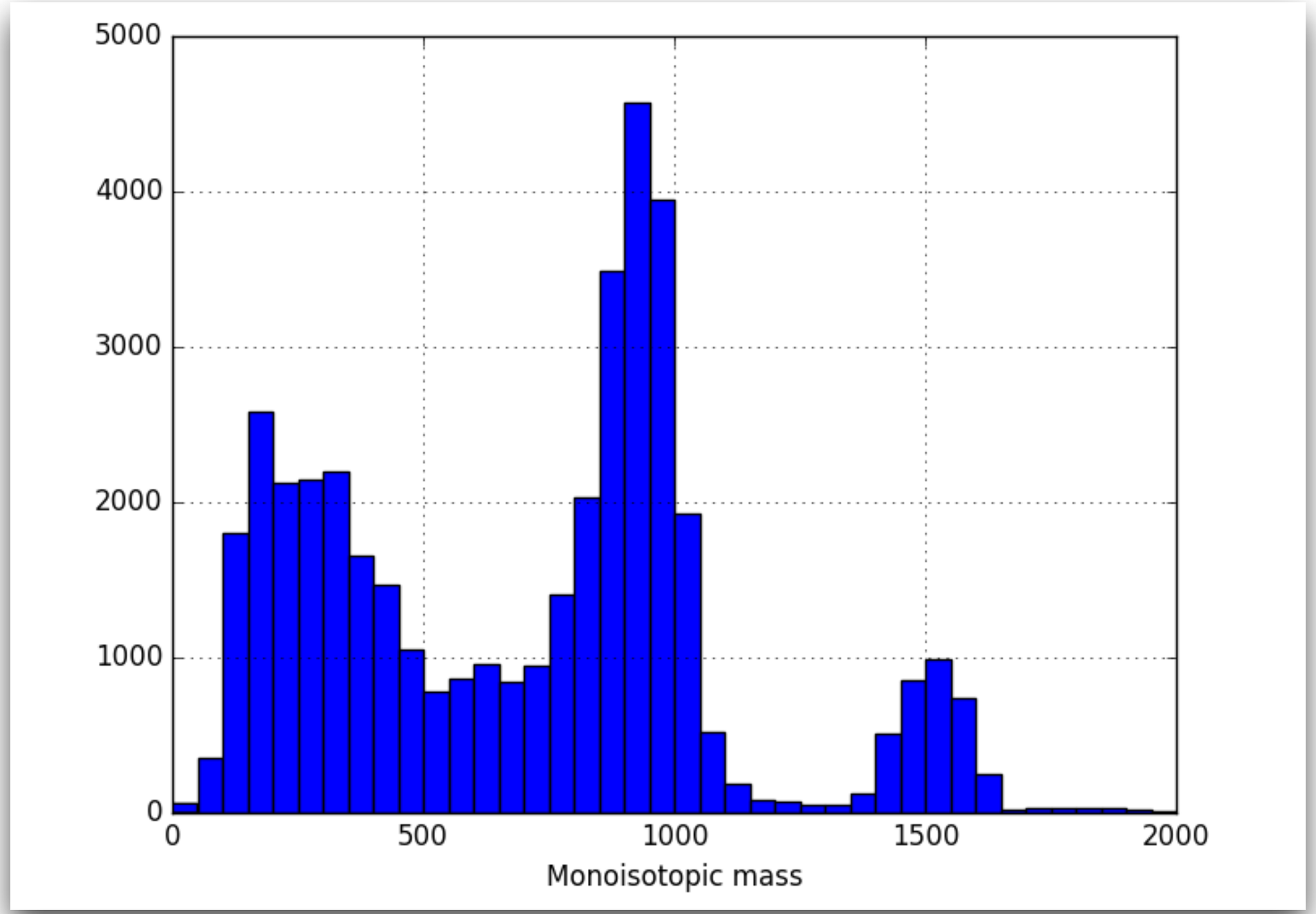
- Overview

Welcome to HMDB Version 3.6

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 contains 41,993 metabolite entries including both water-soluble and lipid soluble metabolites as well as metabolites that would be regarded as either abundant (> 1 uM) or relatively rare (< 1 nM). Additionally, 5,701 protein 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, PDB, UniProt, and GenBank) and a variety of structure and pathway viewing applets. The HMDB database supports extensive text, sequence, chemical structure and relational query searches. Four additional databases, DrugBank, T3DB, SMPDB and FooDB are also part of the HMDB suite of databases. DrugBank contains equivalent information on ~1600 drug and drug metabolites, T3DB contains information on ~3600 common toxins and environmental pollutants, SMPDB contains pathway diagrams for ~700 human metabolic and disease pathways, while FooDB contains equivalent information on ~28,000 food components and food additives.

HMDB

- Breakdown by mass



HMDB

- One metabocard

Showing metabocard for L-Proline (HMDB00162)

Identification | Taxonomy | Ontology | Physical properties | Spectra | Biological properties | Concentrations | Links | References | XML

enzymes (29) | transporters (1) | Show 30 proteins

Show Metabolites with Similar Structures

Record Information	
Version	3.6
Creation Date	2005-11-16 15:48:42 UTC
Update Date	2016-06-10 21:23:05 UTC
HMDB ID	HMDB00162
Secondary Accession Numbers	None
Metabolite Identification	
Common Name	**L-Proline**
Description	L-Proline is one of the twenty amino acids used in living organisms as the building blocks of proteins. Proline is sometimes called an imino acid, although the IUPAC definition of an imine requires a carbon-nitrogen double bond. Proline is a non-essential amino acid that is synthesized from glutamic acid. It is an essential component of collagen and is important for proper functioning of joints and tendons.
Structure	

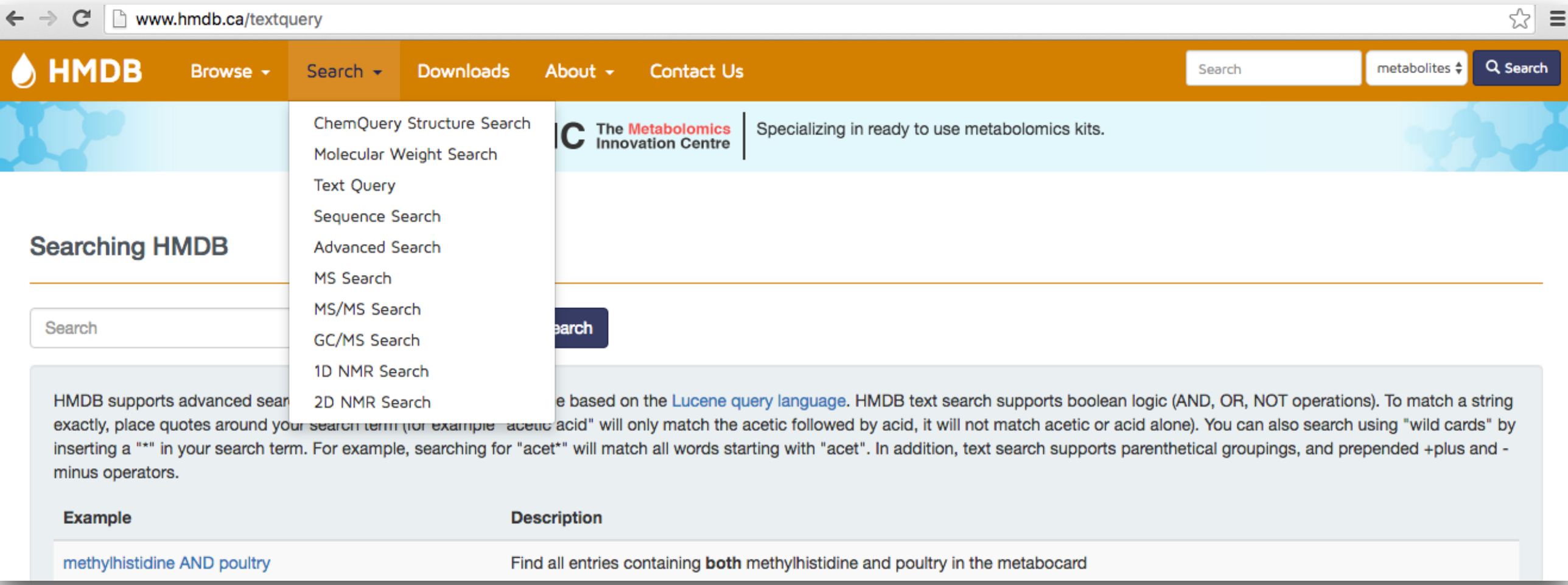
HMDB

- One metabocard

Chemical Formula	$C_5H_9NO_2$
Average Molecular Weight	115.1305
Monoisotopic Molecular Weight	115.063328537
IUPAC Name	(2S)-pyrrolidine-2-carboxylic acid
Traditional Name	L-proline
CAS Registry Number	147-85-3
SMILES	OC(=O)[C@@H]1CCCN1
InChI Identifier	InChI=1S/C5H9NO2/c7-5(8)4-2-1-3-6-4/h4,6H,1-3H2,(H,7,8)/t4-/m0/s1
InChI Key	InChIKey=ONIBWKKTOPOVIA-BYPYZUCNSA-N

HMDB

- Searches



HMDB

- Downloads

Protein/Gene Sequences (in FASTA Format)

Data Set	Released on	Protein Sequences	Gene Sequences
All Metabolite Metabolizing Enzymes	2016-07-10	Download	Download

Structures (in SDF Format)

Data Set	Released on	SDF File
Metabolite Structures	2016-07-10	Download

Metabolite and Protein Data (in XML format)

Data Set	Released on	XML File
All Metabolites	2016-07-10	Download
All Proteins	2016-07-09	Download

Spectra

Data Set	Download Link
Mass Spectra Image Files	Download
GC/MS Peak Lists	Download
NMR Spectra FIDS Files	Download
NMR Spectra Peaklist Files	Download
Spectra information (in XML format)	Download

HMDB

- Download all metabolites

Name	Date Modified	Size	Kind
hmdb_metabolites.xml	Jul 10, 2016, 2:58 AM	664.4 MB	XML
HMDB00001.xml	Apr 13, 2014, 12:50 AM	60 KB	XML
HMDB00002.xml	Apr 13, 2014, 12:50 AM	22 KB	XML
HMDB00005.xml	Apr 13, 2014, 12:50 AM	24 KB	XML
HMDB00008.xml	Apr 13, 2014, 12:50 AM	22 KB	XML
HMDB00010.xml	Apr 13, 2014, 12:50 AM	19 KB	XML
HMDB00011.xml	Apr 13, 2014, 12:50 AM	37 KB	XML
HMDB00012.xml	Apr 13, 2014, 12:50 AM	21 KB	XML
HMDB00014.xml	Apr 13, 2014, 12:50 AM	20 KB	XML
HMDB00015.xml	Apr 13, 2014, 12:50 AM	19 KB	XML
HMDB00016.xml	Apr 13, 2014, 12:50 AM	21 KB	XML
HMDB00017.xml	Apr 13, 2014, 12:50 AM	21 KB	XML
HMDB00019.xml	Apr 13, 2014, 12:50 AM	27 KB	XML
HMDB00020.xml	Apr 13, 2014, 12:50 AM	55 KB	XML
HMDB00021.xml	Apr 13, 2014, 12:50 AM	13 KB	XML
HMDB00022.xml	Apr 13, 2014, 12:50 AM	23 KB	XML
HMDB00023.xml	Apr 13, 2014, 12:50 AM	29 KB	XML
HMDB00024.xml	Apr 13, 2014, 12:50 AM	36 KB	XML
HMDB00026.xml	Apr 13, 2014, 12:50 AM	17 KB	XML
HMDB00027.xml	Apr 13, 2014, 12:50 AM	27 KB	XML
HMDB00030.xml	Apr 13, 2014, 12:50 AM	38 KB	XML
HMDB00031.xml	Apr 13, 2014, 12:50 AM	33 KB	XML
HMDB00032.xml	Apr 13, 2014, 12:50 AM	20 KB	XML
HMDB00033.xml	Apr 13, 2014, 12:50 AM	35 KB	XML
HMDB00034.xml	Apr 13, 2014, 12:50 AM	30 KB	XML
HMDB00036.xml	Apr 13, 2014, 12:50 AM	29 KB	XML

Outline

- Comprehensive metabolomics databases
- Compound databases
- Spectral databases
- Metabolic pathway databases
- Drug databases
- Disease & physiology databases
- Raw data databases

Compound databases

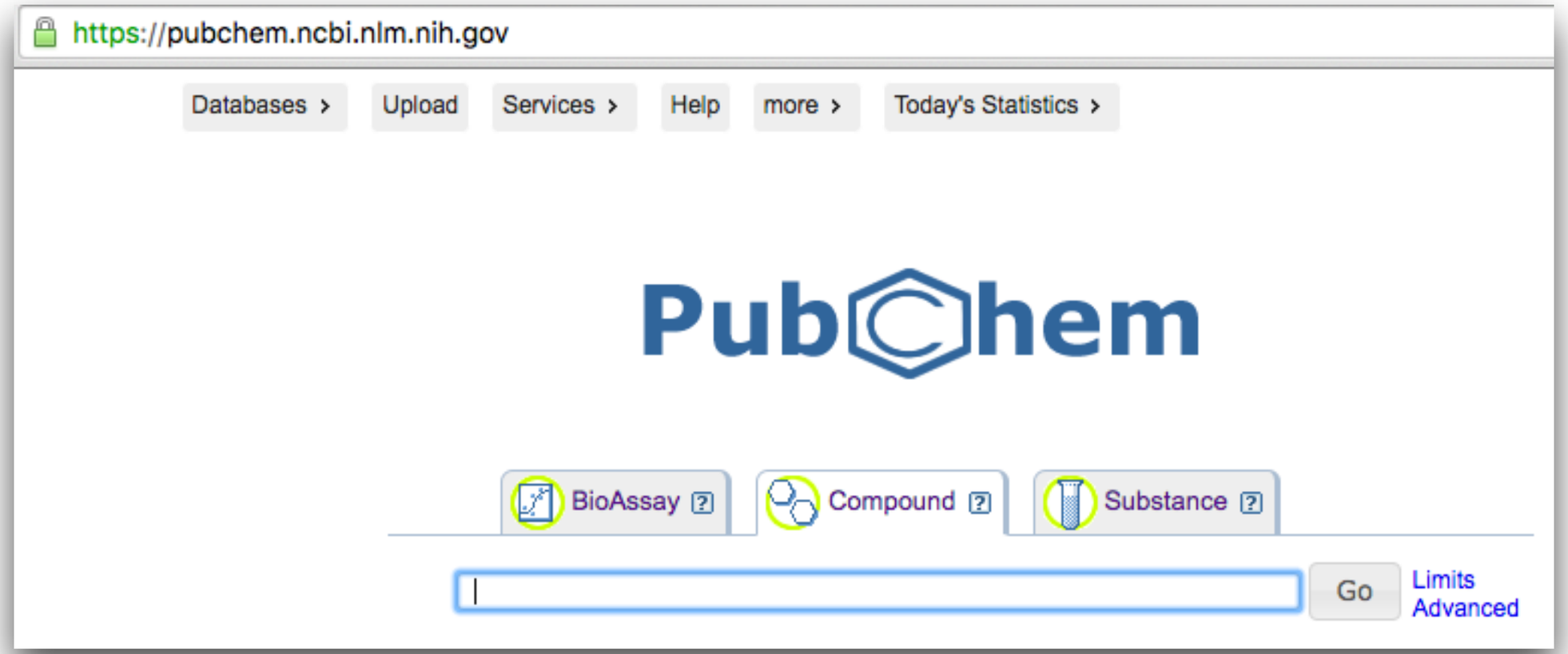
- PubChem
- ChemSpider
- ChEBI
- KEGG Glycan
- IIMDB

Compound databases

- PubChem
- ChemSpider
- ChEBI
- KEGG Glycan
- IIMDB

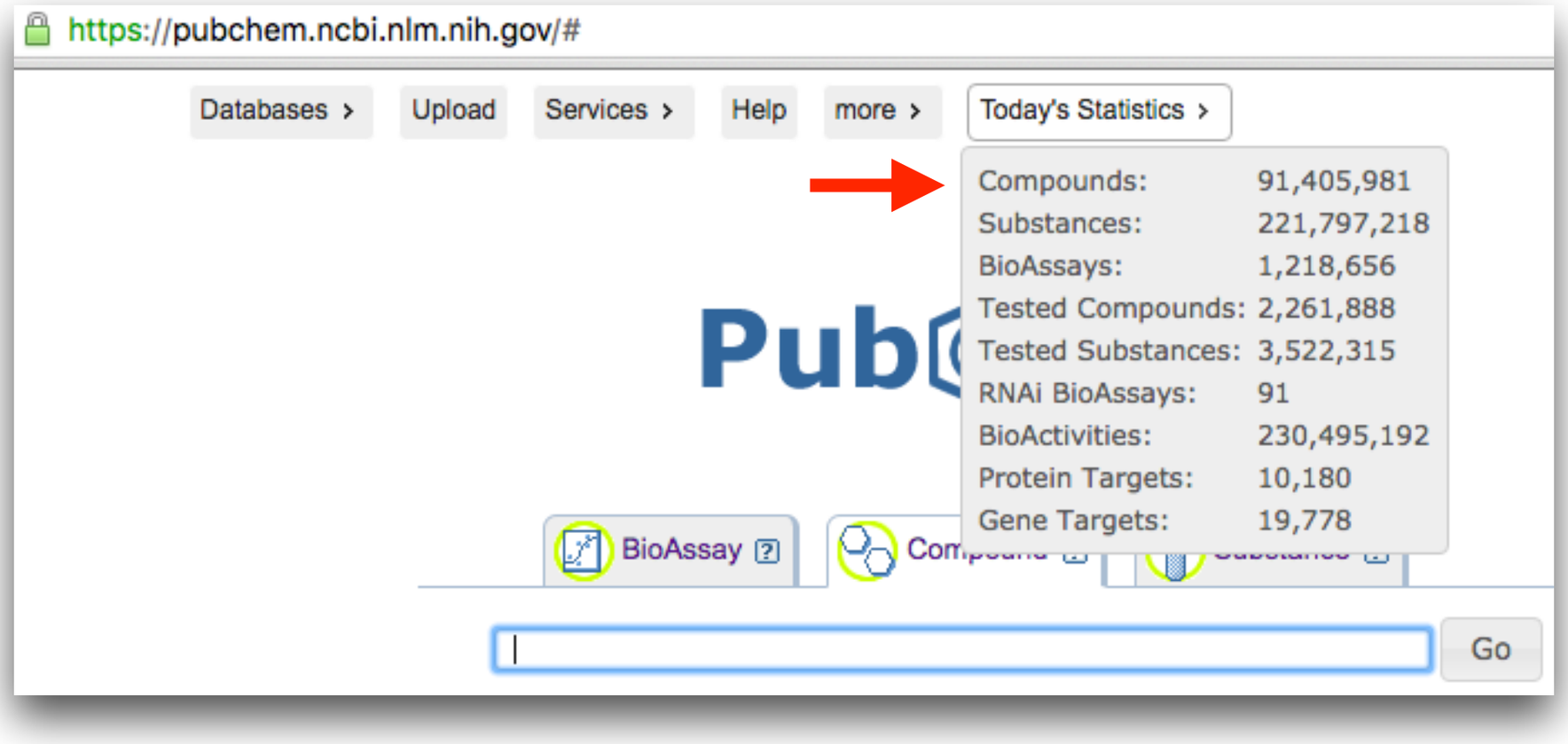
PubChem

- Website



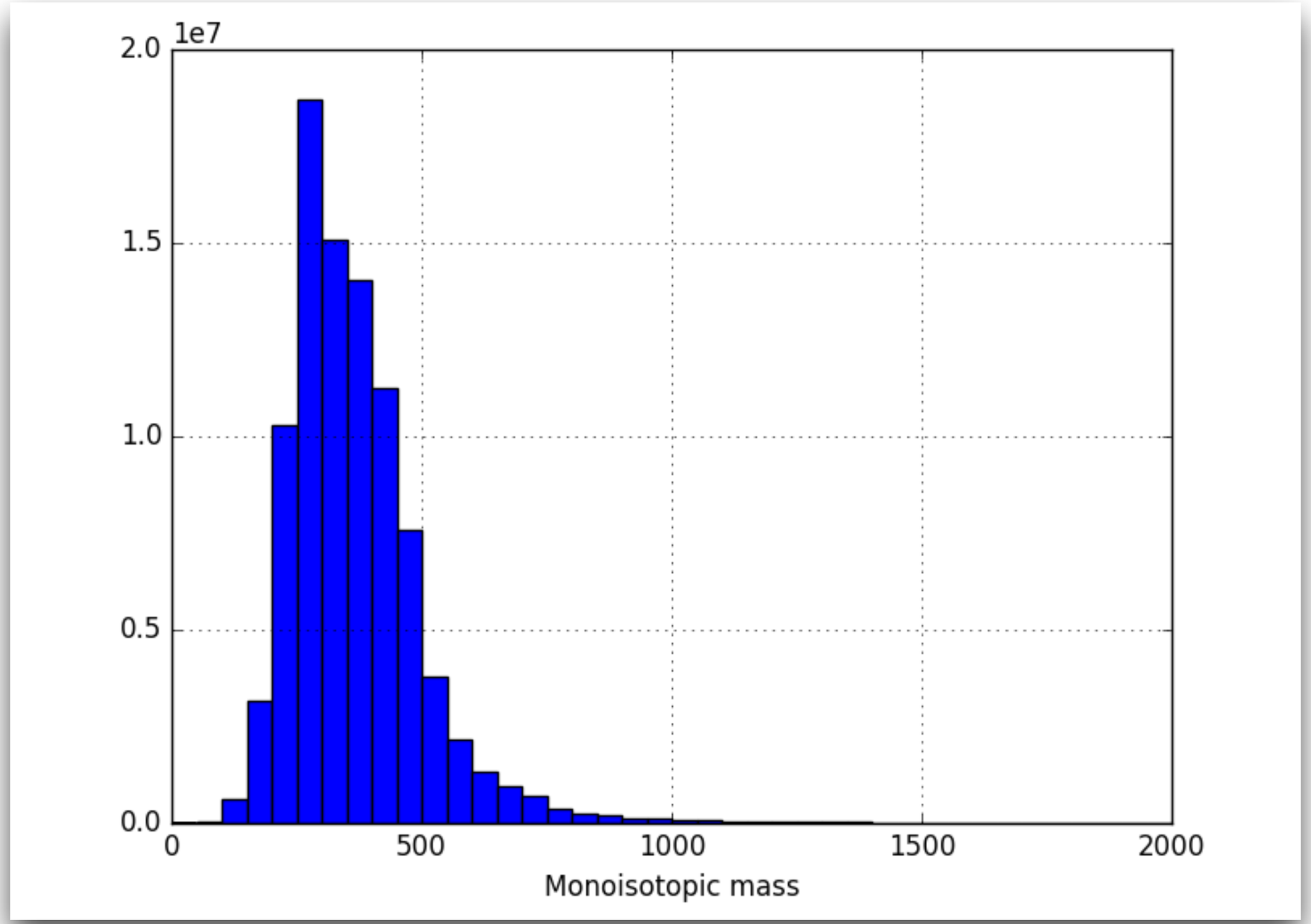
PubChem

- Statistics (July 12, 2016)



PubChem

- Breakdown by mass



PubChem

- Information on one compound

L-proline

Cite this Record

Vendors | Drug Information | Pharmacology | Literature | Patents | Bioactivities

PubChem CID:	145742
Chemical Names:	L-proline; 147-85-3; L-(-)-Proline; Proline; (S)-Pyrrolidine-2-carboxylic acid; (-)-(S)-Proline; More...
Molecular Formula:	$C_5H_9NO_2$
Molecular Weight:	115.13046 g/mol
InChI Key:	ONIBWKKTOPOVIA-BYPYZUCNSA-N
UNII:	9DLQ4CIU6V
Modify Date:	2016-07-09
Create Date:	2004-09-16

L-proline is a non-essential amino acid that is synthesized from GLUTAMIC ACID. It is an essential component of COLLAGEN and is important for proper functioning of joints and tendons.

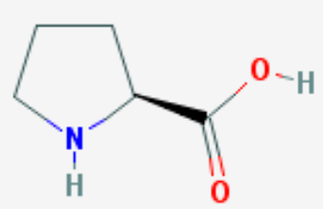
PubChem

- Information on one compound

Contents

1 2D Structure

Search Download Get Image



Magnify

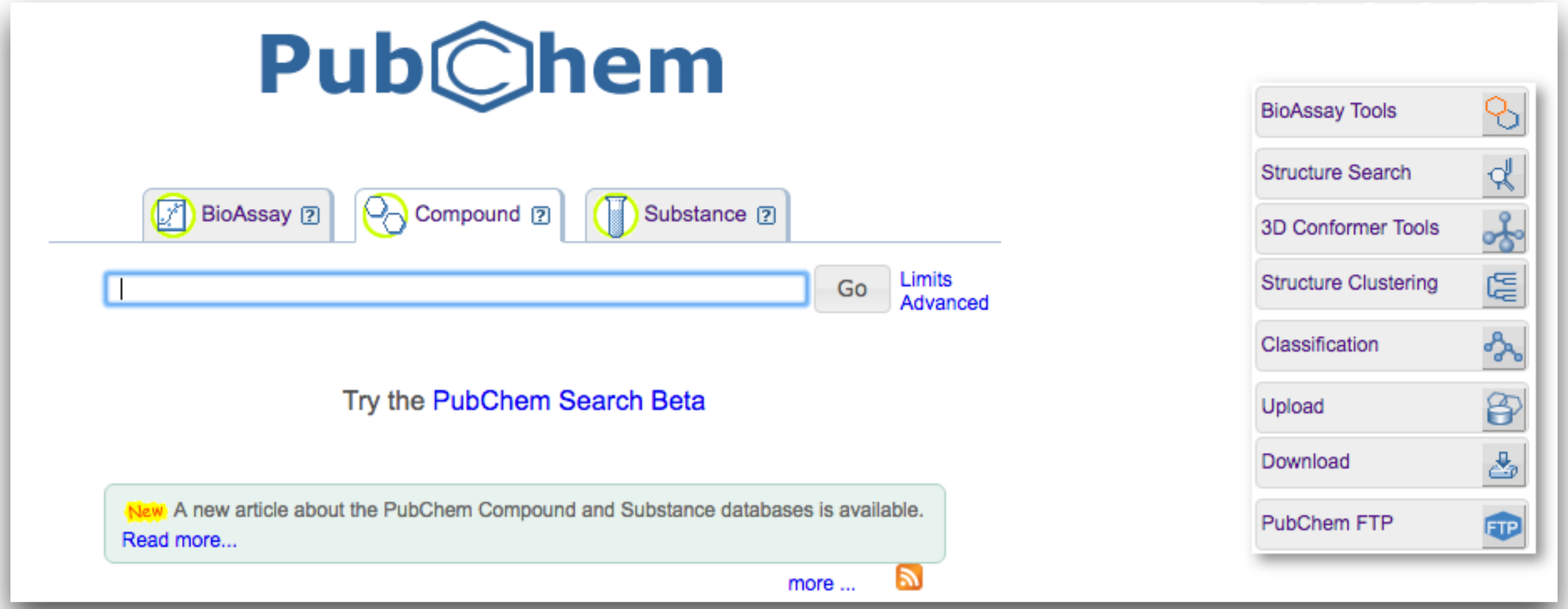
from PubChem

2 3D Conformer

Search Download Get Image

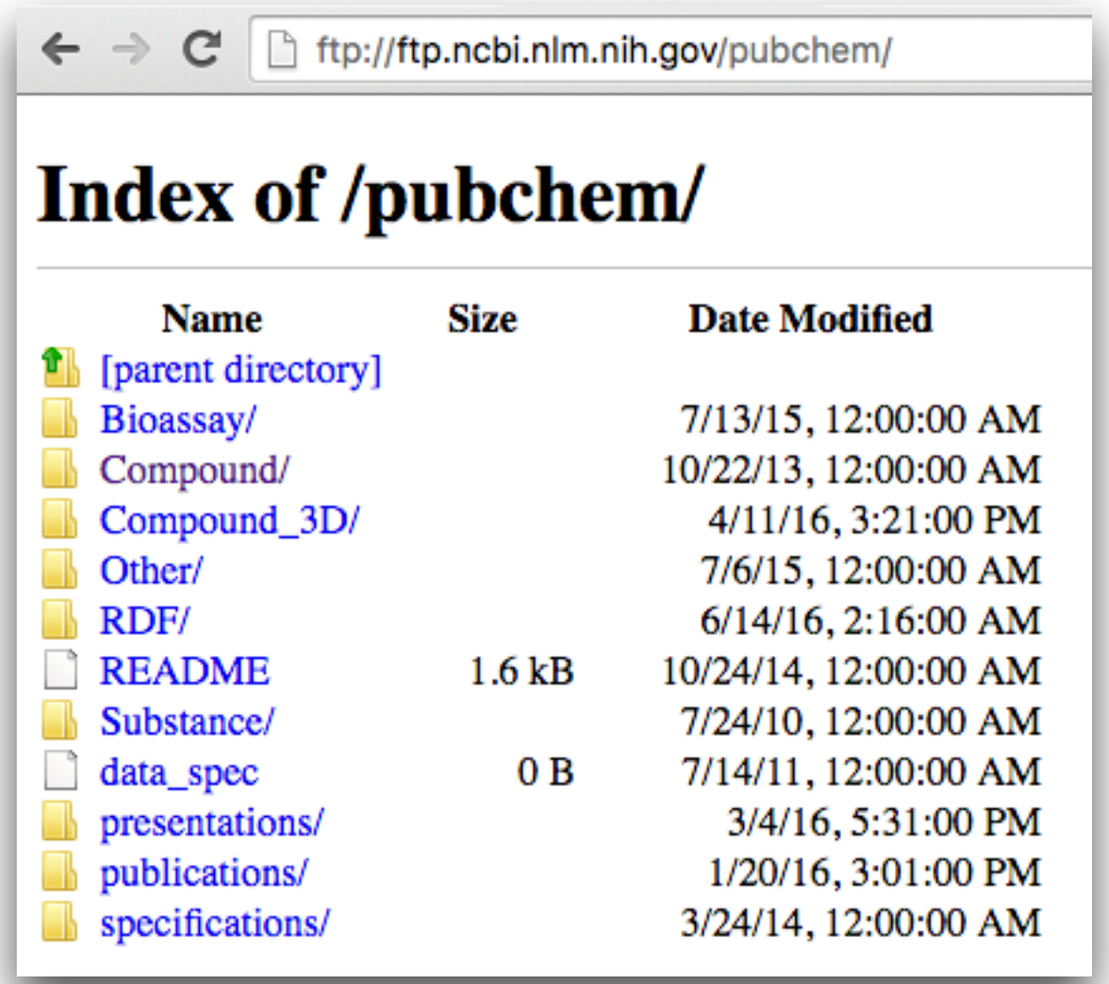
PubChem

- Search and FTP



PubChem

- FTP


ftp://ftp.ncbi.nlm.nih.gov/pubchem/
Index of /pubchem/
Name
Size
Date Modified
[parent directory]
Bioassay/ 7/13/15, 12:00:00 AM
Compound/ 10/22/13, 12:00:00 AM
Compound_3D/ 4/11/16, 3:21:00 PM
Other/ 7/6/15, 12:00:00 AM
RDF/ 6/14/16, 2:16:00 AM
README 1.6 kB 10/24/14, 12:00:00 AM
Substance/ 7/24/10, 12:00:00 AM
data_spec 0 B 7/14/11, 12:00:00 AM
presentations/ 3/4/16, 5:31:00 PM
publications/ 1/20/16, 3:01:00 PM
specifications/ 3/24/14, 12:00:00 AM

PubChem

- Compound XML files

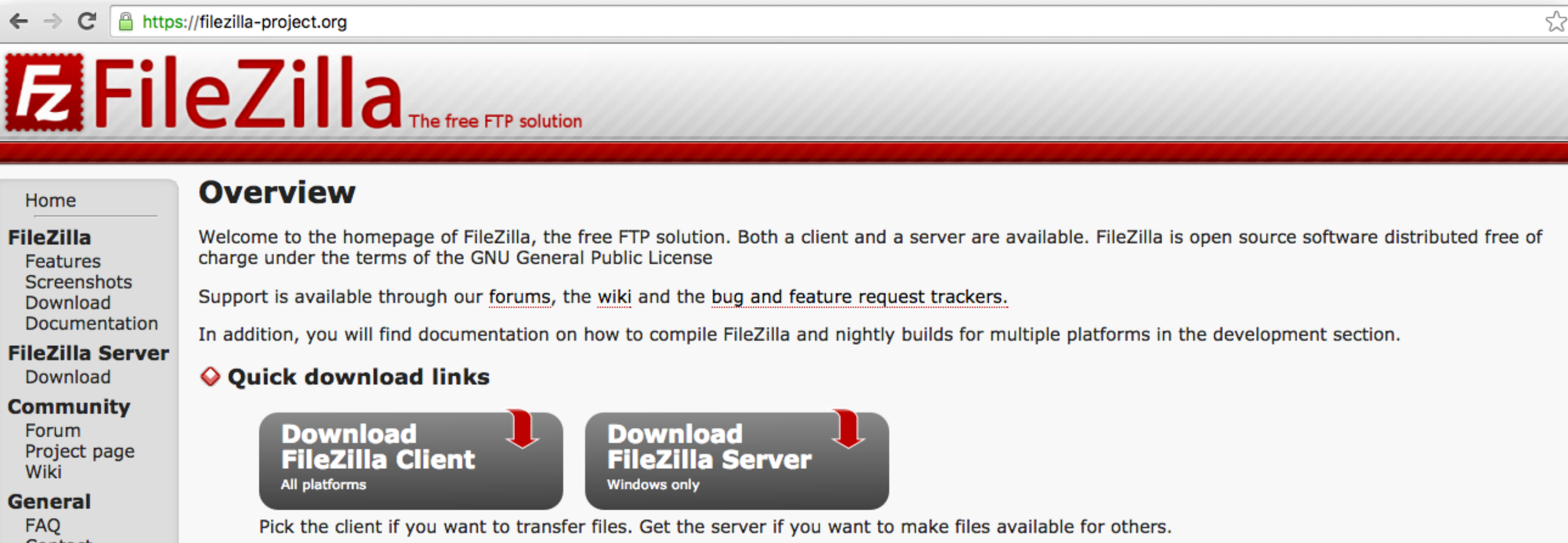
ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound/CURRENT-Full/XML/

Index of /pubchem/Compound/CURRENT-Full/XML/

Name	Size	Date Modified
[parent directory]		
Compound_000000001_000025000.xml.gz	41.4 MB	6/7/16, 11:37:00 AM
Compound_000025001_000050000.xml.gz	46.6 MB	5/20/16, 8:46:00 AM
Compound_000050001_000075000.xml.gz	46.4 MB	6/7/16, 11:35:00 AM
Compound_000075001_000100000.xml.gz	38.7 MB	6/7/16, 11:35:00 AM
Compound_000100001_000125000.xml.gz	48.8 MB	6/7/16, 11:35:00 AM
Compound_000125001_000150000.xml.gz	43.5 MB	6/7/16, 11:35:00 AM
Compound_000150001_000175000.xml.gz	50.1 MB	6/7/16, 11:35:00 AM
Compound_000175001_000200000.xml.gz	52.8 MB	6/7/16, 11:36:00 AM
Compound_000200001_000225000.xml.gz	50.0 MB	4/11/16, 5:49:00 PM
Compound_000225001_000250000.xml.gz	44.4 MB	12/19/15, 12:00:00 AM
Compound_000250001_000275000.xml.gz	46.4 MB	6/7/16, 11:35:00 AM
Compound_000275001_000300000.xml.gz	47.0 MB	12/25/15, 12:00:00 AM
Compound_000300001_000325000.xml.gz	46.9 MB	6/7/16, 11:35:00 AM
Compound_000325001_000350000.xml.gz	48.2 MB	6/7/16, 11:35:00 AM
Compound_000350001_000375000.xml.gz	51.4 MB	6/7/16, 11:35:00 AM
Compound_000375001_000400000.xml.gz	56.9 MB	6/7/16, 11:35:00 AM
Compound_000400001_000425000.xml.gz	51.1 MB	6/15/16, 9:46:00 AM
Compound_000425001_000450000.xml.gz	53.8 MB	6/7/16, 11:35:00 AM
Compound_000450001_000475000.xml.gz	58.2 MB	6/7/16, 11:36:00 AM
Compound_000475001_000500000.xml.gz	62.7 MB	6/7/16, 11:35:00 AM
Compound_000500001_000525000.xml.gz	55.1 MB	2/24/16, 4:21:00 PM
Compound_000525001_000550000.xml.gz	41.5 MB	6/7/16, 11:35:00 AM

PubChem

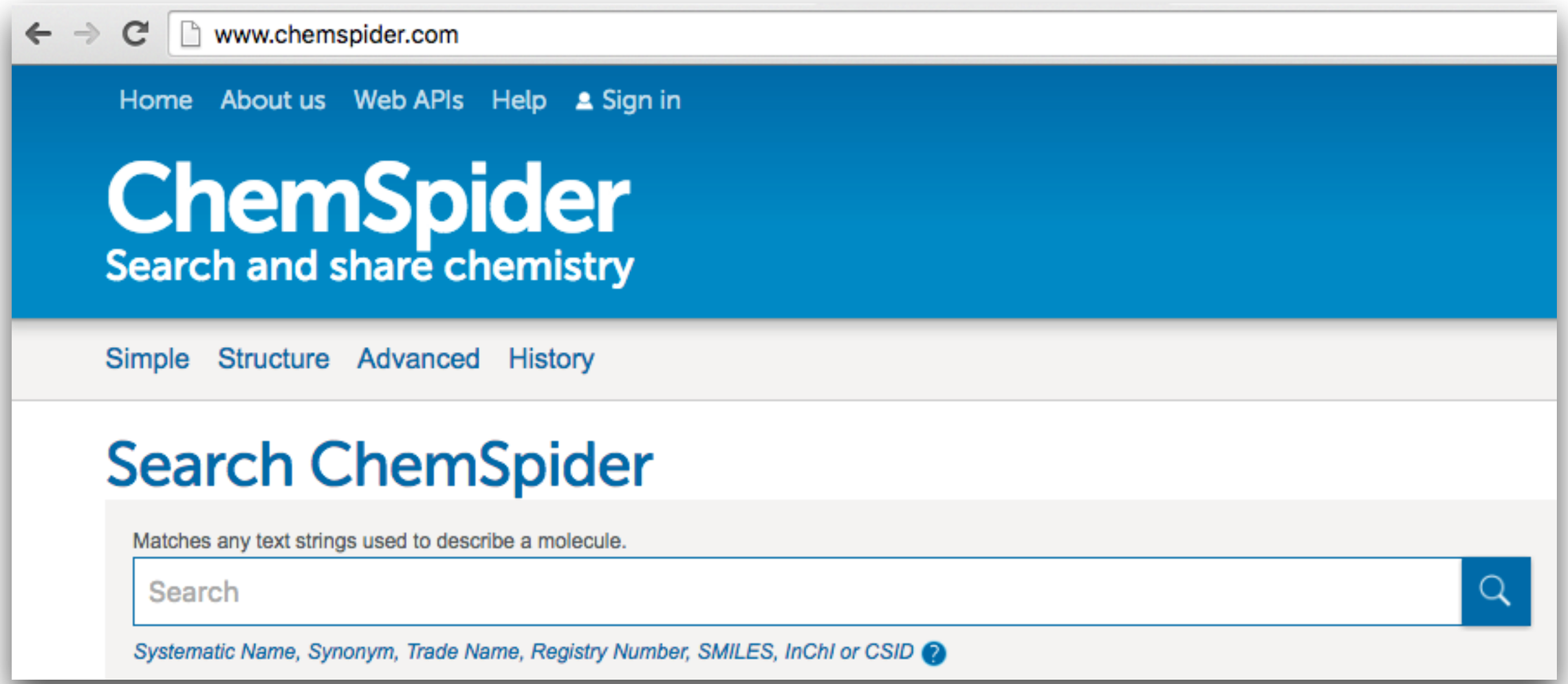
- Download compound XML files



- Total number of XML files: 4,849
- Number of compounds in each XML file: 25,000

ChemSpider

- Website



ChemSpider

- Information

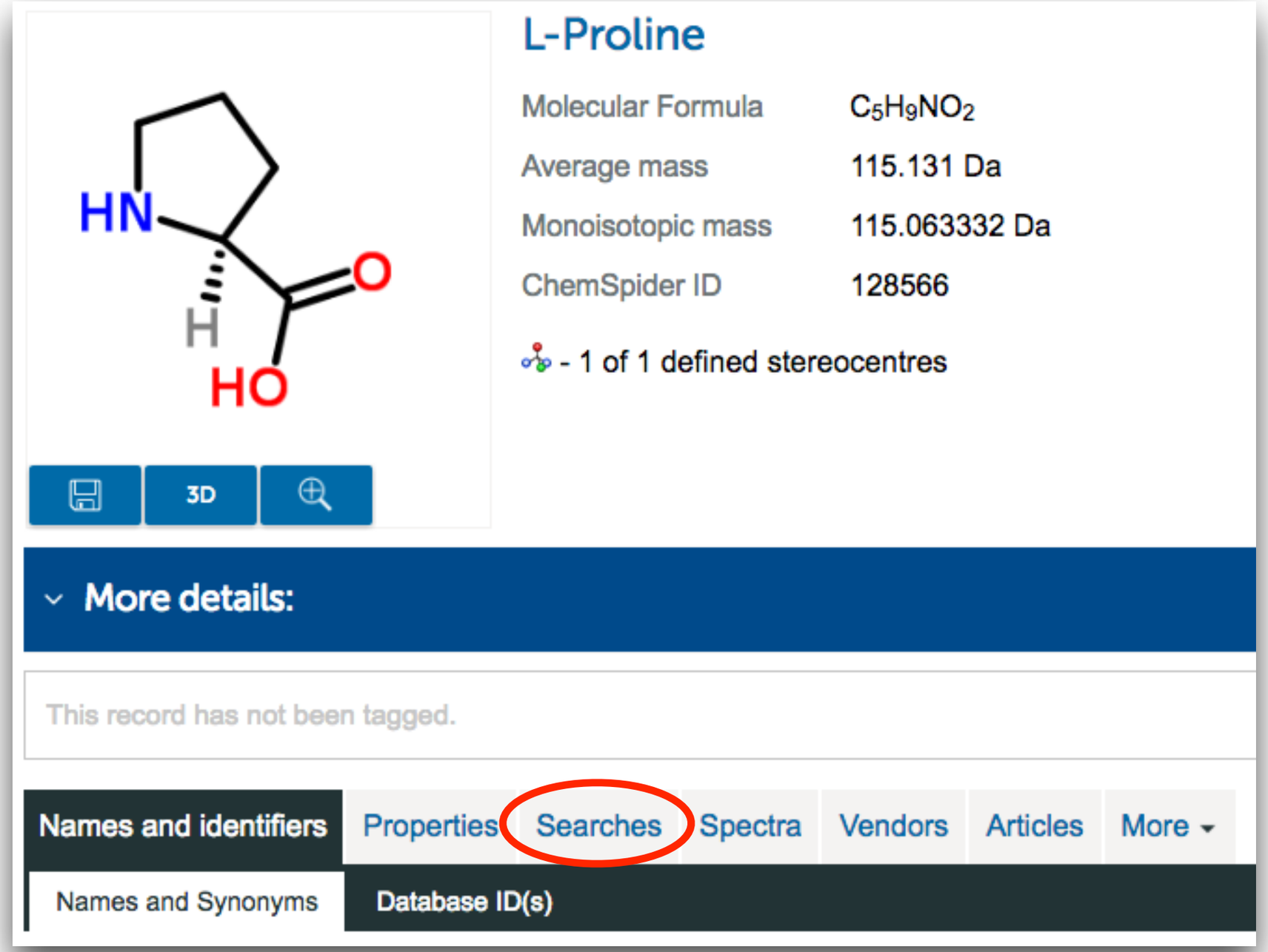
What is ChemSpider?	Search by chemical names	Search by chemical structure	Find important data
ChemSpider is a free chemical structure database providing fast text and structure search access to over 55 million structures from hundreds of data sources.	• Systematic names • Synonyms • Trade names • Database identifiers	• Create structure-based queries • Draw structures in the web page • Use structure files from your computer	• Literature references • Physical properties • Interactive spectra • Chemical suppliers

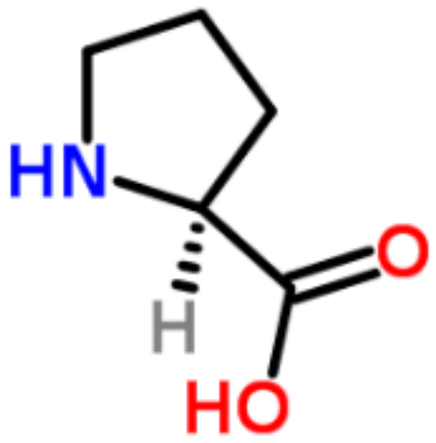
57 Million
chemical structures

517
Data sources

ChemSpider

- One compound





3D

L-Proline

Molecular Formula	$C_5H_9NO_2$
Average mass	115.131 Da
Monoisotopic mass	115.063332 Da
ChemSpider ID	128566

- 1 of 1 defined stereocentres

More details:

This record has not been tagged.

Names and identifiers | Properties | **Searches** | Spectra | Vendors | Articles | More

Search ChemSpider:

- Compounds with the same molecular formula
- Compounds with the same skeleton
- Use this molecule in a structure search

Search Google:

- Search Google Scholar (by synonym)
- Search Google for exact structure
- Search Google for structures with same skeleton

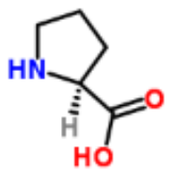
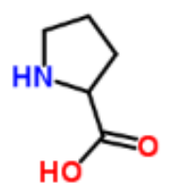
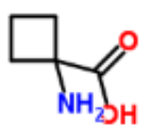
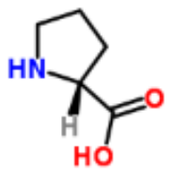
ChemSpider

- Compounds with the same molecular formula

Found 501 results

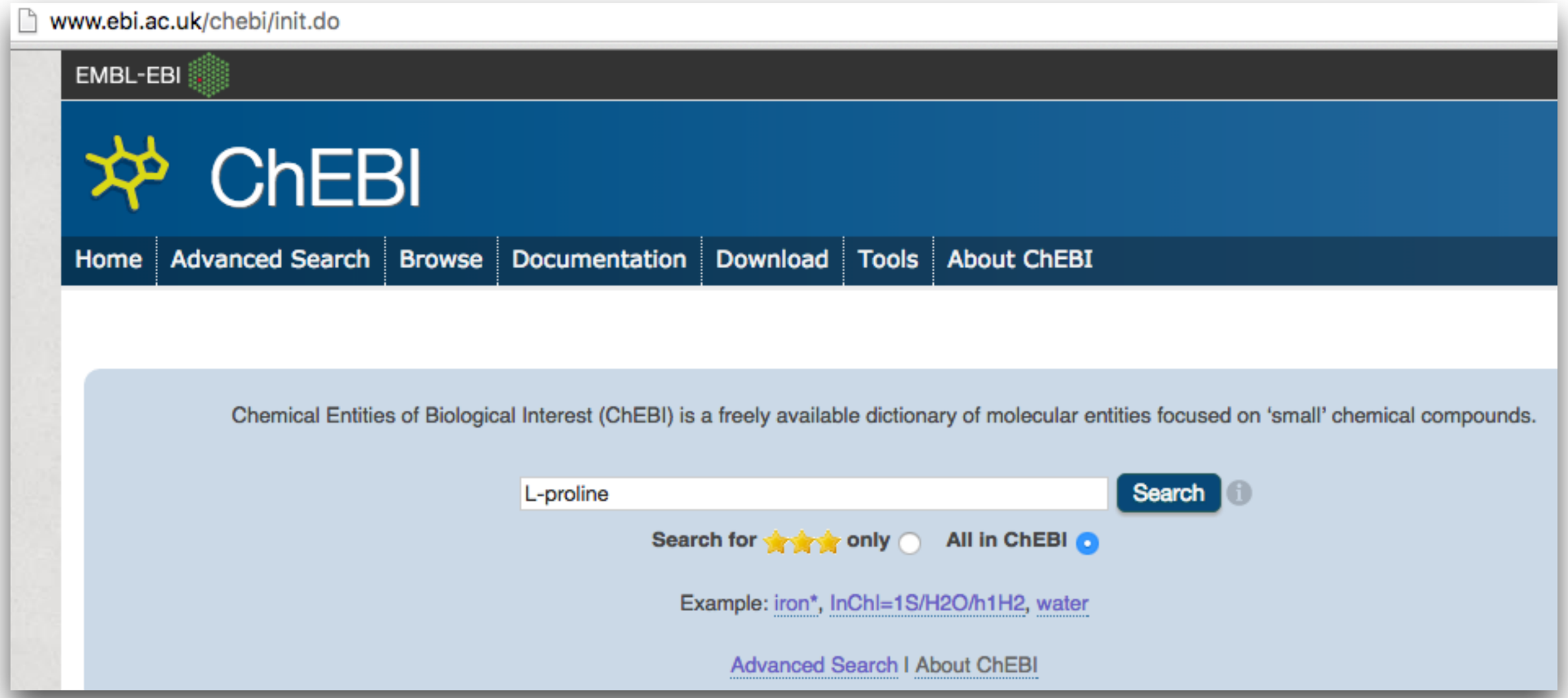
Search term: **MF = 'C_{5}H_{9}NO_{2}'**

1 2 3 4 5

ID	Structure	Molecular Formula	Molecular Weight	# of Data Sources	# of References	# of PubMed	# of RSC
128566 - 1/1 defined		$C_5H_9NO_2$	115.1305	136	11954	28480	3548
594 - 0/1 defined		$C_5H_9NO_2$	115.1305	131	737	28034	3319
80908		$C_5H_9NO_2$	115.1305	110	185	8	5
8640 - 1/1 defined		$C_5H_9NO_2$	115.1305	103	293	701	268

ChEBI

- Chemical Entities of Biological Interest



ChEBI

- About

ChEBI > About ChEBI

1. Introduction

Chemical **E**ntities of **B**iological **I**nterest (**ChEBI**) is a freely available dictionary of molecular entities focused on 'small' chemical compounds. The term 'molecular entity' refers to any constitutionally or isotopically distinct atom, molecule, ion, ion pair, radical, radical ion, complex, conformer, etc., identifiable as a separately distinguishable entity. The molecular entities in question are either products of nature or synthetic products used to intervene in the processes of living organisms.

ChEBI incorporates an ontological classification, whereby the relationships between molecular entities or classes of entities and their parents and/or children are specified.

ChEBI uses nomenclature, symbolism and terminology endorsed by the following international scientific bodies:

- International Union of Pure and Applied Chemistry (IUPAC)
- Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB)

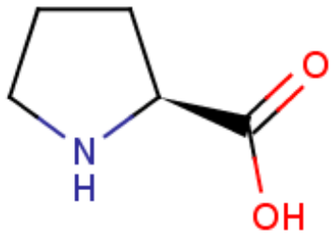
Molecules directly encoded by the genome (e.g. nucleic acids, proteins and peptides derived from proteins by cleavage) are *not* as a rule included in ChEBI.

ChEBI

- One compound

CHEBI:17203 - L-proline

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models



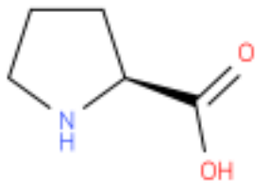
ChEBI Name	**L-proline**
ChEBI ID	**CHEBI:17203**
ChEBI ASCII Name	L-proline
Definition	Pyrrolidine in which the *pro-S* hydrogen at position 2 is substituted by a carboxylic acid group. L-Proline is the only one of the twenty DNA-encoded amino acids which has a secondary amino group α to the carboxyl group. It is an essential component of collagen and is important for proper functioning of joints and tendons. It also helps maintain and strengthen heart muscles.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:45159, CHEBI:45100, CHEBI:45040, CHEBI:42067, CHEBI:184637, CHEBI:6286, CHEBI:13154, CHEBI:21373
Supplier Information	eMolecules:524642, ZINC00895360

Download Molfile

- Find compounds which contain this structure
- Find compounds which resemble this structure
- Take structure to the Advanced Search

Search Results for All in ChEBI

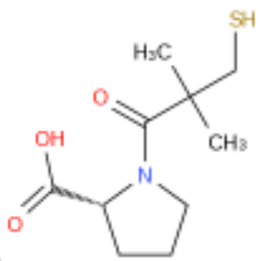
substructure



Edit Search

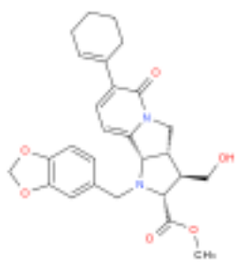
314 entries found, displaying 1 to 15.

(2R)-1-(3-mercapto-2,2-dimethyl-1-oxopropyl)-2-pyrrolidinecarboxylic acid



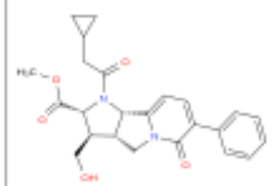
CHEBI:95254
Stars: ★★☆

(2R,3R,3aS,9bS)-1-(1,3-benzodioxol-5-ylmethyl)-7-(1-cyclohexenyl)-3-(hydroxymethyl)-6-oxo-3,3a,4,9b-tetrahydro-2H-pyrrolo[2,3-a]indolizine-2-carboxylic acid methyl ester

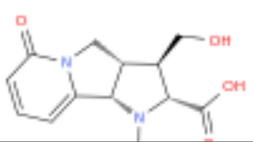


CHEBI:98944
Stars: ★★☆

(2R,3R,3aS,9b
(hydroxymeth
tetrahydro-2H
acid methyl es



(2R,3R,3aS,9bS)-1-(cyclopentylcarbamoyl)-3-(hydroxymethyl)-6-oxo-3,3a,4,9b-tetrahydro-2H-pyrrolo[2,3-a]indolizine-2-carboxylic acid



CHEBI:98447
Stars: ★★☆

(2R,3R,3aS,9bS)-1-(cyclopentylmethyl)-3-(hydroxymethyl)-7-(2-methoxyphenyl)-6-oxo-3,3a,4,9b-tetrahydro-2H-pyrrolo[2,3-a]indolizine-2-carboxylic acid methyl ester

CHEBI:131226
Stars: ★★☆

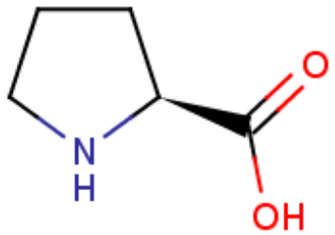
(2R,3R,3aS,9b
(hydroxymeth
tetrahydro-2H
acid

ChEBI

- One compound

CHEBI:17203 - L-proline

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models



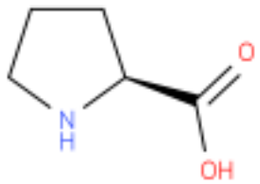
ChEBI Name	**L-proline**
ChEBI ID	**CHEBI:17203**
ChEBI ASCII Name	L-proline
Definition	Pyrrolidine in which the *pro-S* hydrogen at position 2 is substituted by a carboxylic acid group. L-Proline is the only one of the twenty DNA-encoded amino acids which has a secondary amino group α to the carboxyl group. It is an essential component of collagen and is important for proper functioning of joints and tendons. It also helps maintain and strengthen heart muscles.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.
Secondary ChEBI IDs	CHEBI:45159, CHEBI:45100, CHEBI:45040, CHEBI:42067, CHEBI:184637, CHEBI:6286, CHEBI:13154, CHEBI:21373
Supplier Information	eMolecules:524642, ZINC00895360

Download Molfile

- Find compounds which contain this structure
- Find compounds which resemble this structure
- Take structure to the Advanced Search

Search Results for All in ChEBI

similar structures

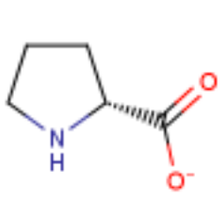


Edit Search

125 entries found, displaying 1 to 15.

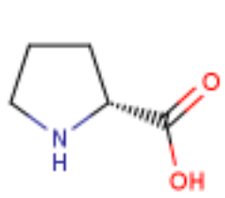
D-prolinate

CHEBI:32867
Stars: ★★★



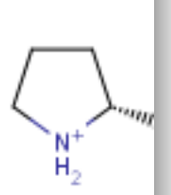
D-proline

CHEBI:16313
Stars: ★★★



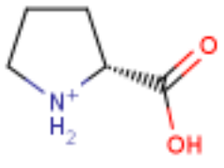
D-proline

Stars: ★★★



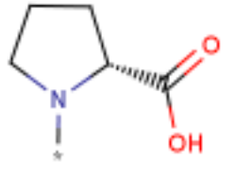
D-prolinium

CHEBI:32868
Stars: ★★★

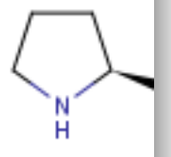


D-prolino group

CHEBI:32870
Stars: ★★★

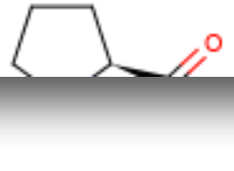


L-prolinat



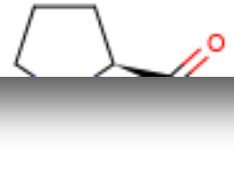
L-proline

CHEBI:17203
Stars: ★★★



L-proline zwitterion

CHEBI:60039
Stars: ★★★



L-prolini

Outline

- Comprehensive metabolomics databases
- Compound databases
- Spectral databases
- Metabolic pathway databases
- Drug databases
- Disease & physiology databases
- Raw data databases

Spectral databases

- NIST 14
- METLIN
- MassBank
- MoNA
- Gold Metabolome Database
- Feign GC-MS database
- HMDB
- BMRB
- Madison Metabolomics Consortium Database
- BML-NMR
- mzCloud

Spectral databases

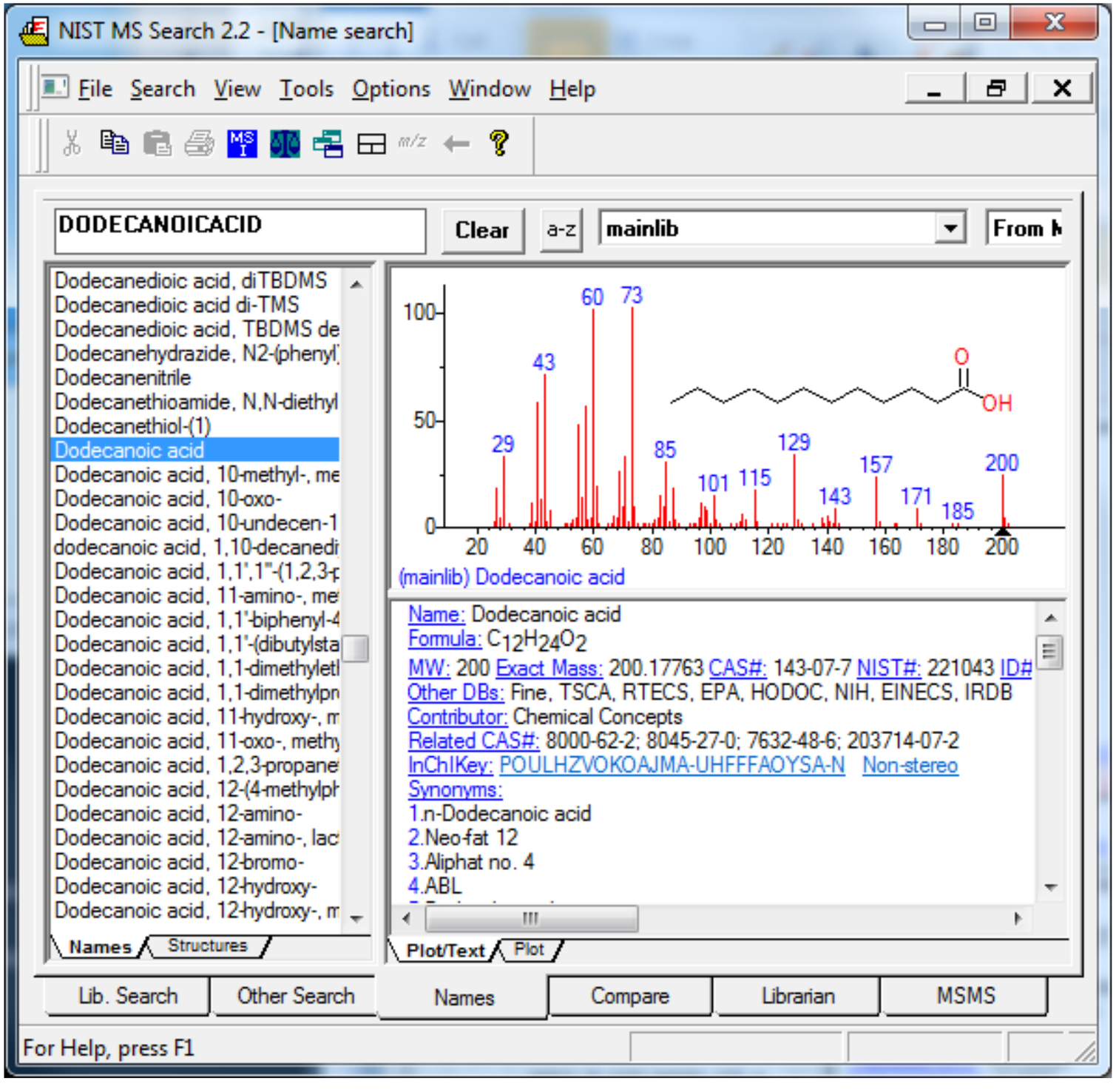
- NIST 14
- METLIN
- MassBank
- MoNA
- Gold Metabolome Database
- Feign GC-MS database
- HMDB
- BMRB
- Madison Metabolomics Consortium Database
- BML-NMR
- mzCloud

NIST 14

- Electron ionization mass spectral library
  - 276,259 spectra of 242,477 unique compounds

- MS/MS library: 234,284 spectra
  - 51,216 ion trap spectra for 42,126 different ions of 8,171 compounds
  - 183,068 collision cell spectra (QTOF and tandem quad) spectra for 14,835 different ions of 7,692 compounds

NIST 14 EI library

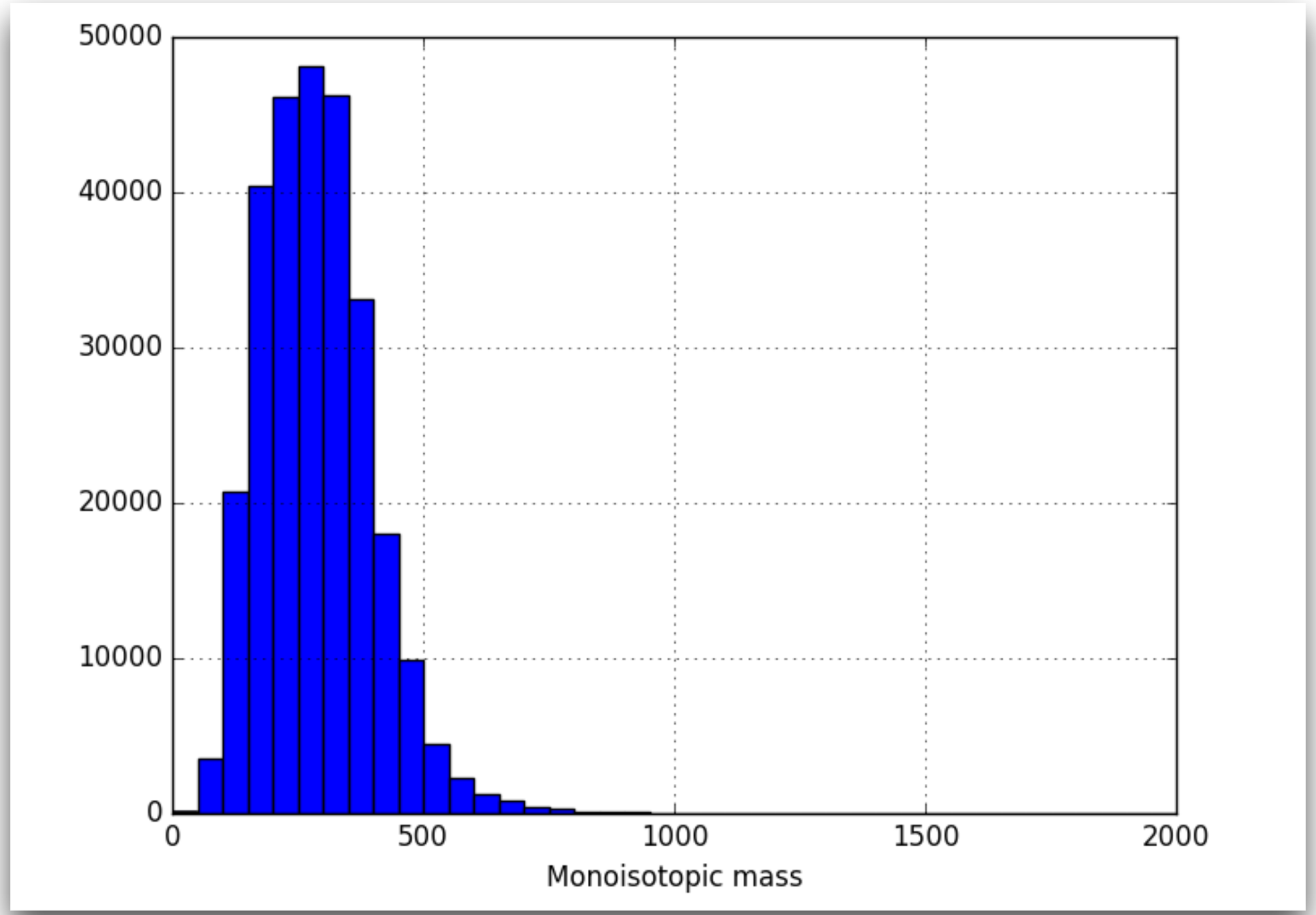


NIST 14 EI library

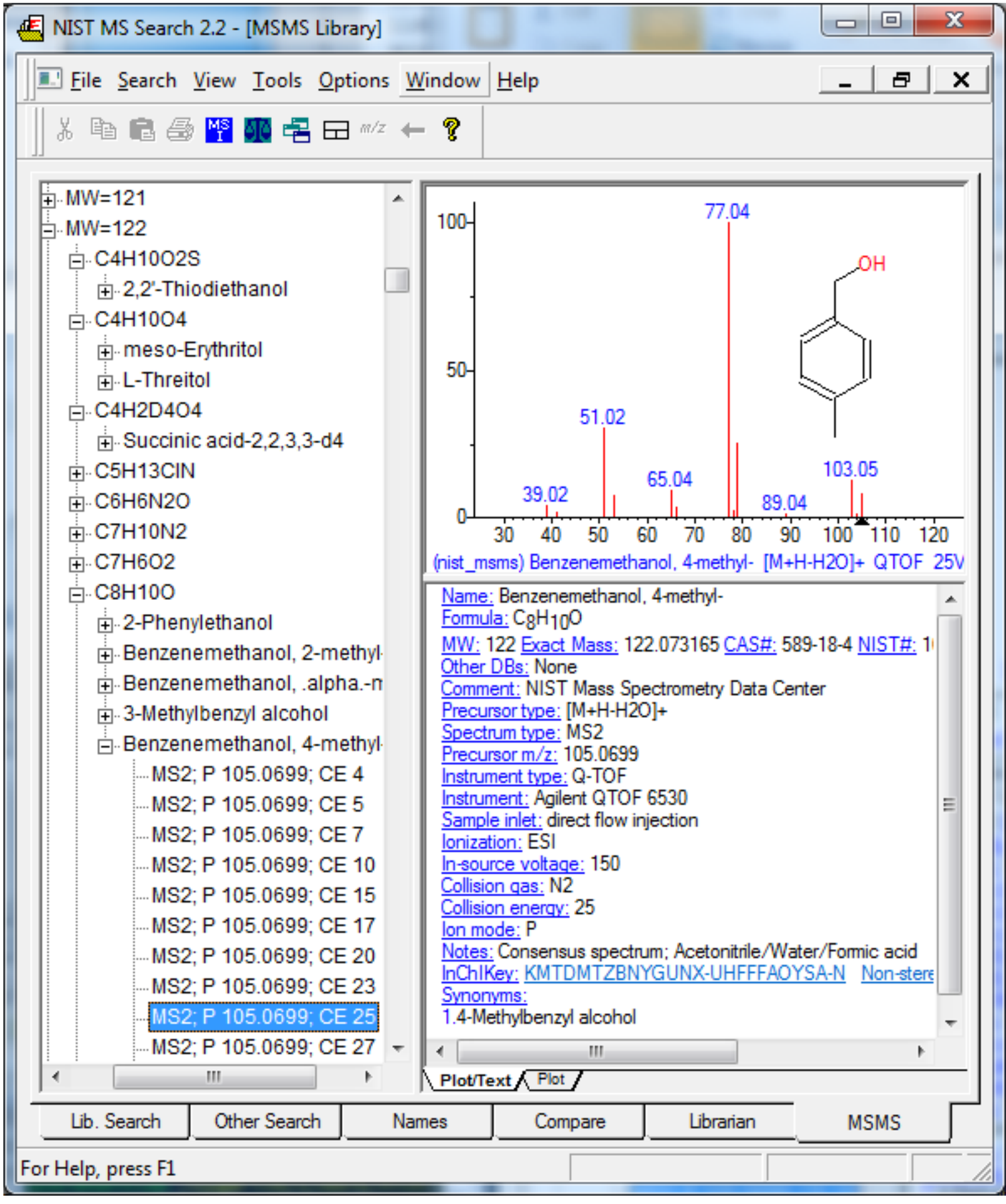
- Focuses on
  - Drugs, metabolites, and poisons
  - Pesticides and fungicides
  - Organics present in soil, water, and air
  - Amino acids, di- and tai-peptides
  - Common sample contaminants
  - Common analytical derivatives of the above

NIST 14 EI library

- Breakdown by mass



NIST 14
MS/MS library

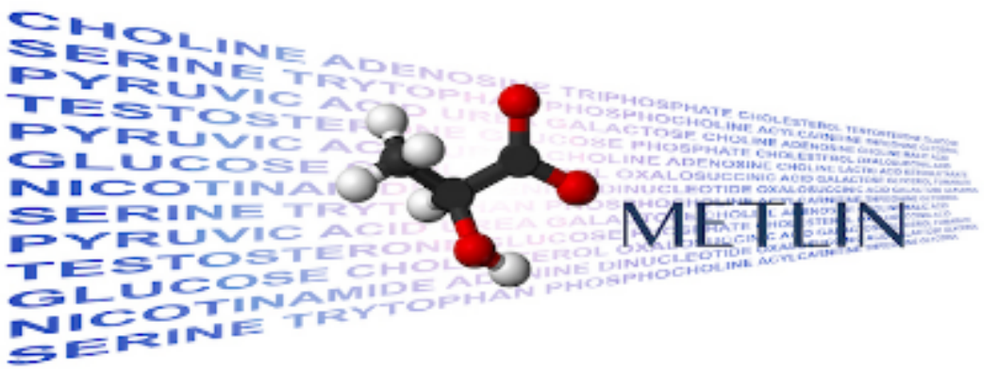


METLIN

https://metlin.scripps.edu/index.php

Scripps Center for Metabolomics

MS HOME | METLIN | XCMS Online | XCMS Institute | XCMS Public | Contact



Statistics

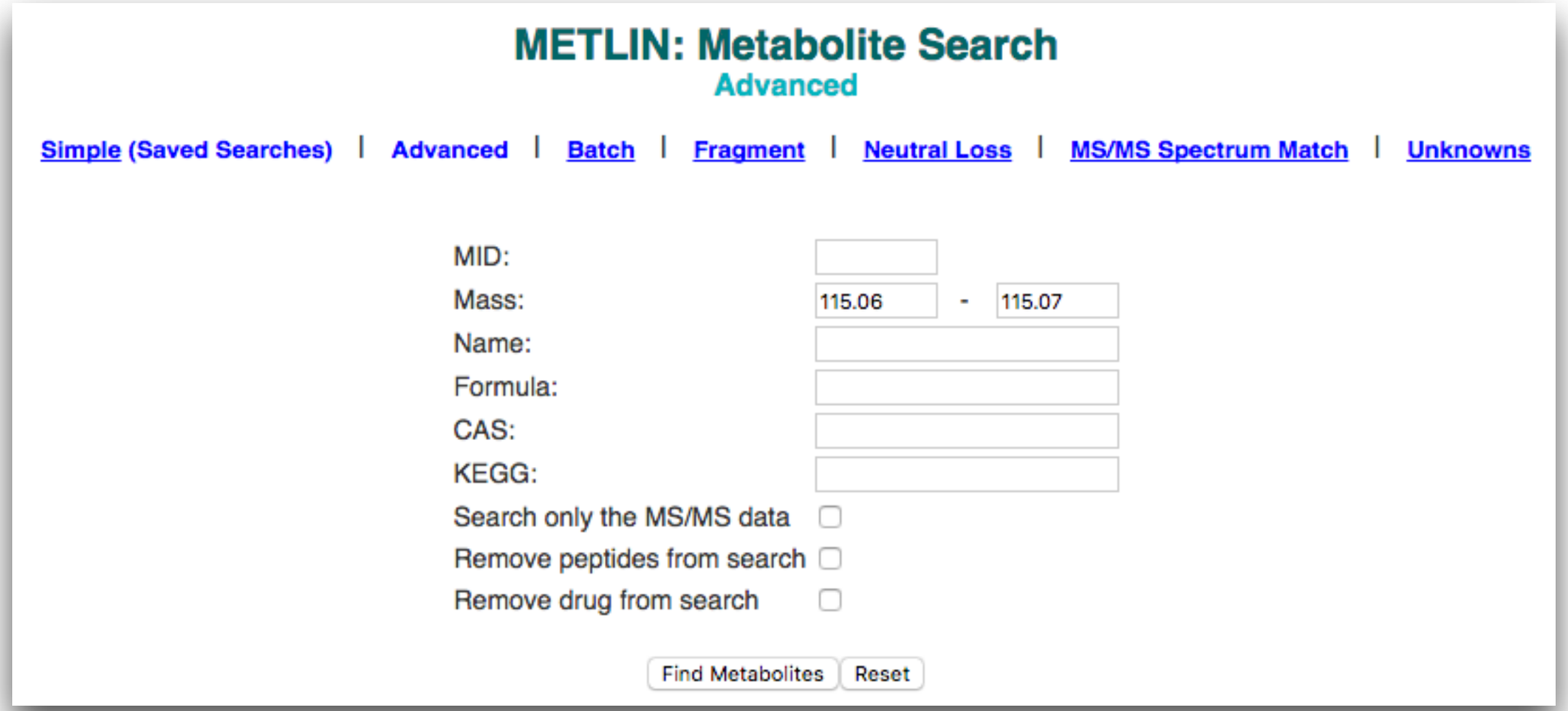
- # Metabolites: 242,032
- # High Resolution MS/MS Spectra: 72,268
- # Metabolites w/ High Resolution MS/MS: 14,034

example | details...

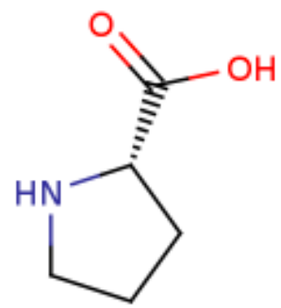
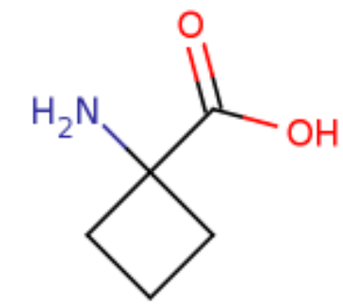
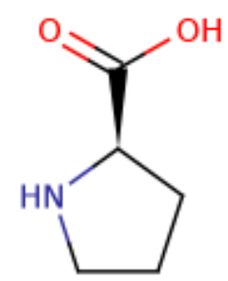
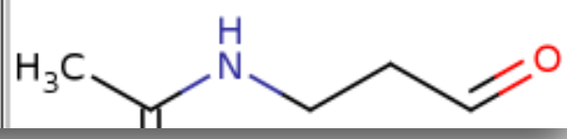
Functionality

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

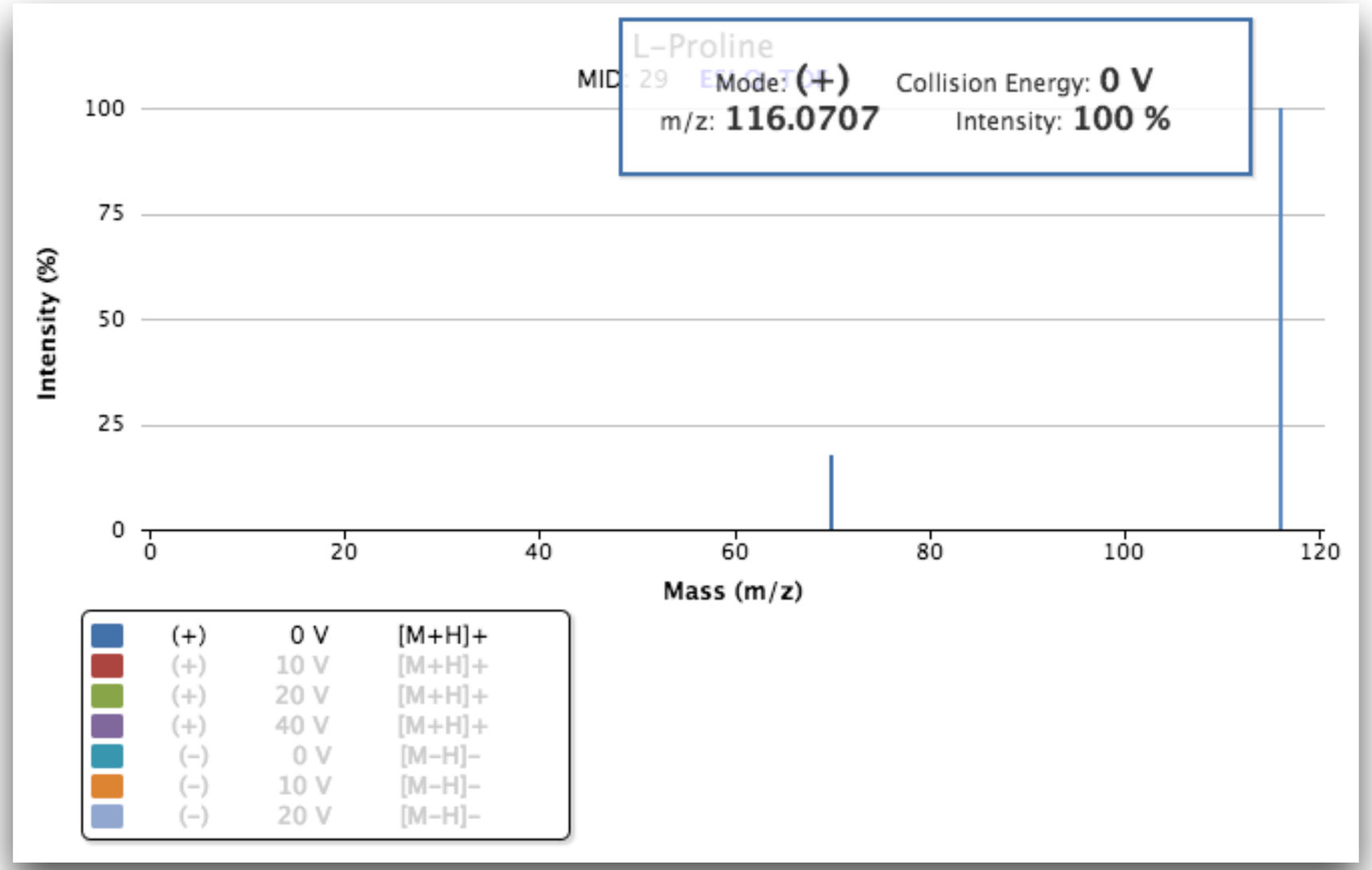
METLIN


METLIN: Metabolite Search
Advanced
Simple (Saved Searches) | Advanced | Batch | Fragment | Neutral Loss | MS/MS Spectrum Match | Unknowns
MID:
Mass:
115.06
-
115.07
Name:
Formula:
CAS:
KEGG:
Search only the MS/MS data
Remove peptides from search
Remove drug from search
Find Metabolites
Reset

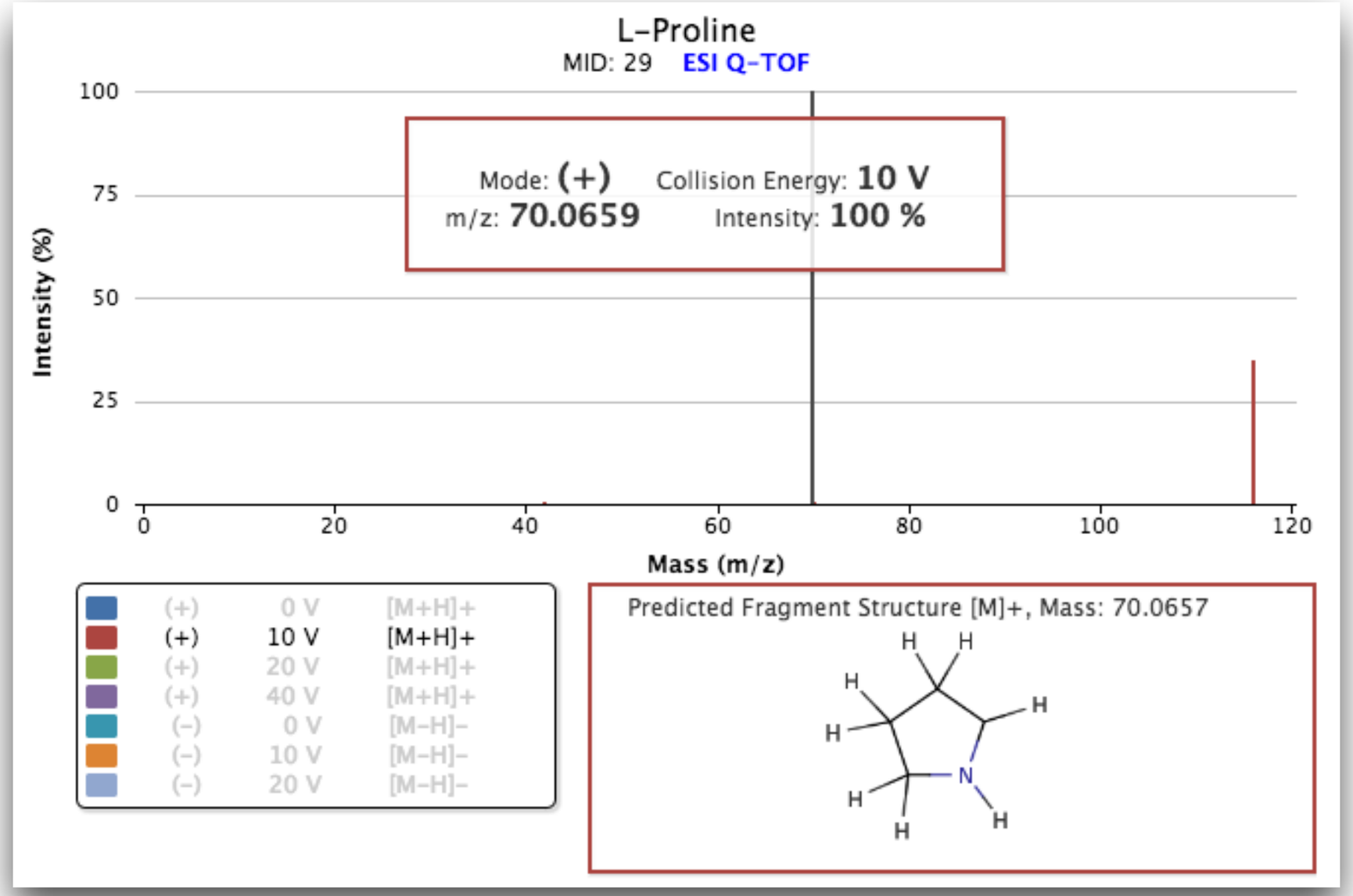
METLIN

MID	Mass	Name	Formula	CAS	KEGG	MS/MS	Structure
29	115.0633	L-Proline	$C_5H_9NO_2$	147-85-3	C00148	View	
44305	115.0633	1-AMINOCYCLOBUTANE CARBOXYLIC ACID	$C_5H_9NO_2$	22264-50-2		View	
58150	115.0633	D-Proline	$C_5H_9NO_2$	344-25-2	C00763	View	
72070	115.0633	3-Acetamidopropanal	$C_5H_9NO_2$		C18170	NO	

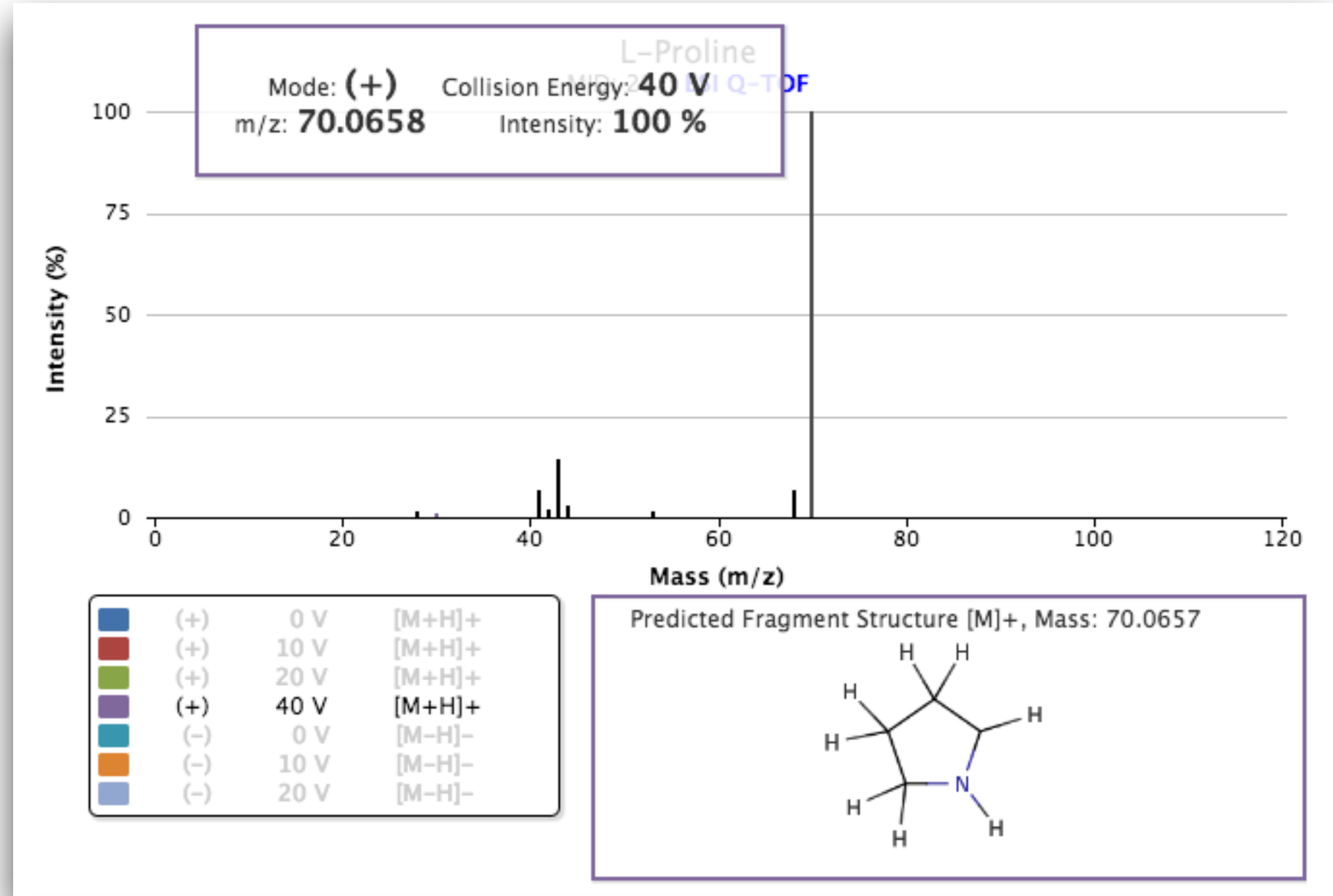
METLIN


L-Proline
Mode: (+)
Collision Energy: 0 V
m/z: 116.0707
Intensity: 100 %
Intensity (%)
100
75
50
25
0
0
20
40
60
80
100
120
Mass (m/z)
(+) 0 V [M+H]+
(+) 10 V [M+H]+
(+) 20 V [M+H]+
(+) 40 V [M+H]+
(-) 0 V [M-H]-
(-) 10 V [M-H]-
(-) 20 V [M-H]-

METLIN

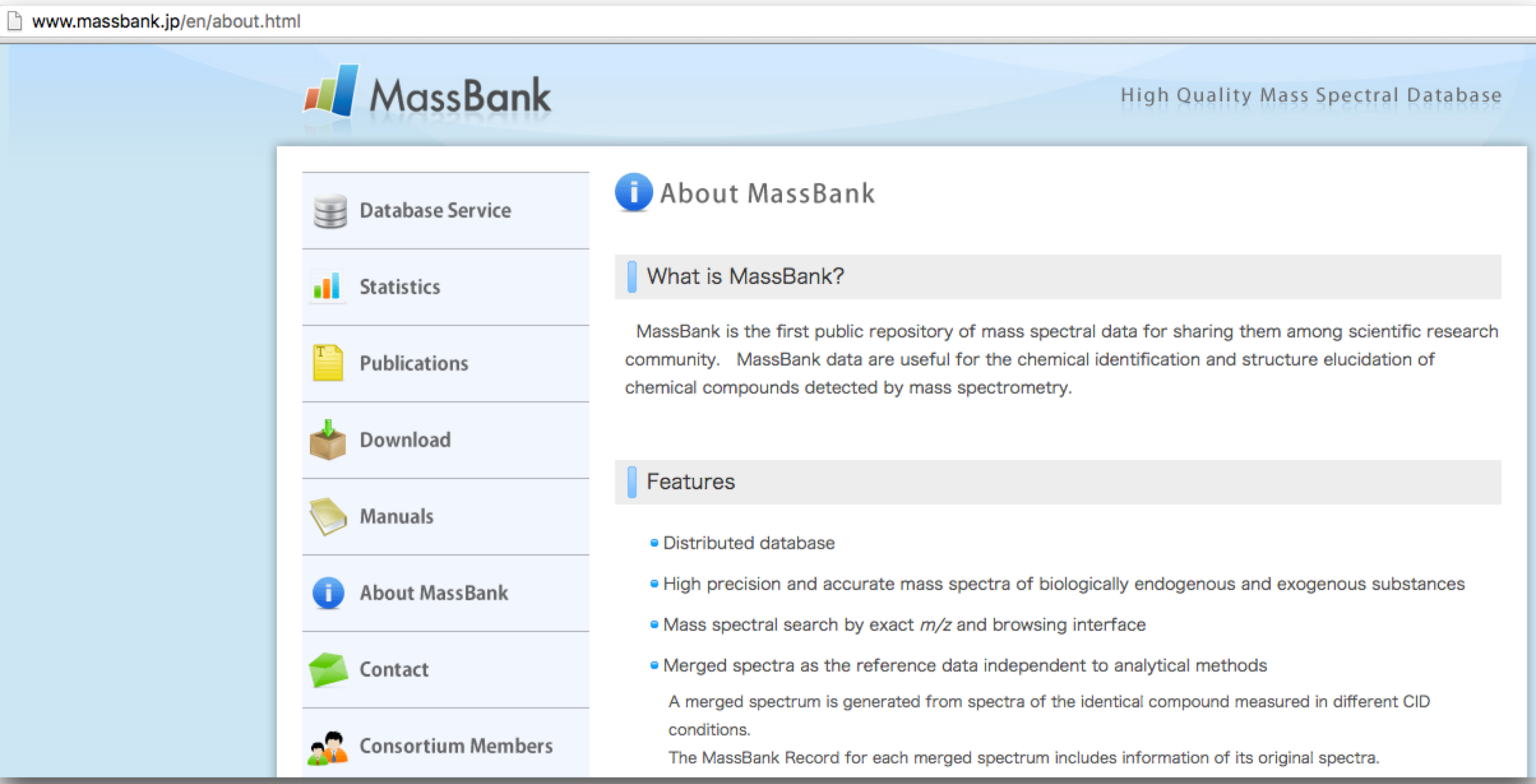

L-Proline
MID: 29 ESI Q-TOF
Mode: (+) Collision Energy: 10 V
m/z: 70.0659 Intensity: 100 %
Intensity (%)
100
75
50
25
0
0
20
40
60
80
100
120
Mass (m/z)
(+) 0 V [M+H]+
(+) 10 V [M+H]+
(+) 20 V [M+H]+
(+) 40 V [M+H]+
(-) 0 V [M-H]-
(-) 10 V [M-H]-
(-) 20 V [M-H]-
Predicted Fragment Structure [M]+, Mass: 70.0657
H
H
H
H
H
N
H
H
H

METLIN


L-Proline
Mode: (+)
Collision Energy: 40 V
m/z: 70.0658
Intensity: 100 %
Intensity (%)
100
75
50
25
0
0
20
40
60
80
100
120
Mass (m/z)
(+) 0 V [M+H]+
(+) 10 V [M+H]+
(+) 20 V [M+H]+
(+) 40 V [M+H]+
(-) 0 V [M-H]-
(-) 10 V [M-H]-
(-) 20 V [M-H]-
Predicted Fragment Structure [M]+, Mass: 70.0657
H
H
H
H
H
H
N
H

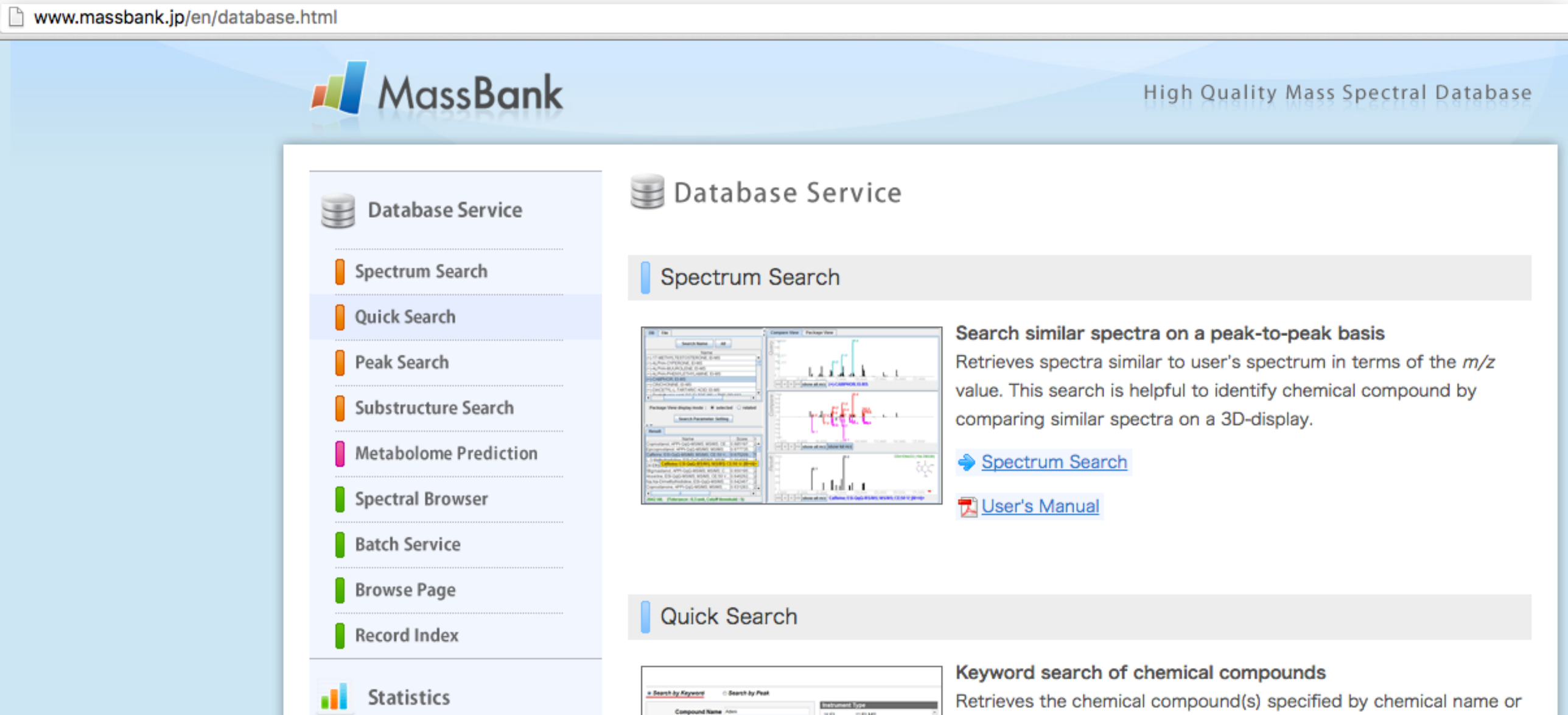
MassBank

- About

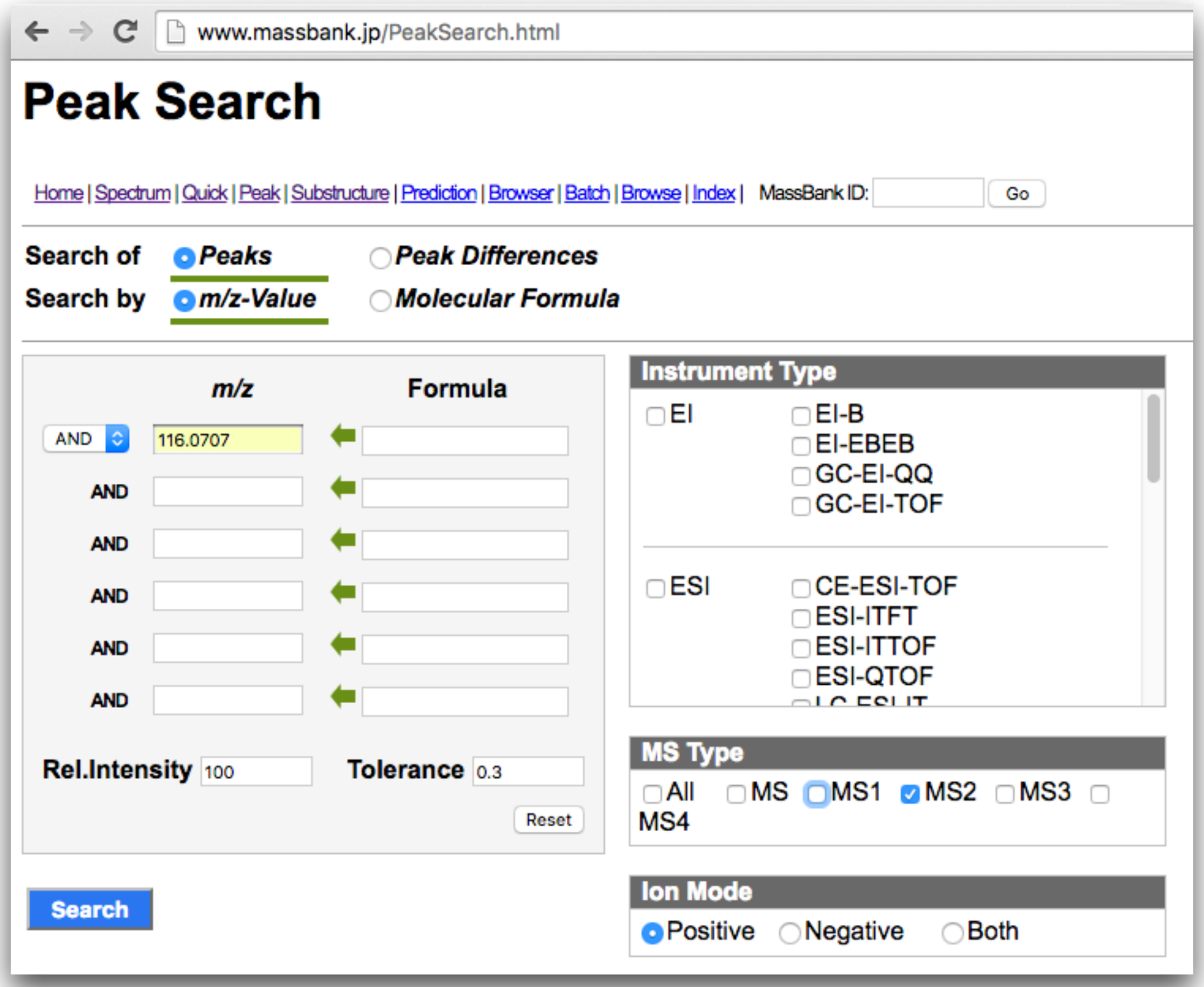


MassBank

- Database services



MassBank


www.massbank.jp/PeakSearch.html
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
m/z
Formula
AND 116.0707
AND
AND
AND
AND
AND
Rel.Intensity 100
Tolerance 0.3
Reset
Search
Instrument Type
EI
EI-B
EI-EBEB
GC-EI-QQ
GC-EI-TOF
ESI
CE-ESI-TOF
ESI-ITFT
ESI-ITTOF
ESI-QTOF
MS Type
All MS MS1 MS2 MS3 MS4
Ion Mode
Positive Negative Both

MassBank

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

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

Search Parameters :
m/z: **116.0707** Rel.Int: **100** Tol.(unit): **0.3**

Instrument Type: **LC-ESI-QTOF**
MS Type: **MS2**
Ion Mode: **Positive**

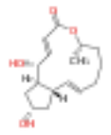
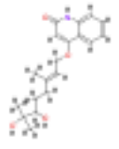
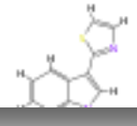
Edit / Resubmit Query

Results : 113 Hit. **(41 - 81 Displayed)**

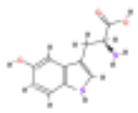
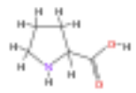
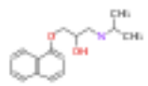
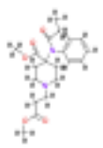
Open All Tree | Multiple Display | Spectrum Search

First Prev 1 **2** 3 Next Last (Total **3** Page)

▼ Results End

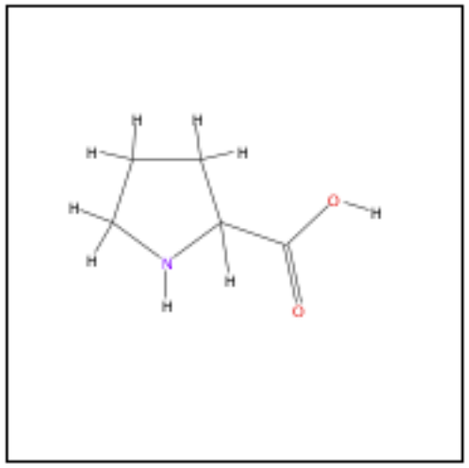
Name	Formula / Structure	ExactMass	ID
Brefeldin-A 1 spectrum	**C16H24O4** 	**280.16746**	
Bucharaine 1 spectrum	**C19H25NO4** 	**331.17836**	
Camalexin 2 spectra	**C11H8N2S** 	**200.04082**	

MassBank

⊞ **Oxitriptan 1** 1 spectrum	**C11H12N2O3** 	**220.08479**	
⊟ **Proline** 2 spectra LC-ESI-QTOF; MS2; CE:10 eV; [M+H]+ LC-ESI-QTOF; MS2; CE:15 eV; [M+H]+	**C5H9NO2** 	**115.06333**	PB000449 PB000450
⊞ **Propranolol** 1 spectrum	**C16H21NO2** 	**259.15720**	
⊞ **Remifentanil** 4 spectra	**C20H28N2O5** 	**376.19982**	
⊞ **S-Lactoylglutathione** 1 spectrum	**C13H21N3O8S** 	**379.10494**	

Proline; LC-ESI-QTOF; MS2; CE:15 eV; [M+H]+

Mass Spectrum Chemical Structure



```
ACCESSION: PB000450
RECORD_TITLE: Proline; LC-ESI-QTOF; MS2; CE:15 eV; [M+H]+
DATE: 2016.01.19 (Created 2008.01.02, modified 2013.06.04)
AUTHORS: Boettcher C, Institute of Plant Biochemistry, Halle, Germany
LICENSE: CC BY-SA
COMMENT: IPB_RECORD: 261
COMMENT: CONFIDENCE: confident structure

CH$NAME: Proline
CH$NAME: pyrrolidine-2-carboxylic acid
CH$COMPOUND_CLASS: Natural Product; amino acid
CH$FORMULA: C5H9NO2
CH$EXACT_MASS: 115.06333
CH$SMILES: C1CC(NC1)C(=O)O
CH$IUPAC: InChI=1S/C5H9NO2/c7-5(8)4-2-1-3-6-4/h4,6H,1-3H2,(H,7,8)
CH$LINK: INCHIKEY ONIBWKKTOPOVIA-UHFFFAOYSA-N
CH$LINK: KEGG C00148
CH$LINK: PUBCHEM CID:614

AC$INSTRUMENT: API QSTAR Pulsar i
AC$INSTRUMENT_TYPE: LC-ESI-QTOF
AC$MASS_SPECTROMETRY: MS_TYPE MS2
AC$MASS_SPECTROMETRY: ION_MODE POSITIVE
AC$MASS_SPECTROMETRY: COLLISION_ENERGY 15 eV
AC$MASS_SPECTROMETRY: IONIZATION ESI

MS$FOCUSED_ION: PRECURSOR_TYPE [M+H]+

PK$SPLASH: splash10-01b9-8900000000-000cb5f6ae87b559450a
```

MassBank

mona.fiehnlab.ucdavis.edu/spectra/splash/splash10-01b9-8900000000-000cb5f6ae87b559450a

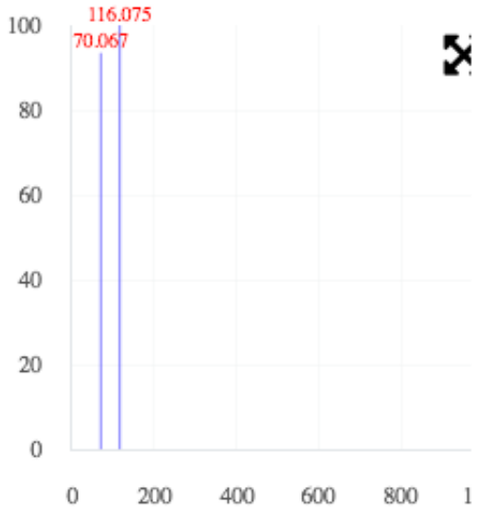
MoNA - MassBank of North America | Browse | Statistics | Resources | Download | Upload | Search...

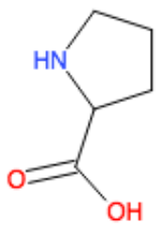
display generated query

1

Proline

Score: ★★★★★★★★★★☆

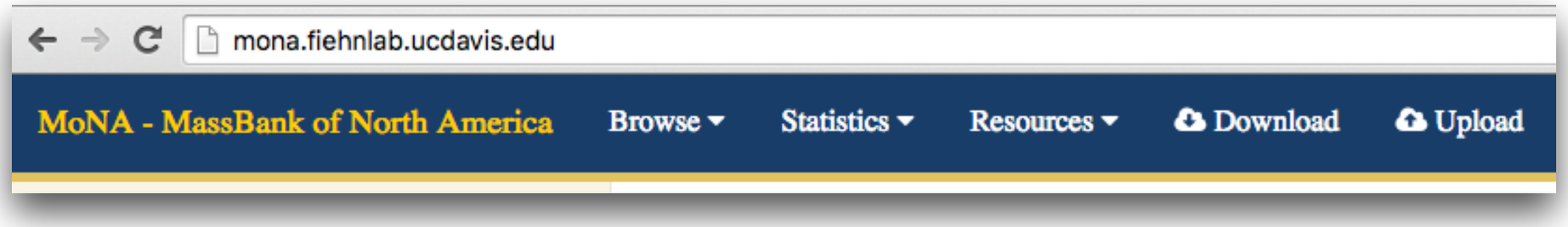




accession	PB000450
authors	Boettcher C, Institute of...
collision energy	15 eV
compound class	Natural Product; amino ac...
exact mass	115.0633
instrument	API QSTAR Pulsar i
instrument type	LC-ESI-QTOF
ion mode	positive
ionization	ESI
license	CC BY-SA
mass accuracy	8754.0881
mass error	-1007.2745
ms level	MS2
origin	PB000450.txt
precursor type	[M+H]+

Originally submitted to the MassBank Spectral Database as PB000450

MoNA



Welcome to MoNA

MassBank of America (MoNA), is an auto curating repository for storing, comparing and querying mass spectra of chemical compounds. It is metadata centric and it was designed to allow easy integration into other tools by utilize its REST based application programming interface. At the current time it contains over 200k predicted and 40k unique experimental mass spectra and their associated metadata. The predicted spectra were obtained by utilizing the lipid blast library and the experimental spectra were acquired from 30 different facilities all over the world, including several major research facilities in the United States and Japan. MoNA is utilizing the InChI Key as unique identifier for chemicals and is designed for easy scalability and expandability. This is realized by utilizing common applications like nginx, grails, AngularJS, postgresSQL and tomcat. MoNA is currently integrated in applications like MSDial, BinBase, MZMine and the statistics package R. This was accomplished by utilizing its REST based API, which is also utilized by its main AngularJS based web interface. We consider MoNA to be highly useful for crosslinking mass spectra in publications, identification of unknowns and integration in data acquisition software.

Metabolic pathway databases

- KEGG
- MetaCyc
- HumanCyc
- BioCyc
- Reactome
- WikiPathways

Drug databases

- DrugBank
- Therapeutic target databases
- PharmGKB
- STITCH
- SuperTarget

Disease & physiology databases

- OMIM
- METAGENE
- OMMBID

Raw data databases

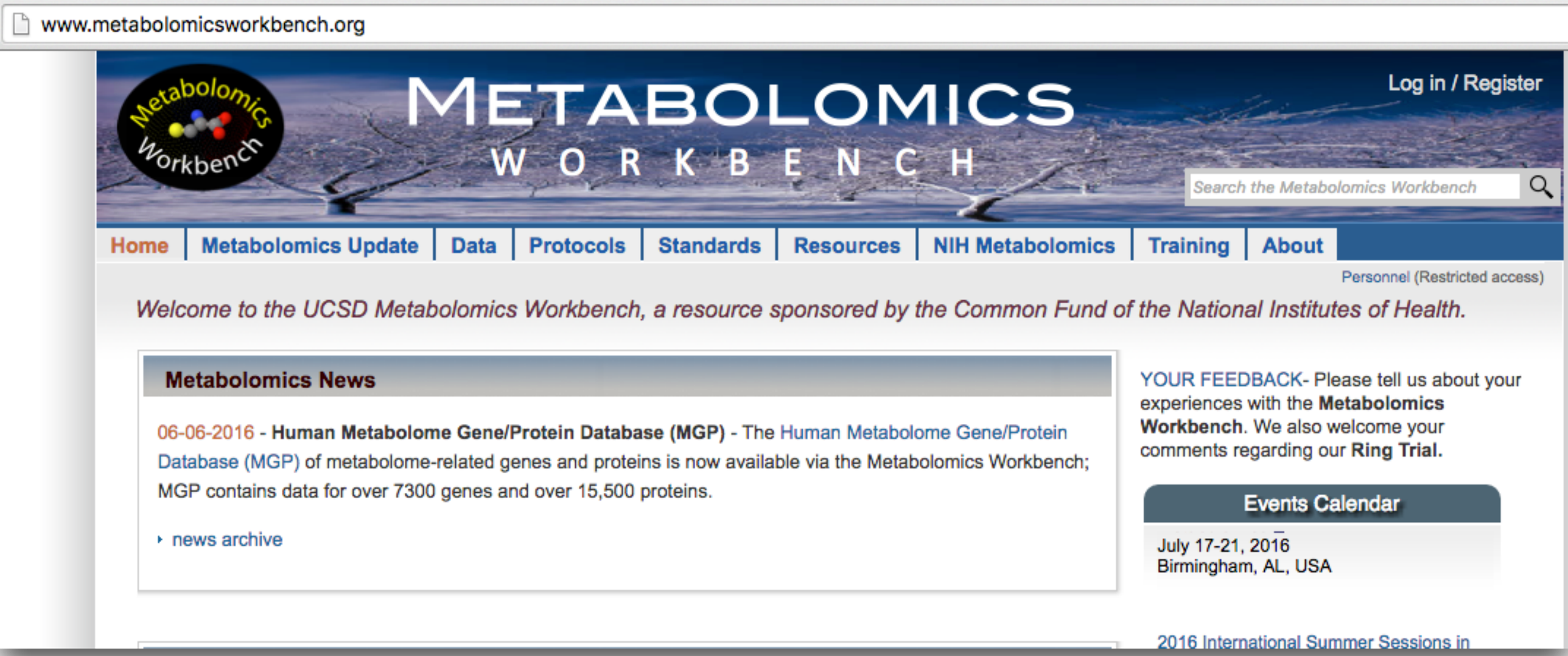
- Metabolomics Workbench
  - Funded by the NIH Common Fund Metabolomics Program
  - Serve as a national and international repository for metabolomics data and metadata
  - Provide access to raw data, metabolite standards, protocols .....

Raw data databases

- Metabolomics Workbench
- MetaboLights

Metabolomics orkbench

- Website



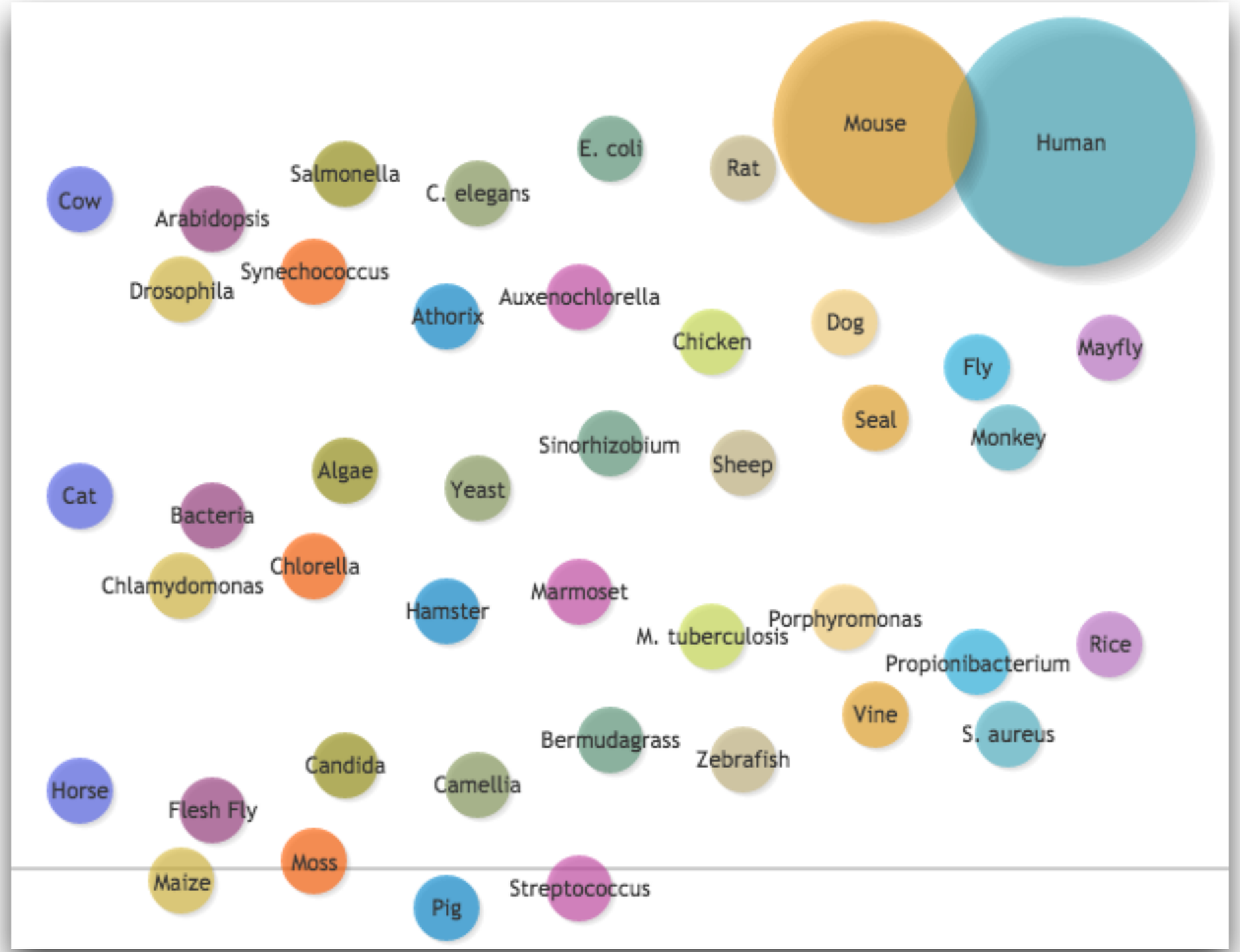
Raw data databases

- Metabolomics Workbench: summary of all studies

Study ID	Study Title	Species	Institute	Analysis	Release Date	Version	Samples	Download (* : Contains raw data)
ST000001	Fatb Induction Experiment (FatBIE)	Arabidopsis thaliana	University of California, Davis	MS	2013-02-14	1	24	Uploaded data (476K)
ST000002	Intestinal Samples II pre/post transplantation	Homo sapiens	University of California, Davis	MS	2013-02-22	1	12	Uploaded data (664K)
ST000003	Metabolomic analysis of mouse embryonic fibroblasts, embryonic stem cells, and induced pluripotent stem cells	Mus musculus	University of California, Davis	MS	2013-02-15	1	18	Uploaded data (5.3G)*
ST000004	Lipidomics studies on NIDDK / NIST human plasma samples	Homo sapiens	LIPID MAPS	MS	2013-03-17	1	8	Uploaded data (48K)
ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Mus musculus	LIPID MAPS	MS	2013-03-22	1	696	Uploaded data (56K)
ST000006	White Wine Study	Vitis vinifera	University of California, Davis	MS	2013-03-23	1	102	Uploaded data (532K)
ST000007	Rice Infection Study	Oryza sativa	University of California, Davis	MS	2013-03-24	1	60	Uploaded data (1.7M)

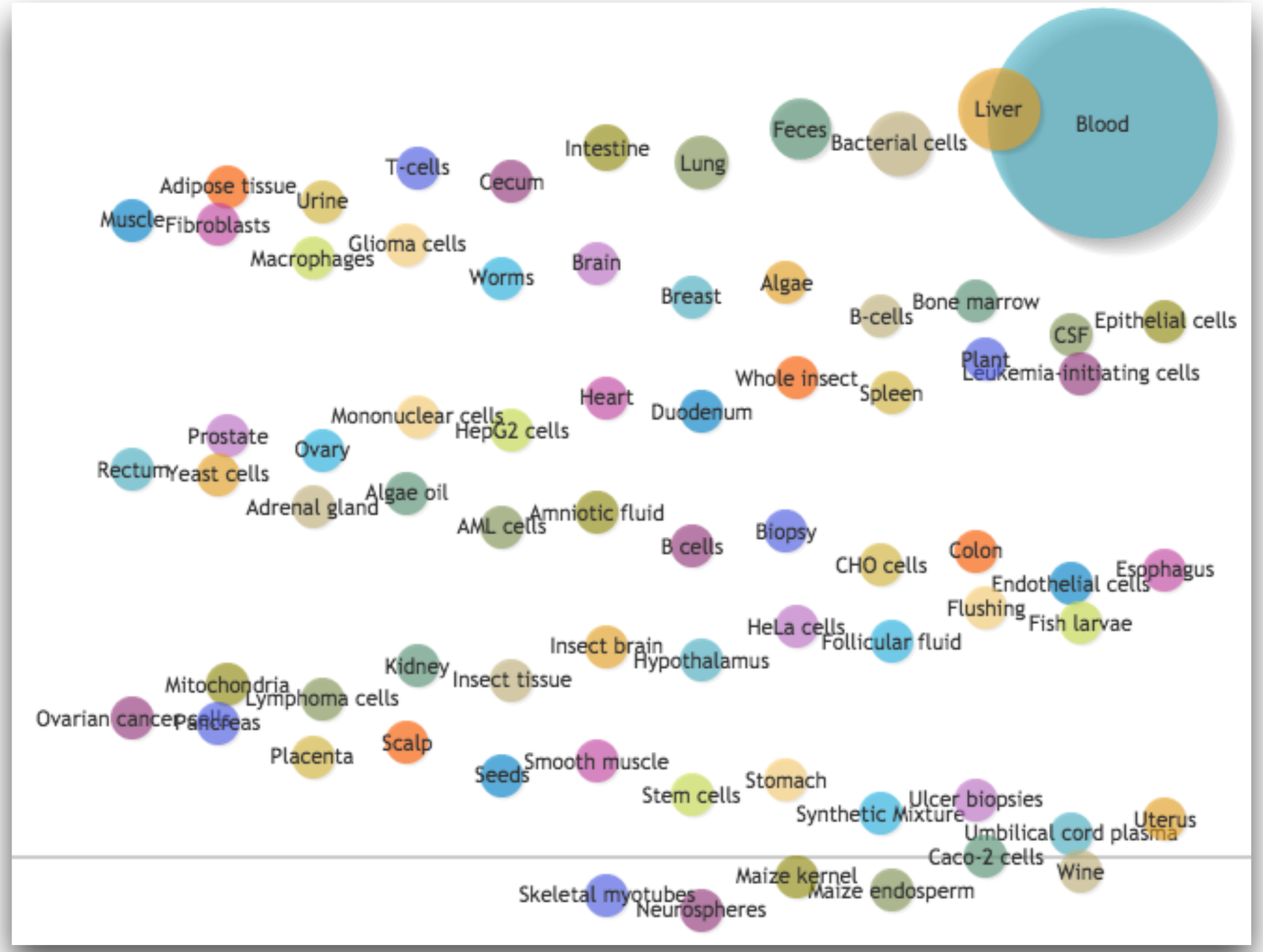
Raw data databases

- Metabolomics Workbench: species



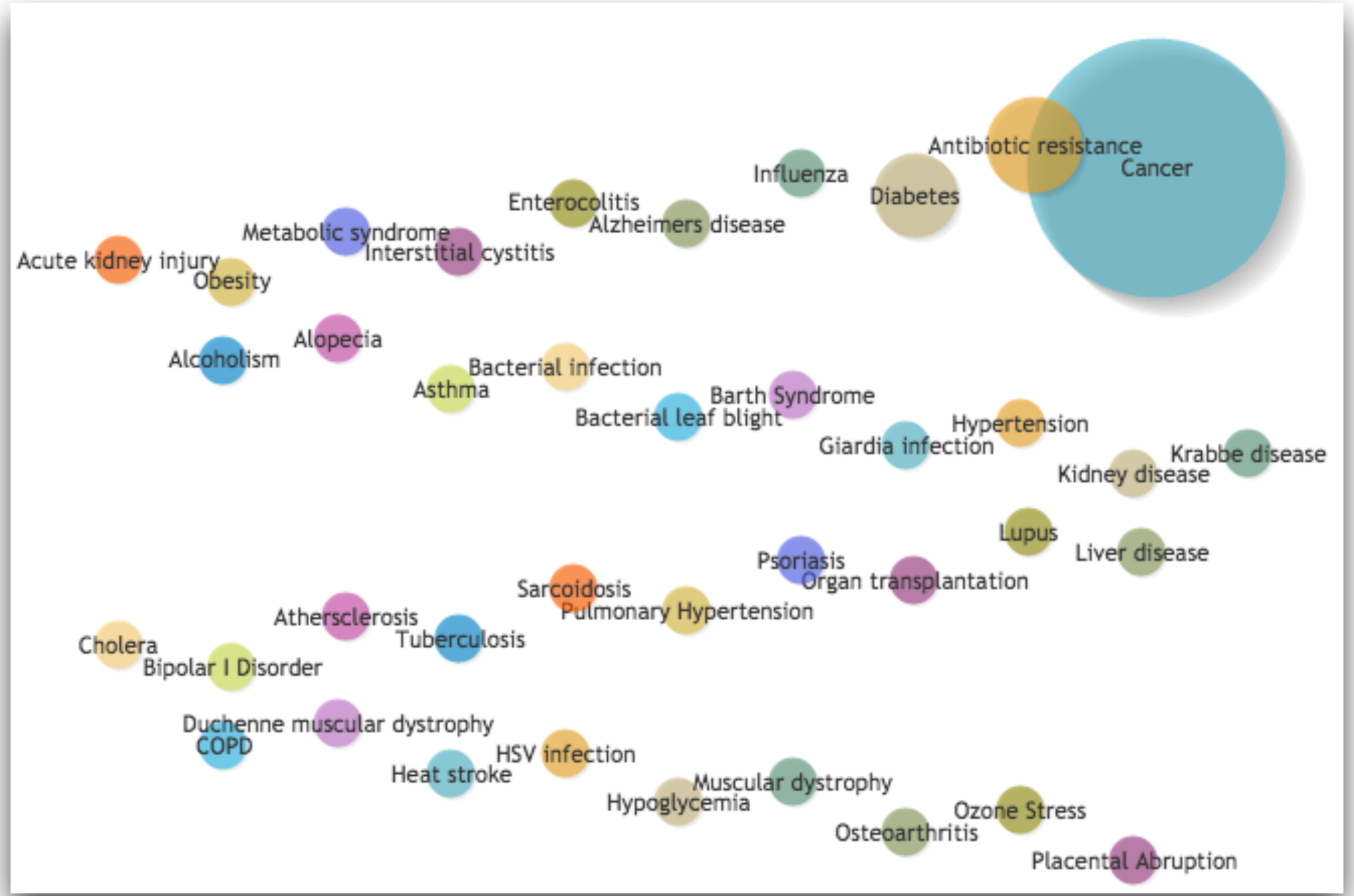
Raw data databases

- Metabolomics Workbench: sample sources



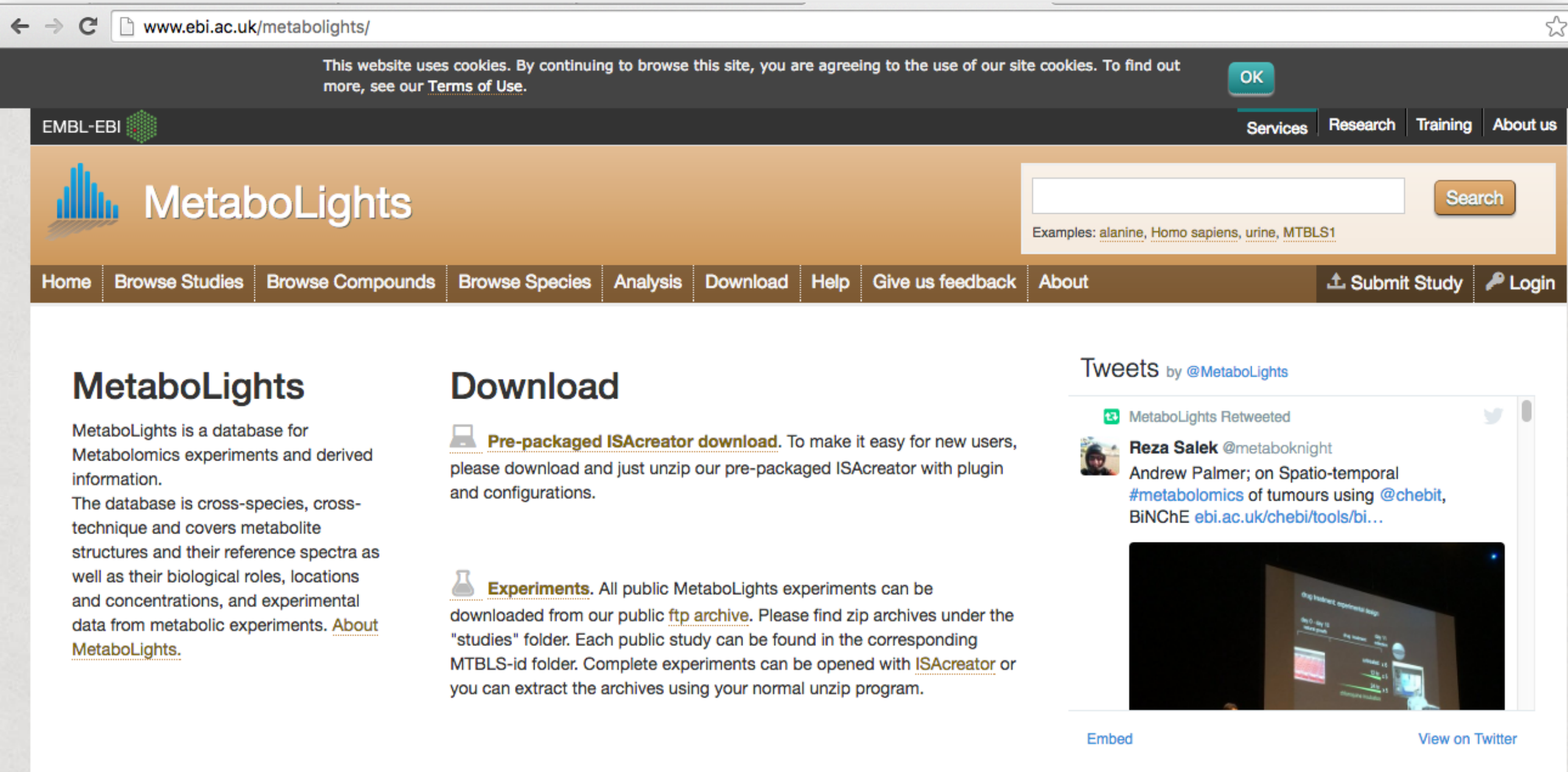
Raw data databases

- Metabolomics Workbench: diseases



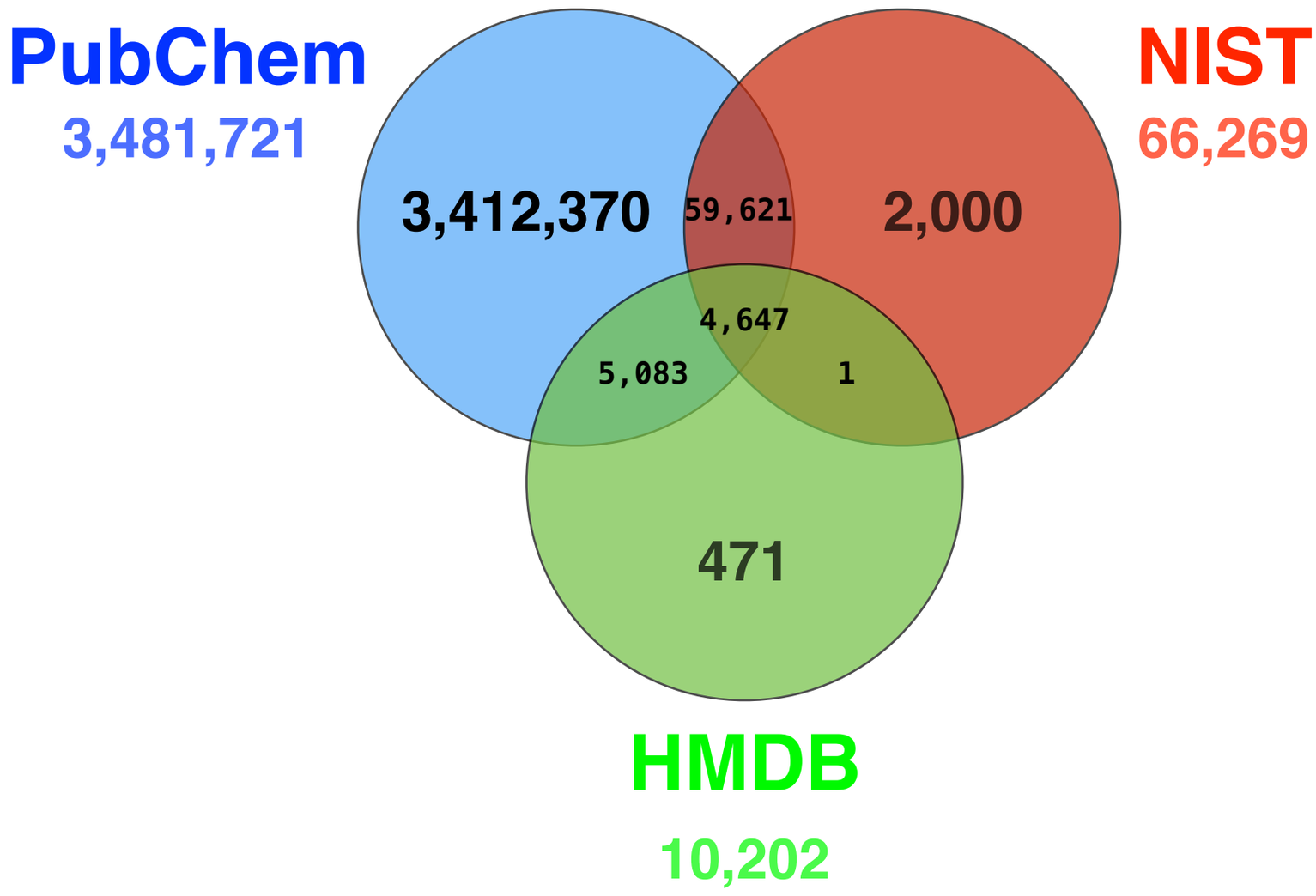
Raw data databases

- MetaboLights



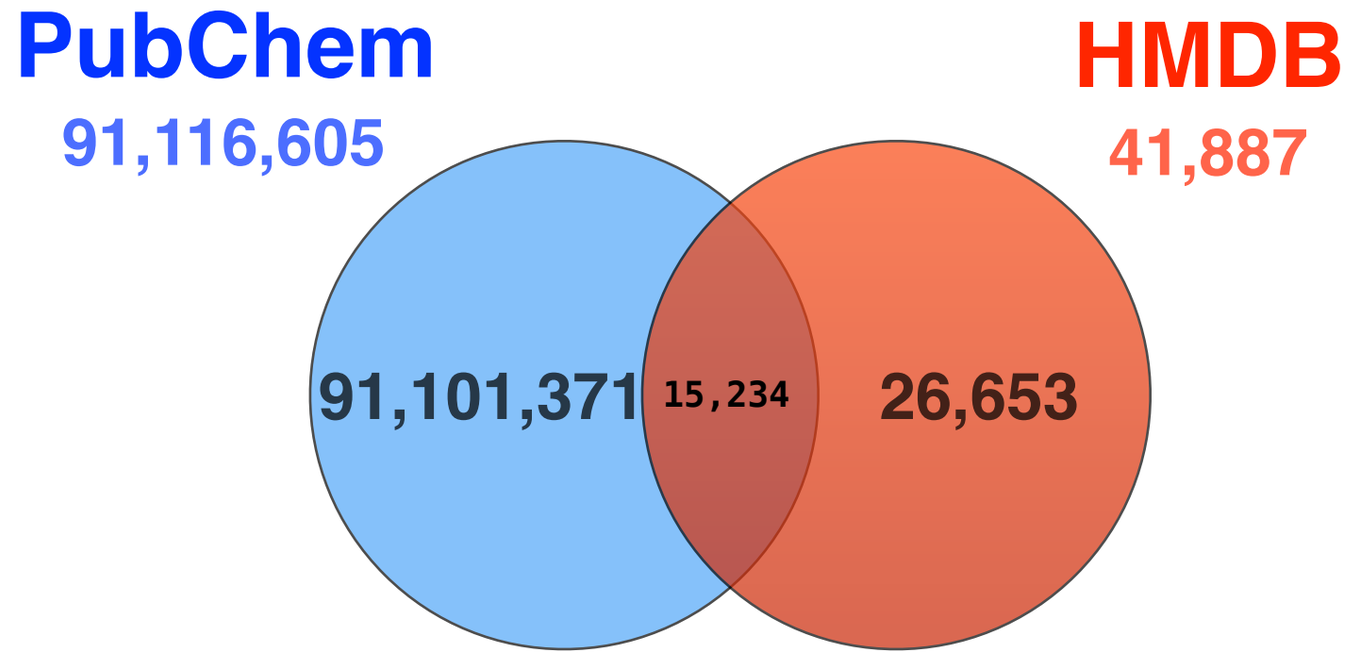
PubChem, NIST, and HMDB, again

- In terms of unique molecular formula



PubChem, NIST, and HMDB, again

- In terms of unique InChi Key



Acknowledgement

- Aleksandr Smirnov

Thank you!