

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 74,462 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, 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: one metabocard

Showing metabocard for L-Proline (HMDB00162)

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

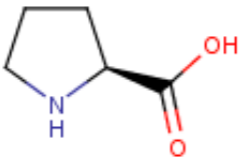
enzymes (29) transporters (1) [Show 30 proteins](#)

[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	 <chem>C1CCNC1C(=O)O</chem>

HMDB: one metabocard

Chemical Formula	C ₅ H ₉ NO 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	<chem>OC(=O)[C@@H]1CCCN1</chem>
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

- Searches

The screenshot shows the HMDB website at www.hmdb.ca/textquery. The navigation bar includes 'Browse', 'Search', 'Downloads', 'About', and 'Contact Us'. A search bar is present with the text 'metabolites' and a 'Search' button. A dropdown menu is open under the 'Search' tab, listing various search methods: ChemQuery Structure Search, Molecular Weight Search, Text Query, Sequence Search, Advanced Search, MS Search, MS/MS Search, GC/MS Search, 1D NMR Search, and 2D NMR Search. Below the dropdown, there is a section titled 'Searching HMDB' with a search input field and a 'Search' button. A text box explains that HMDB supports advanced search based on the Lucene query language, including boolean logic and wild cards. Below this is a table with two columns: 'Example' and 'Description'.

Example	Description
<code>methylhistidine AND poultry</code>	Find all entries containing both methylhistidine and poultry in the metabocard

HMDB: downloads

Protein/Gene Sequences (in FASTA Format)

Data Set	Released on	Protein Sequences	Gene Sequences
All Metabolite Metabolizing Enzymes	2017-07-16	Download	Download

Structures (in SDF Format)

Data Set	Released on	SDF File
Metabolite Structures	2017-07-16	Download

HMDB: downloads

Metabolite and Protein Data (in XML format)

Data Set	Released on	XML File
All Metabolites	2017-07-16	Download
All Proteins	2017-07-16	Download
Urine Metabolites	2017-07-16	Download
Serum Metabolites	2017-07-16	Download
CSF Metabolites	2017-07-16	Download
Saliva Metabolites	2017-07-16	Download
Feces Metabolites	2017-07-16	Download
Sweat Metabolites	2017-07-16	Download

HMDB: downloads

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

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

Secure | <https://pubchem.ncbi.nlm.nih.gov>

Databases >

Upload

Services >

Help

more >

Today's Statistics >

PubChem



BioAssay ?



Compound ?



Substance ?

Go

Limits
Advanced

PubChem

- Statistics (July 19, 2017)

Secure | <https://pubchem.ncbi.nlm.nih.gov>

Databases > Upload Services > Help more > Today's Statistics >

Compounds: 93,553,257
Substances: 234,681,748
BioAssays: 1,252,796
Tested Compounds: 2,395,818
Tested Substances: 3,818,775
RNAi BioAssays: 162
BioActivities: 233,506,437
Protein Targets: 10,340
Gene Targets: 22,094

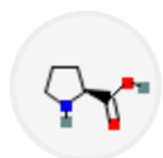
BioAssay ? Compounds Substances

Go Limits Advanced

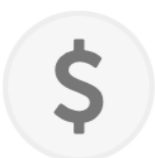
- Information on one compound

L-proline

► Cite this Record



STRUCTURE



VENDORS



DRUG INFO



PHARMACOLOGY



LITERATURE



PATENTS



BIOACTIVITIES

PubChem CID: 145742

Chemical Names: L-proline; 147-85-3; Proline; L-(-)-Proline; 2-pyrrolidinecarboxylic acid; (-)-Proline [More...](#)

Molecular Formula: $C_5H_9NO_2$

Molecular Weight: 115.132 g/mol

InChI Key: ONIBWKKTOPOVIA-BYPYZUCNSA-N

Drug Information: [Drug Indication](#) [Therapeutic Uses](#) [FDA UNII](#)

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

+ Contents



1 2D Structure

2 3D Conformer

3 Biologic Description

4 Names and Identifiers

5 Chemical and Physical Properties

6 Related Records

7 Chemical Vendors

8 Drug and Medication Information

9 Food Additives and Ingredients

10 Pharmacology and Biochemistry

11 Use and Manufacturing

12 Identification

13 Safety and Hazards

14 Toxicity

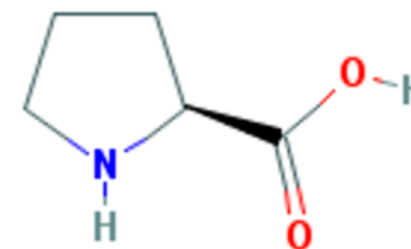
15 Literature

1 2D Structure

🔍 Search

📄 Download

🖼️ Get Image

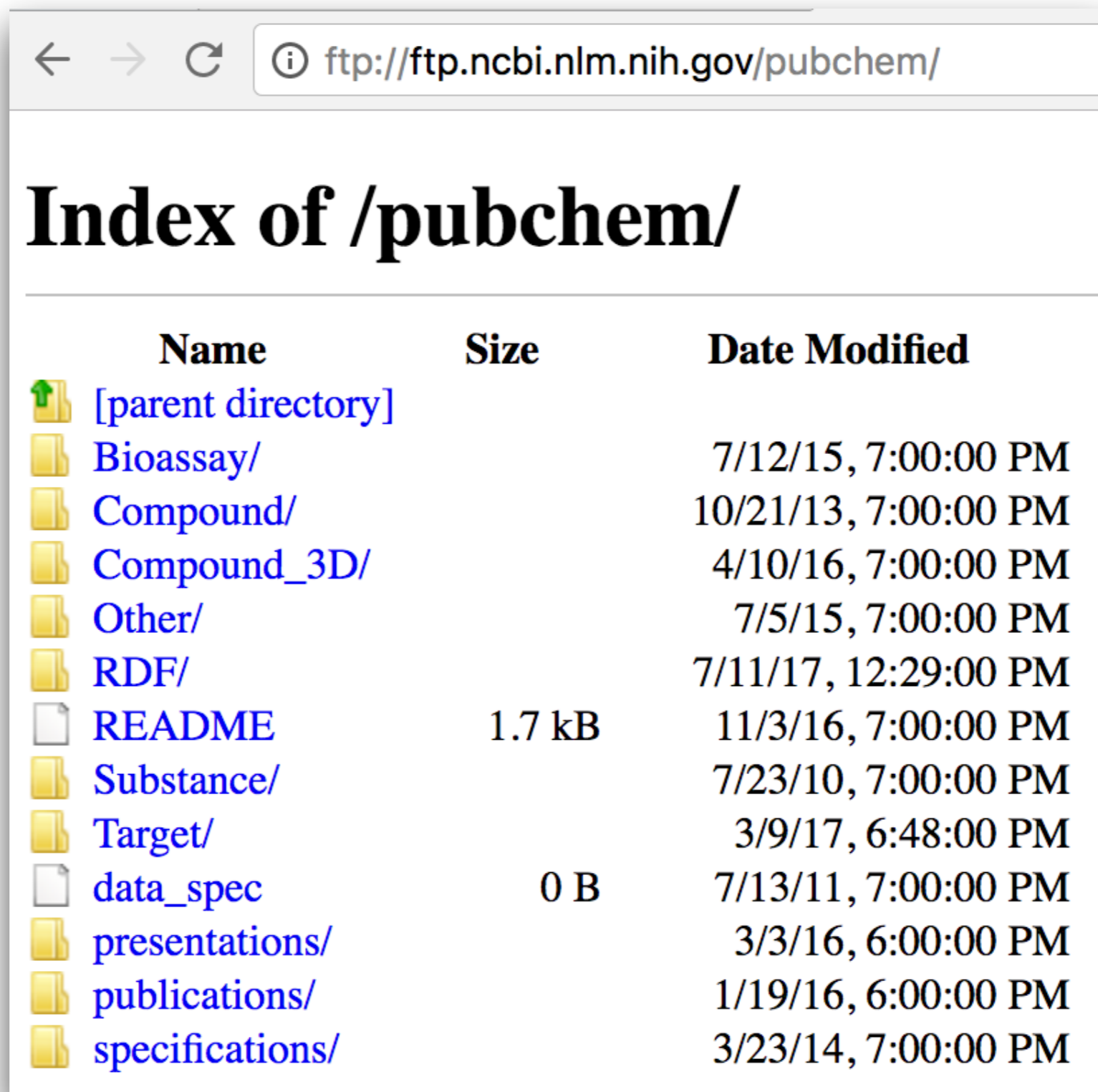


🔍 Magnify

2 3D Conformer














PubChem

- FTP



← → ↻ ⓘ ftp://ftp.ncbi.nlm.nih.gov/pubchem/




















Index of /pubchem/

Name	Size	Date Modified
 [parent directory]		
 Bioassay/		7/12/15, 7:00:00 PM
 Compound/		10/21/13, 7:00:00 PM
 Compound_3D/		4/10/16, 7:00:00 PM
 Other/		7/5/15, 7:00:00 PM
 RDF/		7/11/17, 12:29:00 PM
 README	1.7 kB	11/3/16, 7:00:00 PM
 Substance/		7/23/10, 7:00:00 PM
 Target/		3/9/17, 6:48:00 PM
 data_spec	0 B	7/13/11, 7:00:00 PM
 presentations/		3/3/16, 6:00:00 PM
 publications/		1/19/16, 6:00:00 PM
 specifications/		3/23/14, 7:00:00 PM

PubChem

← → ↻ <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	41.0 MB	3/1/17, 5:00:00 AM
	Compound_000025001_000050000.xml.gz	46.1 MB	6/7/17, 7:44:00 AM
	Compound_000050001_000075000.xml.gz	45.9 MB	7/1/17, 6:11:00 AM
	Compound_000075001_000100000.xml.gz	38.1 MB	4/23/17, 6:30:00 AM
	Compound_000100001_000125000.xml.gz	48.0 MB	6/19/17, 12:51:00 PM
	Compound_000125001_000150000.xml.gz	42.7 MB	5/25/17, 7:45:00 AM
	Compound_000150001_000175000.xml.gz	49.4 MB	6/19/17, 12:51:00 PM
	Compound_000175001_000200000.xml.gz	52.2 MB	4/12/17, 9:04:00 AM
	Compound_000200001_000225000.xml.gz	49.6 MB	4/23/17, 6:30:00 AM
	Compound_000225001_000250000.xml.gz	44.1 MB	3/1/17, 4:59:00 AM
	Compound_000250001_000275000.xml.gz	46.2 MB	1/27/17, 12:18:00 PM
	Compound_000275001_000300000.xml.gz	46.7 MB	12/16/16, 6:00:00 PM
	Compound_000300001_000325000.xml.gz	46.6 MB	12/16/16, 6:00:00 PM
	Compound_000325001_000350000.xml.gz	48.0 MB	12/16/16, 6:00:00 PM
	Compound_000350001_000375000.xml.gz	51.2 MB	4/12/17, 9:04:00 AM
	Compound_000375001_000400000.xml.gz	56.7 MB	4/12/17, 9:04:00 AM
	Compound_000400001_000425000.xml.gz	50.8 MB	1/27/17, 12:19:00 PM
	Compound_000425001_000450000.xml.gz	53.5 MB	6/20/17, 6:05:00 AM

ChemSpider

← → ↻ ⓘ www.chemspider.com

[Home](#) [About us](#) [Web APIs](#) [Help](#) [Sign in](#)

ChemSpider

Search and share chemistry

[Simple](#) [Structure](#) [Advanced](#) [History](#)

Search ChemSpider

Matches any text strings used to describe a molecule.



Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID [?](#)

[What is ChemSpider?](#)

[Search by chemical names](#)

[Search by chemical structure](#)

[Find important data](#)

ChemSpider

What is ChemSpider?

ChemSpider is a free chemical structure database providing fast text and structure search access to over 58 million structures from hundreds of data sources.

Search by chemical names

- Systematic names
- Synonyms
- Trade names
- Database identifiers

Search by chemical structure

- Create structure-based queries
- Draw structures in the web page
- Use structure files from your computer

Find important data

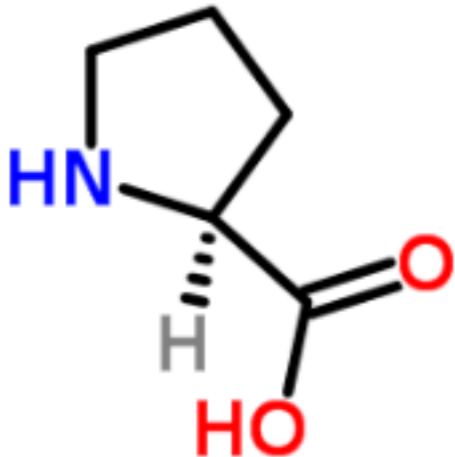
- Literature references
- Physical properties
- Interactive spectra
- Chemical suppliers

59 Million
chemical structures

487
Data sources

ChemSpider


- One compound






The image shows the chemical structure of L-Proline, a five-membered ring containing one nitrogen atom (HN) and one carboxylic acid group (-COOH). The hydroxyl group (-OH) is shown in red, and the hydrogen atom (-H) is shown in grey with a dashed bond, indicating its stereochemistry.

L-Proline

Molecular Formula	C ₅ H ₉ NO ₂
Average mass	115.131 Da
Monoisotopic mass	115.063332 Da
ChemSpider ID	128566

 - 1 of 1 defined stereocentres

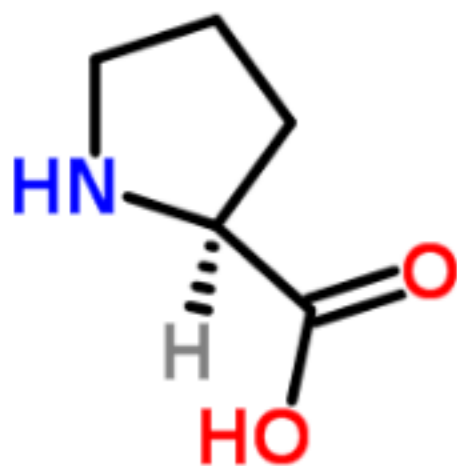
  

▼ **More details:**

This record has not been tagged.


Names and identifiers | [Properties](#) | [Searches](#) | [Spectra](#) | [Vendors](#) | [Articles](#) | [More](#) ▼

Names and Synonyms | Database ID(s)



L-Proline

Molecular Formula	C ₅ H ₉ NO ₂
Average mass	115.131 Da
Monoisotopic mass	115.063332 Da
ChemSpider ID	128566

 - 1 of 1 defined stereocentres



3D



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


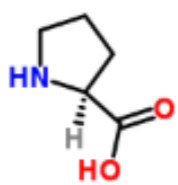

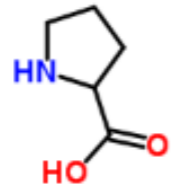
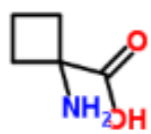

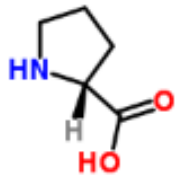
- Compounds with the same molecular formula

Found 501 results

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




1 2 3 4 5

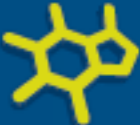
ID	Structure	Molecular Formula	Molecular Weight	# of Data Sources	# of References	# of PubMed	# of RSC
128566  - 1/1 defined		C ₅ H ₉ NO ₂	115.1305	136	11954	28480	3548
594  - 0/1 defined		C ₅ H ₉ NO ₂	115.1305	131	737	28034	3319
80908		C ₅ H ₉ NO ₂	115.1305	110	185	8	5
8640  - 1/1 defined		C ₅ H ₉ NO ₂	115.1305	103	293	701	268

ChEBI

- Chemical Entities of Biological Interest


www.ebi.ac.uk/chebi/init.do


EMBL-EBI 

 ChEBI

Home | Advanced Search | Browse | Documentation | Download | Tools | About ChEBI

Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds.

L-proline 

Search for  only All in ChEBI

Example: [iron*](#), [InChI=1S/H2O/h1H2](#), [water](#)

[Advanced Search](#) | [About ChEBI](#)

ChEBI

- About

[ChEBI](#) > About ChEBI

1. Introduction

Chemical Entities of Biological Interest (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

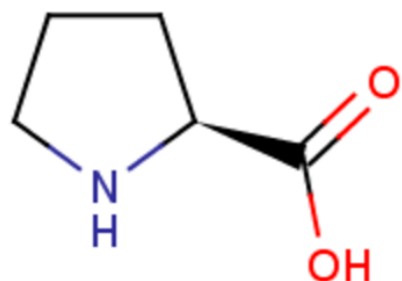
ChEBI Ontology

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](#), [ZINC000000895360](#)

Download

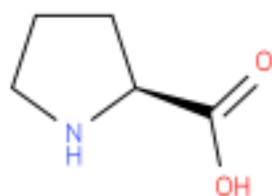
 [Molfile XML SDF](#)

- [Find compounds which contain this structure](#)
- [Find compounds which resemble this structure](#)
- [Take structure to the Advanced Search](#)

[more structures >>](#)

Search Results for All in ChEBI

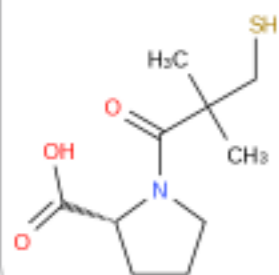
substructure



Edit Search

314 entries found, displaying 1 to 15.

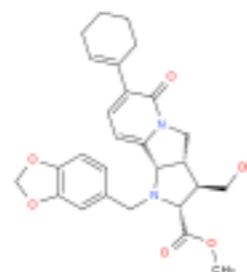
[\(2R\)-1-\(3-mercapto-2,2-dimethyl-1-oxopropyl\)-2-pyrrolidinecarboxylic acid](#)



CHEBI:95254

Stars: ★★☆☆

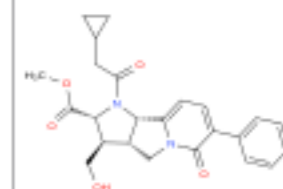
[\(2R,3R,3aS,9bS\)-1-\(1,3-benzodioxol-5-ylmethyl\)-7-\(1-cyclohexenyl\)-3-\(hydroxymethyl\)-6-oxo-3,3a,4,9b-tetrahydro-2H-pyrrolo\[2,3-a\]indolizine-2-carboxylic acid methyl ester](#)



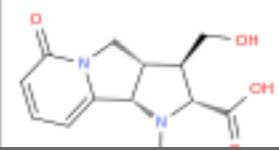
CHEBI:98944

Stars: ★★☆☆

[\(2R,3R,3aS,9bS\)-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](#)



[\(2R,3R,3aS,9bS\)-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: ★★☆☆

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



ChEBI

- One compound

CHEBI:17203 - L-proline

Main

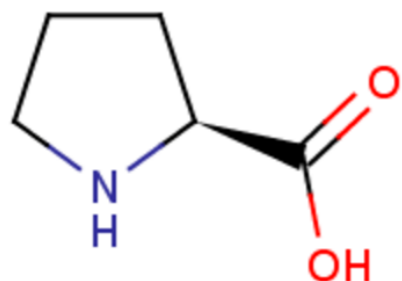
ChEBI Ontology

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](#), [ZINC000000895360](#)

Download

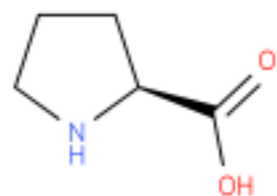
 [Molfile XML SDF](#)

- [Find compounds which contain this structure](#)
- [Find compounds which resemble this structure](#)
- [Take structure to the Advanced Search](#)

[more structures >>](#)

Search Results for All in ChEBI

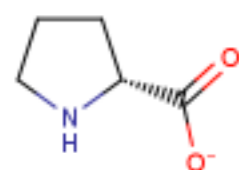
similar structures



Edit Search

125 entries found, displaying 1 to 15.

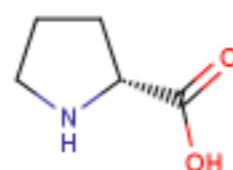
D-prolinate



CHEBI:32867

Stars: ★★★★★

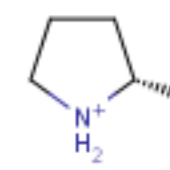
D-proline



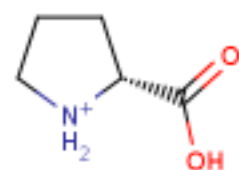
CHEBI:16313

Stars: ★★★★★

D-proline



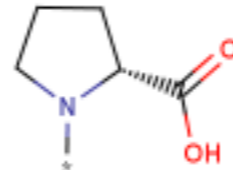
D-prolinium



CHEBI:32868

Stars: ★★★★★

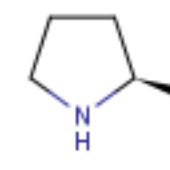
D-prolino group



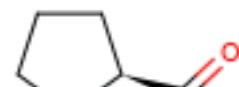
CHEBI:32870

Stars: ★★★★★

L-prolinate



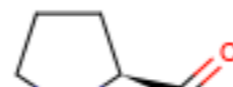
L-proline



CHEBI:17203

Stars: ★★★★★

L-proline zwitterion



CHEBI:60039

Stars: ★★★★★

L-prolinium



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 17
- METLIN
- MassBank
- MoNA
- Gold Metabolome Database
- Feign GC-MS database
- HMDB
- BMRB
- Madison Metabolomics Consortium Database
- BML-NMR
- mzCloud

NIST 17

- Electron ionization mass spectral library
 - 306,622 spectra of 267,376 unique compounds

- MS/MS library: 652,475 spectra
 - 176,594 ion trap spectra for 120,346 different ions of 14,351 compounds
 - 475,881 collision cell spectra (QTOF and tandem quad) spectra for 39,158 different ions of 14,073 compounds

NIST 17

- New ways to identify unknowns
 - Hybrid search
 - Annotated recurring spectral libraries
 - High mass accuracy MS Interpreter

NIST 17 EI library

NIST MS Search 2.3 - [Name search]

File Search View Tools Options Window Help

MSI

DODECANOICACID Clear a-z mainlib From MA

Dodecanethiol-(1)
Dodecanoic acid
Dodecanoic acid, 10-methyl-, meth
Dodecanoic acid, 10-oxo-
Dodecanoic acid, 10-undecen-1-y
dodecanoic acid, 1,10-decanediyl
Dodecanoic acid, 1,1',1''-(1,2,3-pr
Dodecanoic acid, 11-amino-, meth
Dodecanoic acid, 1,1'-biphenyl-4-y
Dodecanoic acid, 1,1'-(dibutylstani
Dodecanoic acid, 1,1-dimethylethy
Dodecanoic acid, 1,1-dimethylprop
Dodecanoic acid, 11-hydroxy-, me
Dodecanoic acid, 11-oxo-, methyl
Dodecanoic acid, 1,2,3-propanetri
Dodecanoic acid, 12-(4-methylphe
Dodecanoic acid, 12-amino-
Dodecanoic acid, 12-amino-, lacta
Dodecanoic acid, 12-bromo-
Dodecanoic acid, 12-hydroxy-
Dodecanoic acid, 12-mercapto-
Dodecanoic acid, 12-(methylaminc
Dodecanoic acid, 1a,2,5,5a,6,9,11
Dodecanoic acid, 1a,2,5,5a,6,9,11
Dodecanoic acid, 1-(hydroxymethy
Dodecanoic acid, 1-methyldecyl e:
Dodecanoic acid, 1-methylethyl es
Dodecanoic acid, 1-methyloctyl es

Names Structures

(mainlib) Dodecanoic acid

Name: Dodecanoic acid
Formula: C₁₂H₂₄O₂
MW: 200 Exact Mass: 200.17763 CAS#: 143-07-7 NIST#: 221043 ID#:
Other DBs: Fine, TSCA, RTECS, EPA, HODOC, NIH, EINECS, IRDB
Contributor: Chemical Concepts
Related CAS#: 8000-62-2; 8045-27-0; 7632-48-6; 203714-07-2
InChIKey: POULHZVOKOAJMA-UHFFFAOYSA-N Non-stereo
10 largest peaks:
73 999 | 60 987 | 43 693 | 41 567 | 57 550 |
55 463 | 129 323 | 29 320 | 71 320 | 85 295 |

Synonyms:
1,n-Dodecanoic acid

Plot/Text Plot

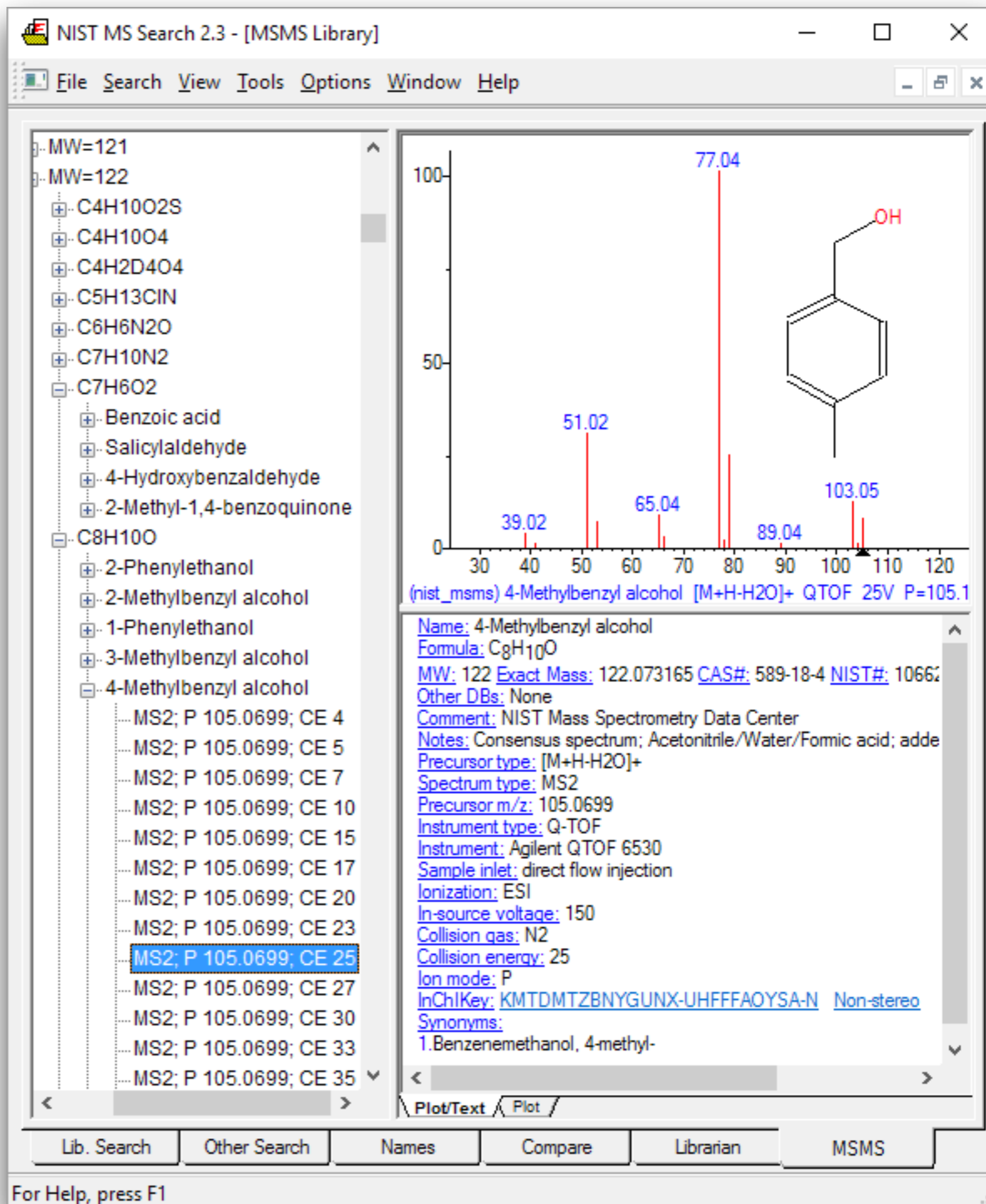
Lib. Search Other Search Names Compare Librarian MSMS

For Help, press F1

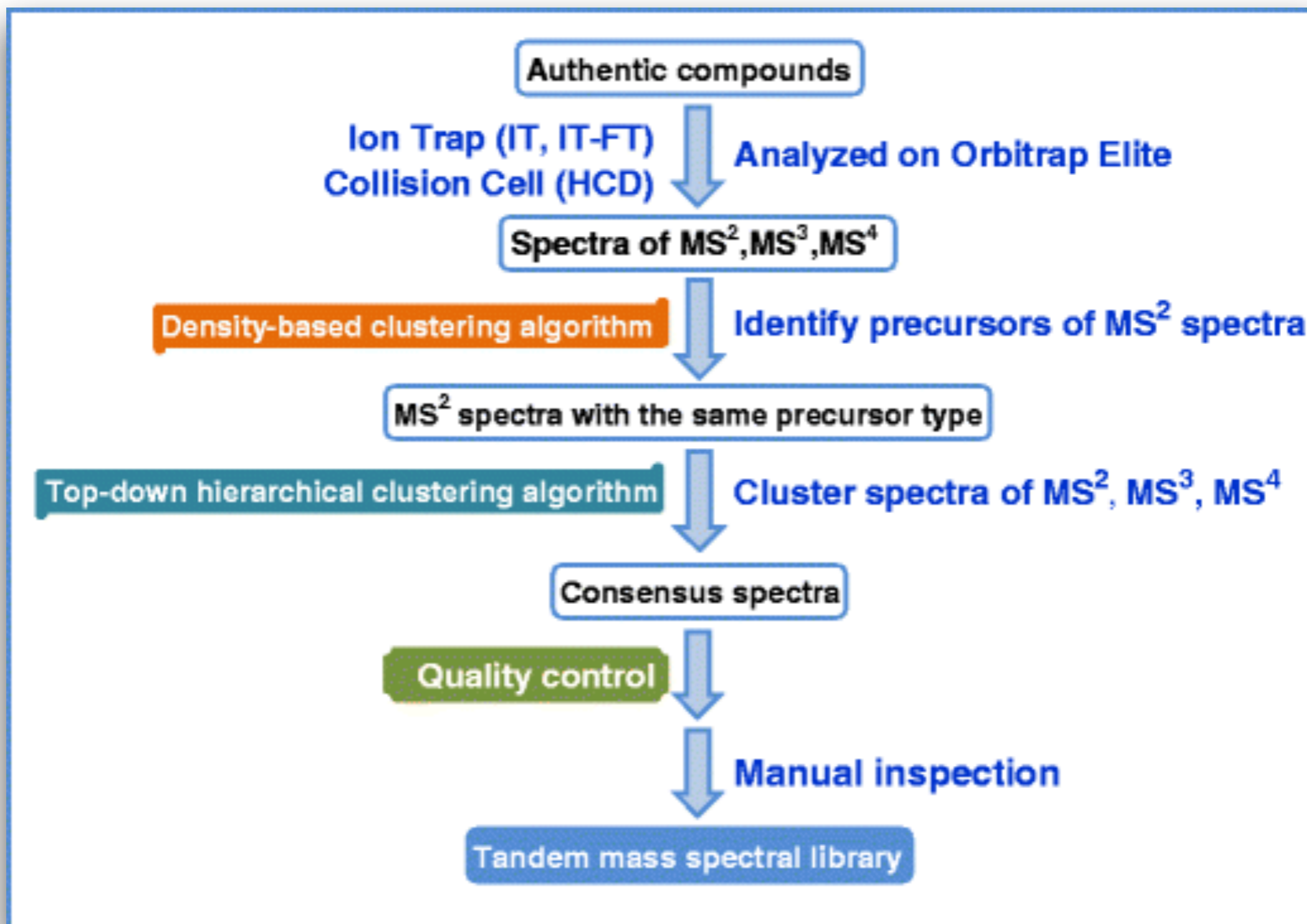
NIST 17 EI library

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

NIST 17 MS/MS library



NIST 17 MS/MS library



NIST 17 MS/MS library

- 13,045 precursor precursor ions
[M+H]⁺, [M+2H]²⁺, [M+H-H₂O]⁺, [M+H-NH₃]⁺,
[M+H-OH]⁺, [M+H+H₂O]⁺, [M+NH₄]⁺, [2M+H]⁺,
[3M+H]⁺, [M+Na]⁺, [M-H+2Na]⁺, [M-2H+3Na]⁺,
[M+K]⁺, [M-H+2K]⁺, [M-2H+3K]⁺, [M+Li]⁺, [M-H+2Li]⁺
- 6,001 negative precursors ions
[M-H]⁻, [M-2H]²⁻, [M-H-H₂O]⁻, [M-H-NH₃]⁻,
[M-H+H₂O]⁻, [M-H+NH₃]⁻, [2M-H]⁻, [3M-H]⁻

NIST 17 MS/MS library

- MS3 and MS4 spectra of the most intense peaks in the MS2 and MS3 spectra, respectively
- New precursors:
 - In-source fragments
 - $[M+H-\text{neutral}]^+$ and $[M-H-\text{neutral}]^-$
 - Fragments from the original target compound within the ESI source

METLIN

Secure https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage

[Home](#)

[Metlin](#) ▾

[isoMETLIN](#)

[Sign Up](#)

[Forgot Password?](#)

CHOLINE ADENOSINE TRIPHOSPHATE CHOLESTEROL TESTOSTERONE GLUCOSE
SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYLCARNITINE THREONINE GLYCEROL
PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE CHOLINE MALIC ACID
TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL OXALOSUCCINIC ACID
PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE CHOLINE MALIC ACID
GLUCOSE PHOSPHATE CHOLESTEROL OXALOSUCCINIC ACID
NICOTINAMIDE ADENINE DINUCLEOTIDE OXALOSUCCINIC ACID GALACTOSE GLYCEROL
SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYLCARNITINE THREONINE GLYCEROL
PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE CHOLINE MALIC ACID
TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL OXALOSUCCINIC ACID GALACTOSE GLYCEROL FUMARATE
GLUCOSE CHOLESTEROL OXALOSUCCINIC ACID GALACTOSE GLYCEROL FUMARATE
NICOTINAMIDE ADENINE DINUCLEOTIDE OXALOSUCCINIC ACID GALACTOSE GLYCEROL
SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYLCARNITINE THREONINE GLYCEROL

METLIN

The original and most comprehensive MS/MS metabolite database

METLIN

Metabolite Searching

METLIN has multiple searching capabilities including single, batch, precursor ion, neutral loss, accurate mass, and fragment searches. The popular [similarity search algorithm](#) for unknown characterization, another METLIN search option, originated on METLIN in 2008.

Tandem Mass Spectrometry

METLIN represents the largest MS/MS collection of data with the database generated at multiple collision energies and in positive and negative ionization modes. The data is generated on multiple instrument types including SCIEX, Agilent, Bruker and Waters QTOF mass spectrometers.

Metabolites

Created in 2004, METLIN now includes 961,829 molecules ranging from lipids, steroids, plant & bacteria metabolites, small peptides, carbohydrates, exogenous drugs/metabolites, central carbon metabolites and toxicants. Over 14,000 metabolites have been individually analyzed and another 200,000 has *in silico* MS/MS data.

METLIN

Secure | https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage#

Home

Metlin ▾

isoMETLIN

 Sign Up

Forgot Password?

Enter email address

Enter password

Login

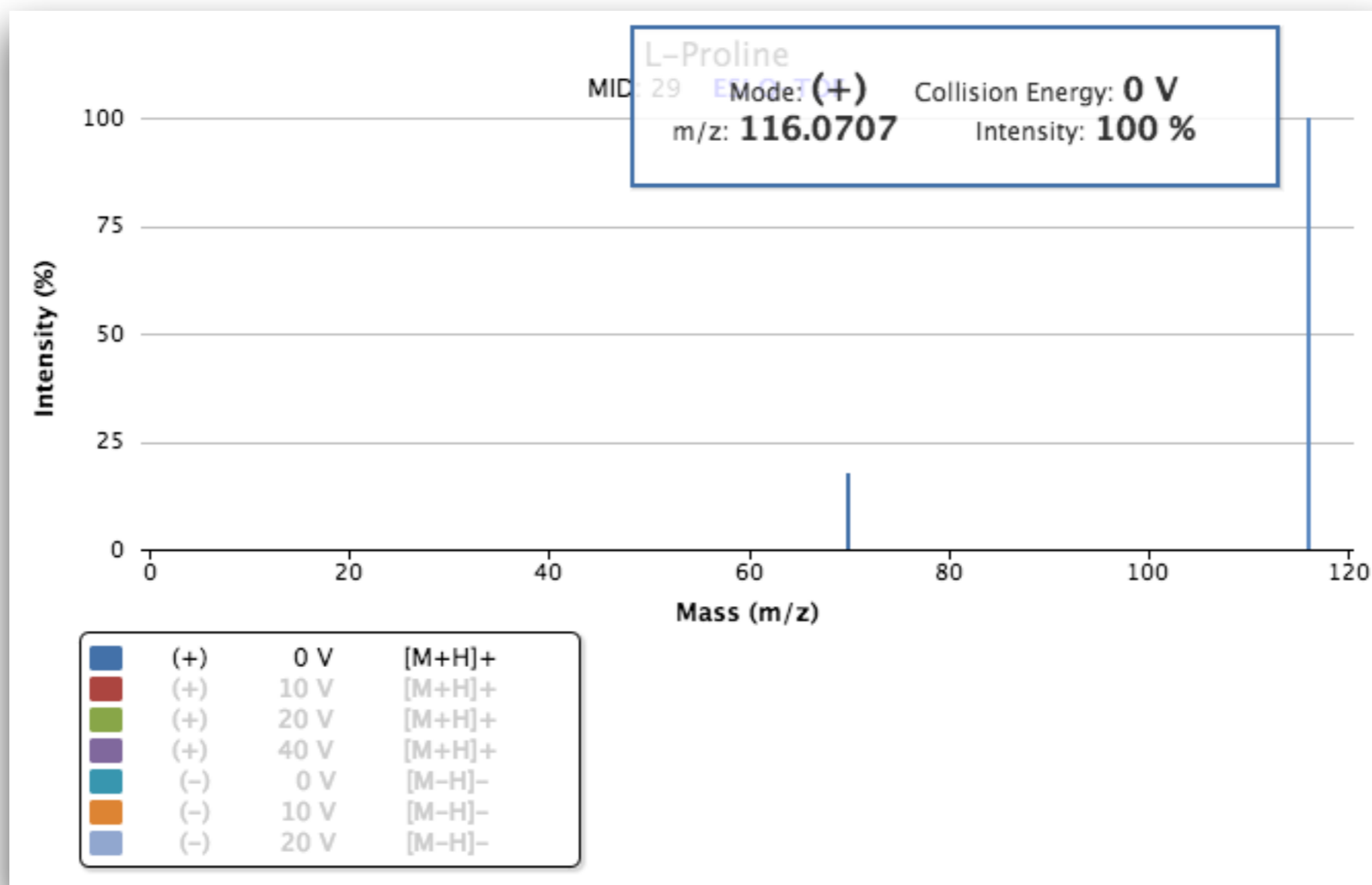
- Simple
- Advanced
- Batch
- Fragment
- Neutral Loss
- MS/MS Spectrum Match

METLIN

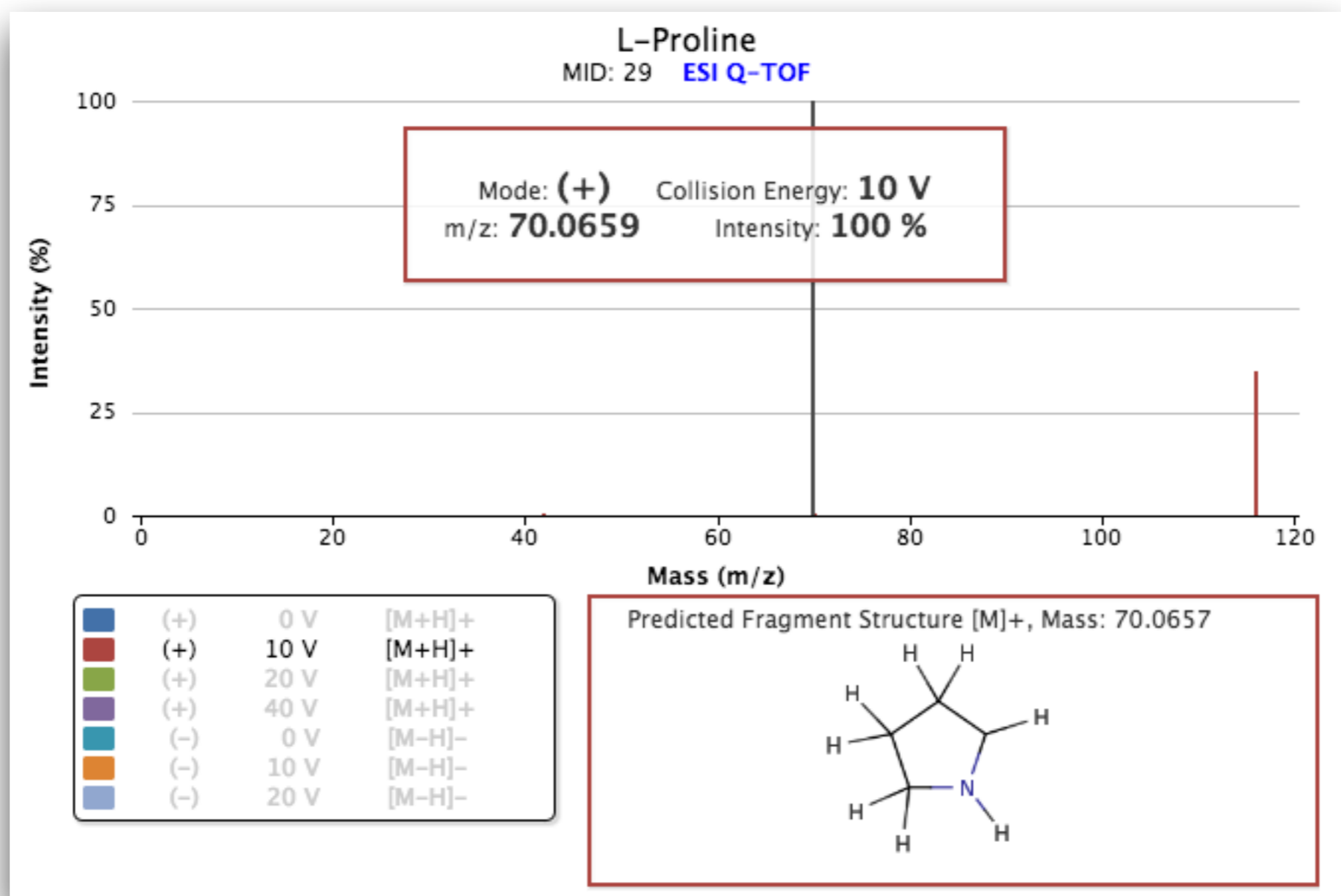
NICOTINAMIDE ADENOSINE TRIPHOSPHATE CHOLESTEROL TESTOSTERONE GLUCOSE
SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYLCARNITINE THREONINE GLYCEROL
PYRUVIC ACID UREA GALACTOSE CHOLINE ADENOSINE CHOLINE MALIC ACID
TESTOSTERONE GLUCOSE PHOSPHATE CHOLESTEROL OXALOSUCCINIC ACID
GLUCOSE CHOLESTEROL OXALOSUCCINIC ACID GALACTOSE GLYCEROL FUMARATE
NICOTINAMIDE ADENINE DINUCLEOTIDE OXALOSUCCINIC ACID GALACTOSE GLYCEROL
SERINE TRYPTOPHAN PHOSPHOCHOLINE ACYLCARNITINE THREONINE GLYCEROL

The original and most comprehensive MS/MS metabolite database

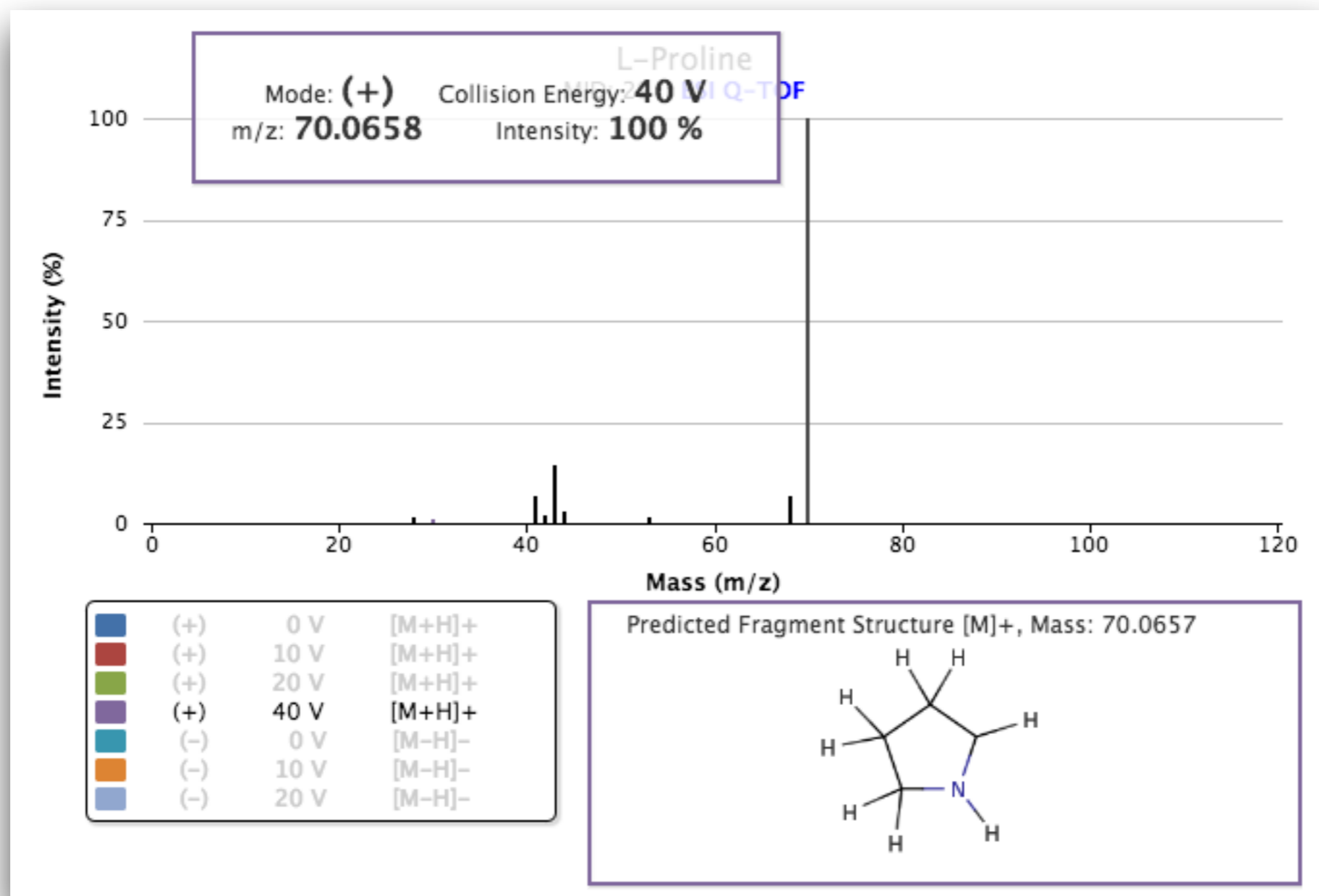
METLIN



METLIN



METLIN



MassBank

www.massbank.jp/?lang=en



High Quality Mass Spectral Database



Database Service

- Spectrum Search
- Quick Search
- Peak Search
- Substructure Search
- Metabolite Prediction
- Spectral Browser
- Batch Service
- Browse Page
- Record Index

MassBank

Database Service

Spectrum Search

Quick Search

Peak Search

Substructure Search

Metabolome Prediction

Spectral Browser

Batch Service

Browse Page

Record Index

Statistics

Publications

Database Service

Spectrum Search



Search similar spectra on a peak-to-peak basis

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

[Spectrum Search](#)

[User's Manual](#)

Quick Search

Keyword search of chemical compounds

Retrieves the chemical compound(s) specified by chemical name or molecular formula, and displays its spectra.

[Quick Search](#)

[User's Manual](#)

MassBank

← → ↻

Peak Search

[Home](#) | [Spectrum](#) | [Quick](#) | [Peak](#) | [Substructure](#) | [Prediction](#) | [Browser](#) | [Batch](#) | [Browse](#) | [Index](#) | MassBank ID:

Search of **Peaks** **Peak Differences**
Search by **m/z-Value** **Molecular Formula**

	m/z	Formula
AND <input type="button" value="v"/>	<input type="text" value="116.0707"/>	<input type="text"/>
AND	<input type="text"/>	<input type="text"/>
AND	<input type="text"/>	<input type="text"/>
AND	<input type="text"/>	<input type="text"/>
AND	<input type="text"/>	<input type="text"/>
AND	<input type="text"/>	<input type="text"/>

Rel.Intensity Tolerance

Instrument Type

EI EI-B
 EI-EBEB
 GC-EI-QQ
 GC-EI-TOF

ESI CE-ESI-TOF
 ESI-ITFT
 ESI-ITTOF
 ESI-QTOF
 LC-ESI-IT

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:

Search Parameters :

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

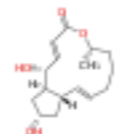
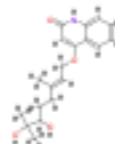
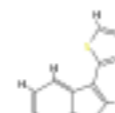
Instrument Type: **LC-ESI-QTOF**
MS Type: **MS2**
Ion Mode: **Positive**

[Edit / Resubmit Query](#)

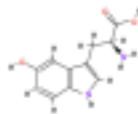
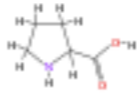
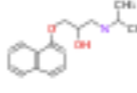
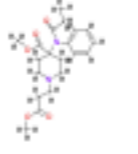

Results : **113 Hit.** (41 - 81 Displayed)

[First](#) [Prev](#) [1](#) [2](#) [3](#) [Next](#) [Last](#) (Total 3 Page)

▼ Results End

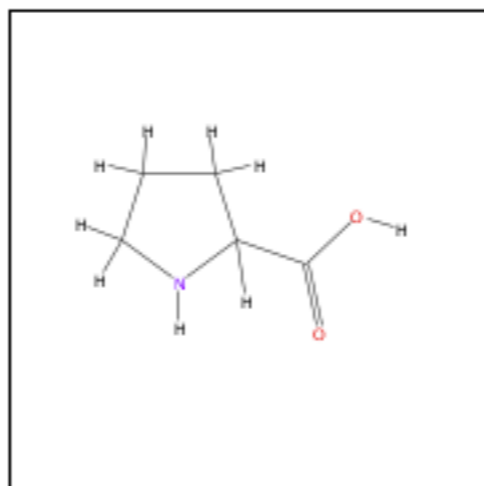
<input type="checkbox"/>	Name ▲	Formula / Structure	ExactMass	ID
<input type="checkbox"/>	<input checked="" type="checkbox"/> Brefeldin-A 1 spectrum	C₁₆H₂₄O₄ 	280.16746	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Bucharaine 1 spectrum	C₁₉H₂₅NO₄ 	331.17836	
<input type="checkbox"/>	<input checked="" type="checkbox"/> Camalexin 2 spectra	C₁₁H₈N₂S 	200.04082	

MassBank

<p><input checked="" type="checkbox"/> Oxitriptan 1</p> <p>1 spectrum</p>	<p>C₁₁H₁₂N₂O₃</p> 	<p>220.08479</p>	
<p><input type="checkbox"/> Proline</p> <p>2 spectra</p> <p> LC-ESI-QTOF; MS2; CE:10 eV; [M+H]⁺ LC-ESI-QTOF; MS2; CE:15 eV; [M+H]⁺ </p>	<p>C₅H₉NO₂</p> 	<p>115.06333</p>	<p>PB000449 PB000450</p>
<p><input checked="" type="checkbox"/> Propranolol</p> <p>1 spectrum</p>	<p>C₁₆H₂₁NO₂</p> 	<p>259.15720</p>	
<p><input checked="" type="checkbox"/> Remifentanil</p> <p>4 spectra</p>	<p>C₂₀H₂₈N₂O₅</p> 	<p>376.19982</p>	
<p><input checked="" type="checkbox"/> S-Lactoylglutathione</p> <p>1 spectrum</p>	<p>C₁₃H₂₁N₃O₈S</p> 	<p>379.10494</p>	

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

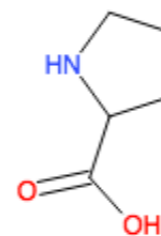
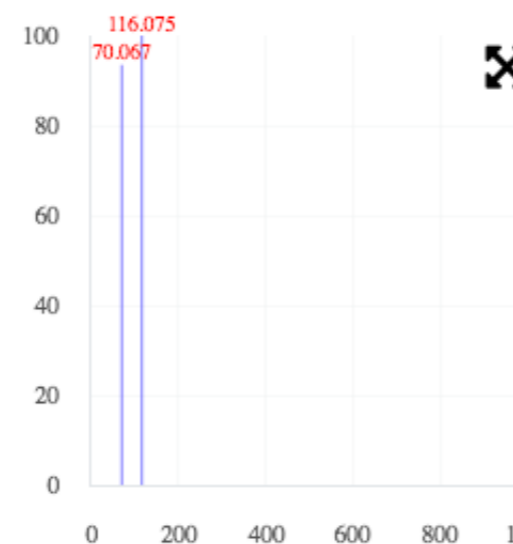
Search...

display generated query



Proline

Score: ★★★★★★★★★★☆☆



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

Originally submitted to the MassBank Spectral Database as [PB000450](#)

MoNA

← → ↻ ⓘ mona.fiehnlab.ucdavis.edu

MoNA - MassBank of North America

 Spectra ▼

 Downloads

 Upload

 Help ▼

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 the 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

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
- MetaboLights

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

Metabolomics workbench

www.metabolomicsworkbench.org



The banner features a circular logo on the left with a molecular structure and the text "Metabolomics Workbench". The main title "METABOLOMICS WORKBENCH" is centered in large white letters. On the right, there is a "Log in / Register" link and a search bar with the placeholder text "Search the Metabolomics Workbench". Below the banner is a navigation menu with links: Home, NIH Data Repository, Databases, Protocols, Standards, Tools, Training / Events, Publications, About, and Search. A welcome message is displayed at the bottom of the banner area.

Metabolomics Workbench

METABOLOMICS
WORKBENCH

Log in / Register

Search the Metabolomics Workbench

[Home](#) | [NIH Data Repository](#) | [Databases](#) | [Protocols](#) | [Standards](#) | [Tools](#) | [Training / Events](#) | [Publications](#) | [About](#) | [Search](#)

Welcome to the UCSD Metabolomics Workbench, a resource sponsored by the Common Fund of the National Institutes of Health.

Metabolomics workbench

The screenshot shows the top navigation bar of the Metabolomics Workbench website. The navigation items are: Home, NIH Data Repository, Databases, Protocols, Standards, Tools, Training / Events, Publications, About, and Search. A dropdown menu is open under 'NIH Data Repository', listing: Overview, Upload / Manage Studies, Browse / Search Studies (circled in red), Analyze Studies, Data Sharing Policy, Tutorials, and FAQ. Below the navigation bar, a welcome message reads: 'Welcome to the Metabolomics Workbench, a resource sponsored by the Common Fund of the National Institutes of Health.' Below this, there are two main buttons: 'Browse and Search Studies' and 'Analyze Studies'. On the right side, there is a 'Quick Links - Key Resources' dropdown menu, a 'YOUR FEEDBACK' section with a request for user input, and an 'Events Calendar' button.

Metabolomics workbench

Browse and Search Studies

- **Browse**
 - [Summary of all projects \(groups of studies\)](#)
 - [Summary of all studies](#)
 - [Bubble plots of studies by disease, sample source, species, pathway and metabolite class](#)
- **Search**
 - **Experimental Projects / Studies**
 - [Data/metadata in experimental projects/studies](#)
 - **Metabolites**
 - [Metabolite data/metadata in experimental studies and Metabolite Database](#)
 - [Untargeted MS data with an m/z value](#)
 - **REST service**
 - [Use the Metabolomics Workbench REST service to retrieve different types of data](#)

Metabolomics workbench

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)
ST000008 (Availability TBA)	Metabolomics Analysis of Population Genetics (PopGen) for Response to Cholera Vaccination)	Homo sapiens	University of North Carolina	NMR	-	-	112	Not available
ST000009	Mixed meal tolerance	Homo sapiens	University of Michigan	MS	2013-04-26	1	228	Uploaded data (8.0M)
ST000010	Lung Cancer Cells 4	Homo sapiens	University of Michigan	MS	2013-05-03	1	78	Uploaded data (2.0M)
ST000011	African Metabolomics	Homo sapiens	University of Michigan	MS	2013-05-04	1	80	Uploaded data (2.2M)

Metabolomics workbench

The image shows a screenshot of the Metabolomics Workbench website. At the top, there is a navigation bar with the following links: Home, NIH Data Repository, Databases, Protocols, Standards, Tools, Training / Events, Publications, About, and Search. Below this, there is a secondary navigation bar with links: Overview, Upload / Manage Studies, Browse / Search Studies, and Analyze Studies. A dropdown menu is open under the 'Tools' link, listing the following options: Overview, Load and analyze your own dataset, Analyze Studies (highlighted with a red oval), MS Searches, REST Service, and External Tools (Links). On the left side of the page, there is a section titled 'Summary of all studies'. Below this section, there is a dashed box containing text: 'Click the Study ID to access detailed study information; download the mwTab (metad... Please refer to our [Data:FAQ](#) and [About:How to Cite](#) pages for information regarding how to cite the metabolomics workbench and datasets that you have uploaded or downloaded.'

Home | NIH Data Repository | Databases | Protocols | Standards | Tools | Training / Events | Publications | About | Search

Overview | Upload / Manage Studies | Browse / Search Studies | Analyze Studies

Summary of all studies

Click the Study ID to access detailed study information; download the mwTab (metad... Please refer to our [Data:FAQ](#) and [About:How to Cite](#) pages for information regarding how to cite the metabolomics workbench and datasets that you have uploaded or downloaded.

Tools dropdown menu:

- Overview
- Load and analyze your own dataset
- Analyze Studies
- MS Searches
- REST Service
- External Tools (Links)

Metabolomics workbench

Analyze Studies

MS/NMR studies identifying named metabolites

Normalization and averaging

- Metabolite averages per experimental factor for studies
- Normalize study data
- Relative log abundance plots

Clustering and correlation

- Hierarchical or heatmap cluster analysis
- Clustered correlation analysis
- Network analysis on correlated metabolites

Univariate analysis

- Volcano plot analysis

Multivariate analysis

- Principal component analysis
- Linear discriminant analysis
- Partial least-squares discriminant analysis

Classification and feature analysis

- OPLS-DA analysis and VIP projection
- Random Forest analysis and VIP projection

Mapping metabolites to biochemical pathways

Metabolomics workbench

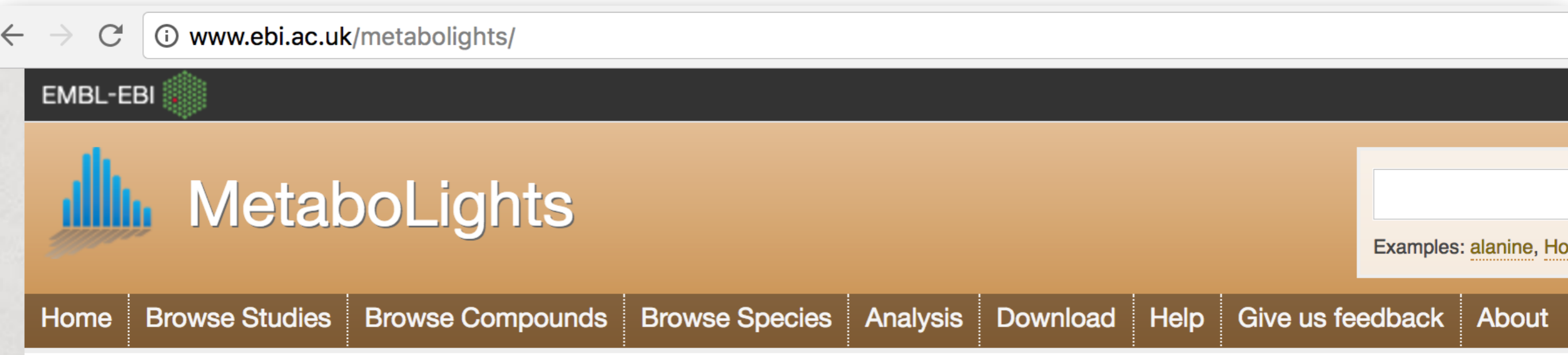
Principal Component Analysis on MS studies

This analysis uses the ["prcomp"](#) function of the R statistics environment
Click on links below to perform analysis.

PCA	Study ID	Study Title	MS Analysis Type
Run	ST000001	Fatb Induction Experiment (FatBIE)	GCMS positive ion mode
Run	ST000002	Intestinal Samples II pre/post transplantation	GCMS positive ion mode
Run	ST000003	Metabolomic analysis of mouse embryonic fibroblasts, embryonic stem cells, and induced pluripotent stem cells	LC/MS positive ion mode
Run	ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Fatty acid/eicosanoid MS quantitative analysis with deuterated standards
Run	ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Phospholipid MS quantitative analysis with odd-chain standards
Run	ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Sphingolipid MS quantitative analysis with C12 standards
Run	ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Sterol MS quantitative analysis with deuterated standards
Run	ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Prenol/Cardiolipin MS quantitative analysis with odd-chain(CL) or stable analog standards
Run	ST000005	Timecourse on RAW 264.7 cells treated with Kdo2-Lipid A and compactin	Triacylglycerol MS quantitative analysis with deuterated standards

Raw data databases

- MetaboLights



The screenshot shows a web browser window with the address bar containing www.ebi.ac.uk/metabolights/. The page header features the EMBL-EBI logo and the MetaboLights logo, which includes a blue bar chart icon. A search input field is present with the text "Examples: alanine, Ho...". The main navigation menu consists of the following links: Home, Browse Studies, Browse Compounds, Browse Species, Analysis, Download, Help, Give us feedback, and About.

MetaboLights

MetaboLights: a database for metabolomics experiments and derived information

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 metabolic experiments. MetaboLights is the recommended Metabolomics repository for a number of [leading journals](#). We will provide search services around spectral similarities and chemical structures.

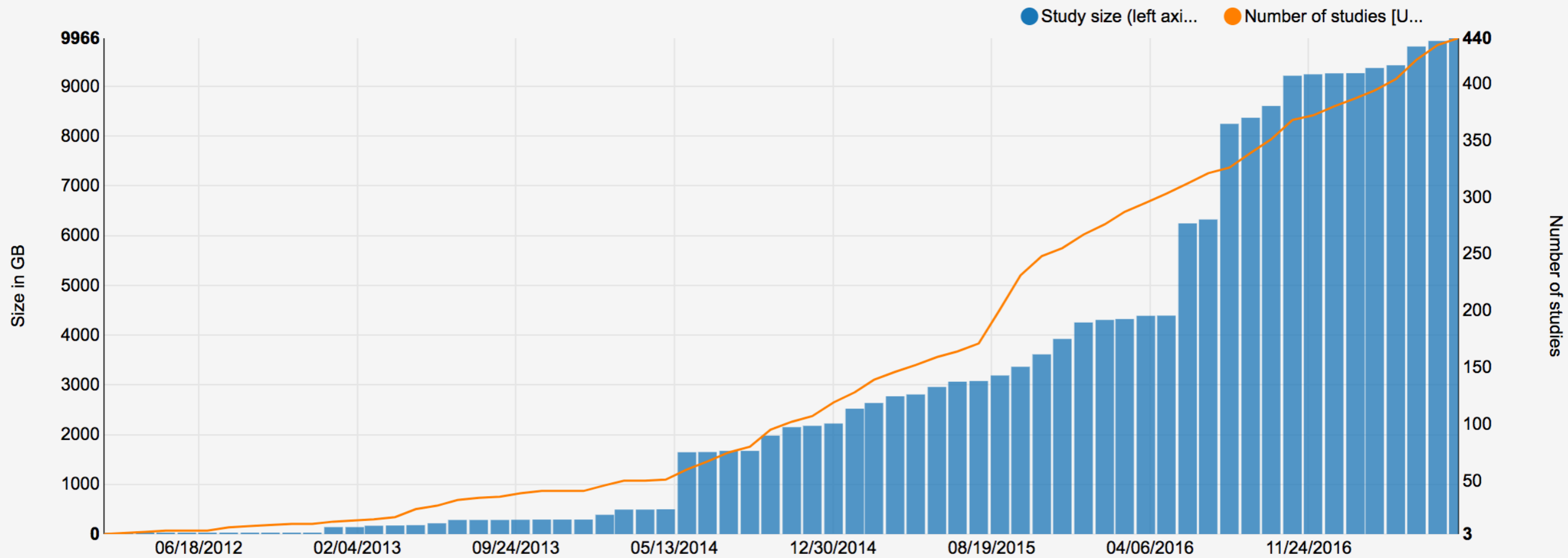
We offer user-submission tools and have strong reporting capabilities. We will utilise and further develop de-facto standard formats where various components are encapsulated, such as the encoded spectral and chromatographic data, and associated information about the chemical structure, as well as metadata describing assays and the study as a whole.

MetaboLights semantic quality will be based on various controlled vocabularies linking data to existing resources such as [BRENDA](#) (tissue ontology) and other established ontologies. We are dedicated to collaborate closely with major parties in world-wide metabolomics communities, such as the [Metabolomics Society](#) and the associated Metabolomics Standards Initiative (MSI).

MetaboLights

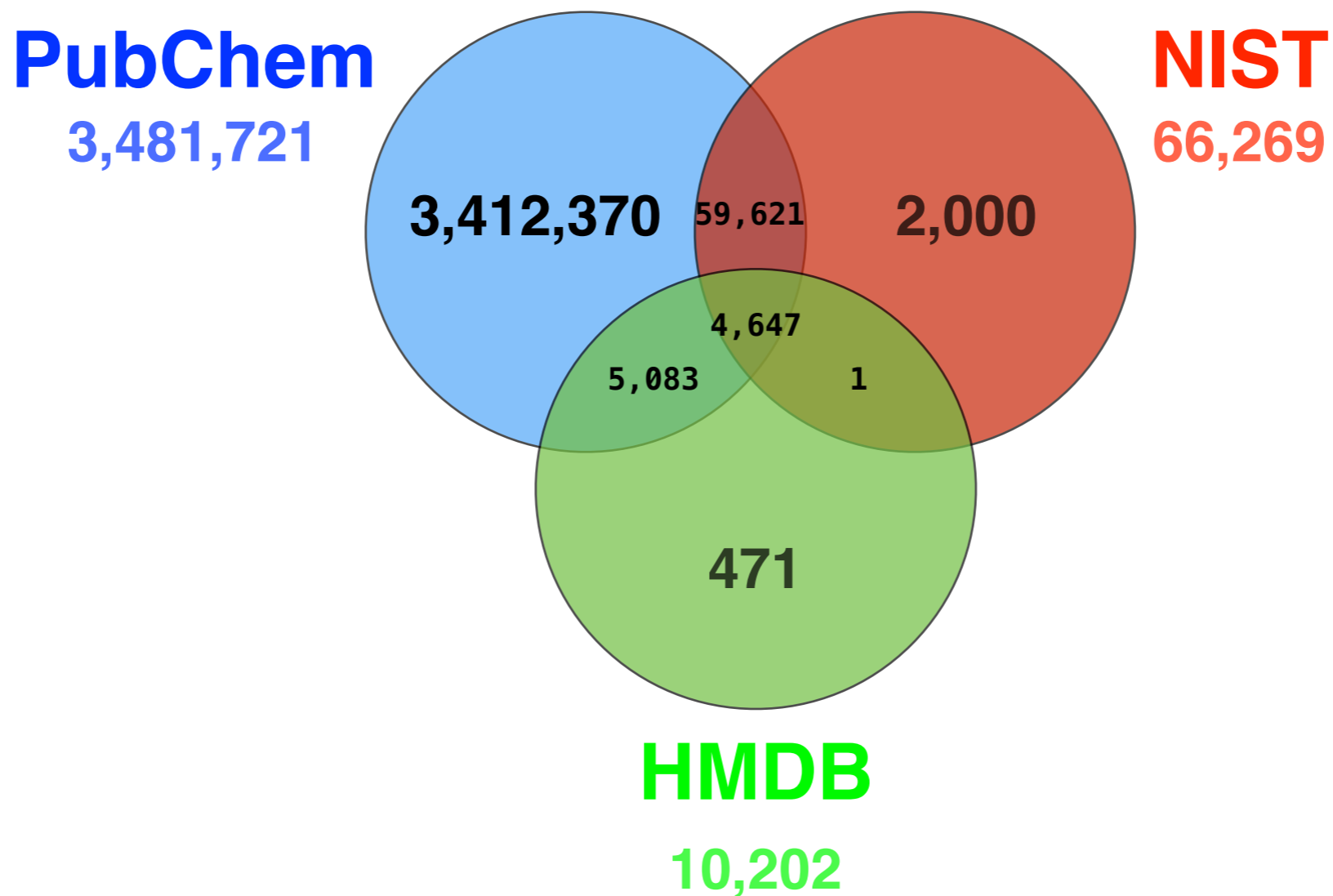
MetaboLights Statistics

Data growth over time



PubChem, NIST, and HMDB, again

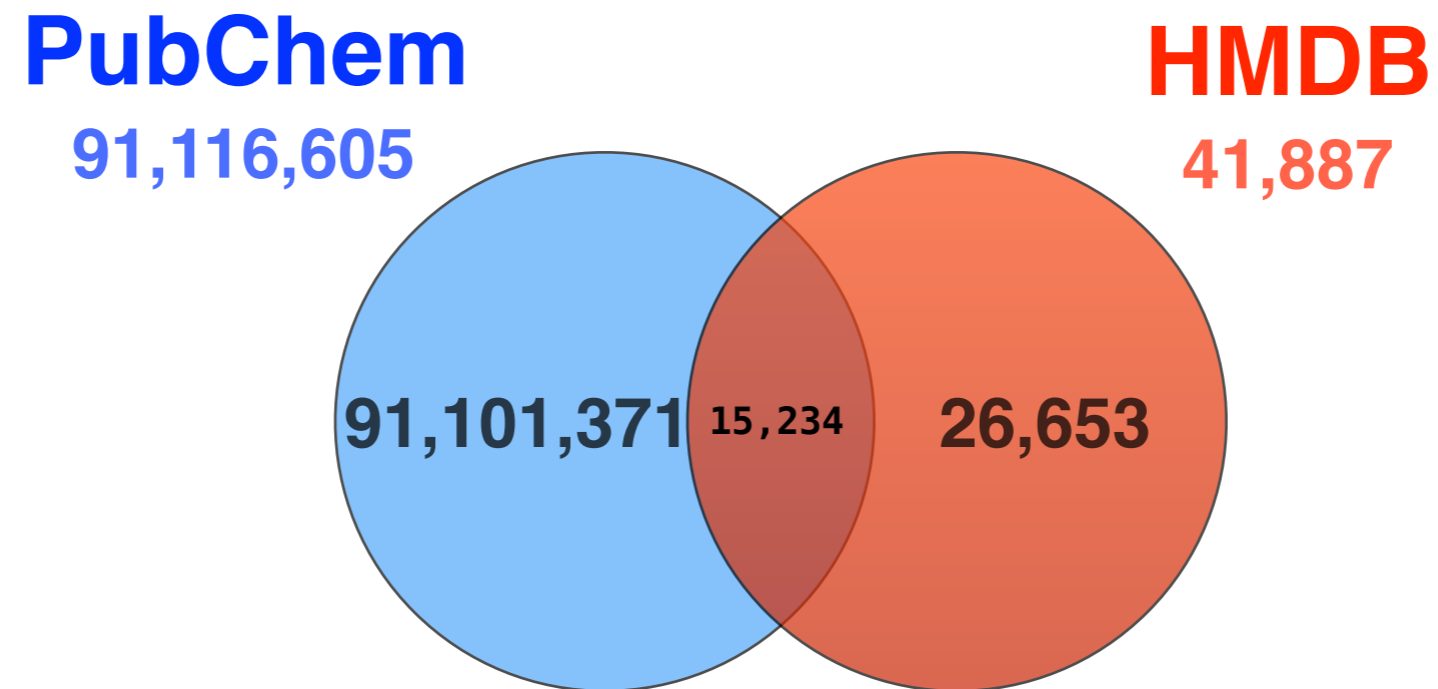
- In terms of unique molecular formula



**based on database
version in 2016**

PubChem, NIST, and HMDB, again

- In terms of unique InChi Key



**based on database
version in 2016**

Acknowledgement

- Aleksandr Smirnov, Ph.D.



Thank you!