

Overview of open resources to support automated structure verification and elucidation

Antony Williams¹ and Emma Schymanski²

1. *National Center for Computational Toxicology, US EPA, Research Triangle Park, Durham, NC, USA.*
2. *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg.*

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

*March 2018
ACS Spring Meeting, New Orleans*

- Our focus for the session:
 - Access to data to support automated structure verification and elucidation – NMR and MS
 - Data quality, curation and validation – and a call to action
 - OPENness is here – Open Access, Data, Source
 - Data standards – we already have them and there are more coming
 - Vendors providing and using available data
 - There are tools USING these data for Structure Elucidation
 - Cannot be an exhaustive review...

- All published structures and spectra will be available from all published articles for repurposing and reuse in standard formats (preferably not necessarily Open!)
- Scientists are building open approaches
 - MS fragmentation
 - NMR shift prediction
 - Structure generators
 - Computer-Assisted Structure Elucidation (CASE)
- Are we there yet???

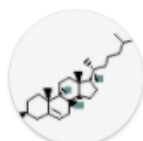
- We have achieved ideal scenario right?
- No – PDF figures in Supplementary Info is still the default position
- There is a need for public databases of spectral data. There ARE some out there.
- Just like Wikipedia we are primarily consumers rather than contributors...

- There are many sites that “share” spectral data. Generally in non-open formats
- There are rich resources
- Cannot easily be used to serve automated structure verification and elucidation.

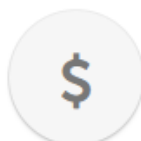


Cholesterol

► Cite this Record



STRUCTURE



VENDORS



DRUG INFO



PHARMACOLOGY



LITERATURE



PATENTS



BIOACTIVITIES

PubChem CID: 5997

Chemical Names: Cholesterol; 57-88-5; Cholesterin; Cholest-5-en-3beta-ol; Cholesteryl alcohol; Cholestrin [More...](#)

Molecular Formula: $C_{27}H_{46}O$

Molecular Weight: 386.664 g/mol

InChI Key: HVYWMOMLDIMFJA-DPAQBDIFSA-N

Drug Information:

[Therapeutic Uses](#)

[Clinical Trials](#)

[FDA UNII](#)

Safety Summary: [Laboratory Chemical Safety Summary \(LCSS\)](#)

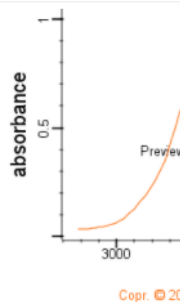
PubChem - Spectral Links

4.4.1 Infrared Spectra

Infrared Spectra: 1 of 1 (ATR-IR Spectra)

Instrument Name	Bio-Rad FTS
Technique	ATR-Neat (DuraSamplIR II)
Source of Spectrum	Forensic Spectral Resear
Source of Sample	Sigma-Aldrich Company
Catalog Number	C8667
Lot Number	SLBC7554V
Copyright	Copyright © 2014-2017

Thumbnail



4.4.2 1D NMR Spectra

1D NMR Spectra: 1 of 16 (1H NMR Spectra)

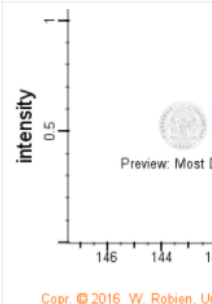
1H NMR Spectra	1D NMR Spectrum 2491 - 1H NMR Spectrum (HMDB0062453)
13C NMR Spectra	1D NMR Spectrum 3188 - 13C NMR Spectrum (HMDB0062453)

from Human Metabolome Database (HMDB)

1D NMR Spectra: 2 of 16 (13C NMR Spectra)

Copyright © 2016 W. Robien Rights Reserved.

Thumbnail



4.4.3 Mass Spectrometry

4.4.3.1 GC-MS

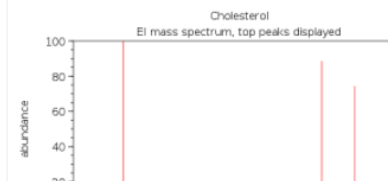
1. GC-MS Spectrum 2260
2. GC-MS Spectrum 4679
3. GC-MS Spectrum 4680
4. GC-MS Spectrum 4681
5. GC-MS Spectrum 4682
6. GC-MS Spectrum 4683

from Human Metabolome Database (HMDB)


<< < 1 of 3 > >>

NIST Number	332884
Library	Main library
Total Peaks	274
m/z Top Peak	43
m/z 2nd Highest	55
m/z 3rd Highest	57

Thumbnail



Spectral Links to Partial Data




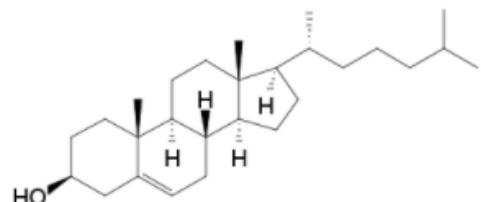
[Buy Now](#) [Log In](#) [BIO-RAD](#)

[Cite](#) [Share](#) [Follow](#) [Feedback](#)

CHOLESTEROL;VITAMIN-D-DERIVATIVE

Compound with one FTIR, fifteen NMR, and one Raman spectra

 Download



SpectraBase Compound ID	DW7SMDXE5b3
InChI	InChI=1S/C27H46O/c1-18(2)7-6-8-19(3)23-11-12-24-22-10-9-20-17-21(28)13-15-26(20,4)25(22)14-16-27(23,24)5/h9,18-19,21-25,28H,6-8,10-17H2,1-5H3/t19-,21+,22+,23-,24+,25+,26+,27-/m1/s1
InChIKey	HVYWMOMLIMFJA-DPAQBDFSA-N Google Search
Mol Weight	386.7 g/mol
Molecular Formula	C27H46O
Exact Mass	386.354866 g/mol

Spectra [Synonyms](#) [Similar Compounds](#) [Literature](#) [Related Ambiguous Compounds](#)

¹³C NMR

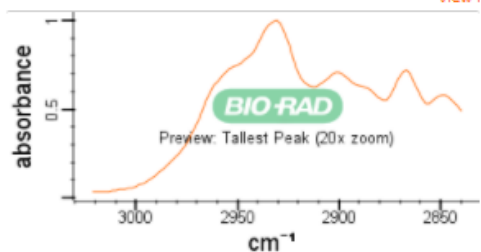
ATR-IR

Raman

15

1

1



absorbance

cm⁻¹

Preview: Tallest Peak (20x zoom)

View Full Record

Copyright	Copyright © 2014-2017 Bio-Rad Laboratories, Inc. All Rights Reserved.
Source of Sample	Sigma-Aldrich Company Llc
Source of Spectrum	Forensic Spectral Research
Catalog Number	C8667

7

SDBS – Free Not Open

Spectral Database for Organic Compounds SDBS

[Japanese](#)[Introduction](#)[Disclaimer](#)[HELP](#)[Contact](#)[What's New](#)[RIO-DB](#)[FAQ](#)[LINK](#)

Welcome to Spectral Database for Organic Compounds, SDBS.

This is a free site organized by [National Institute of Advanced Industrial Science and Technology \(AIST\)](#), Japan.

NMR: *T.Yamaji, T.Saito, K.Hayamizu, M.Yanagisawa and O.Yamamoto*

MS: *N.Wasada*

ESR: *K.Someno*

Access to this database is free of charge. However we request visitors to our database not to download more than 50 spectra and/or compound information in one day. All accesses are recorded. It is prohibited that you use any information of SDBS for profit-making or commercial use without obtaining proper permission from us. If more spectra are required for some specific purpose or commercial use, you should consult us and describe the intended usage or purpose of our SDBS.

Disclaimer

We are doing our best to compile high quality databases. However, there are no such databases without any errors or mistakes. We make no warranties to those effects and shall not be liable for any damage that may result from errors in the database. You can check the page [Known Errors of Mistakes in SDBS](#). When you find new errors or mistakes, please inform us by email (see [this page](#) for contact details).

Access to this database is free of charge. However we request visitors to our database not to download more than 50 spectra and/or compound information in one day. All accesses are recorded. It is prohibited that you use any information of SDBS for profit-making or commercial use without obtaining proper permission from us. If more spectra are required for some specific purpose or commercial use, you should consult us and describe the intended usage or purpose of our SDBS.

SDBS – Free Not Open

Spectral Database for Organic Compounds SDBS

[Japanese](#)[Introduction](#)[Disclaimer](#)[HELP](#)[Contact](#)[What's New](#)[RIO-DB](#)[FAQ](#)[LINK](#)[AIST](#)

SDBS Information

SDBS No.: 887

Compound Name:
cholesterol

Molecular Formula: $C_{27}H_{46}O$

Molecular Weight: 386.6

CAS Registry No.:
57-88-5

Derivatives:
display in a separate page
[SDBS Structures Web \(on trial since 2013-05-14\)](#)

Spectral Code:

Mass :

^{13}C NMR : in $CDCl_3$

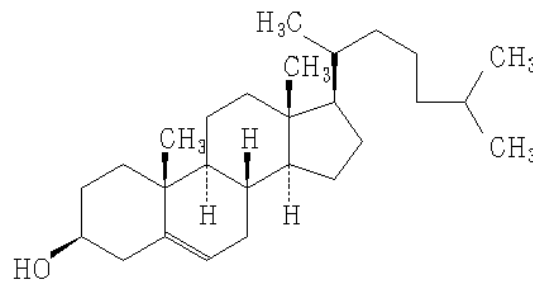
1H NMR : 400 MHz in $CDCl_3$

IR : nujol mull

IR : KBr disc

[Chemical Information:](#)

SDBS No.: 887 CAS Registry No.: 57-88-5
Molecular Formula: $C_{27}H_{46}O$ Molecular Weight: 386.6
SDBS-NO= 887
CHOLESTEROL



Compound Name:
cholesterol
5-cholesten-3beta-ol

[\(c\) National Institute of Advanced Industrial Science and Technology \(AIST\).](#)

SDBS – Free Not Open

Spectral Database for Organic Compounds SDBS

[Japanese](#)[Introduction](#)[Disclaimer](#)[HELP](#)[Contact](#)[What's New](#)[RIO-DB](#)[FAQ](#)[LINK](#)[AIST](#)

SDBS Information

SDBS No.: 887

Compound Name:
cholesterol

Molecular Formula: $C_{27}H_{46}O$

Molecular Weight: 386.6

CAS Registry No.:
57-88-5

Derivatives:
display in a separate page
[SDBS Structures Web \(on trial since 2013-05-14\)](#)

Spectral Code:

[Mass :](#)

[\$^{13}C\$ NMR : in \$CDCl_3\$](#)

[\$^1H\$ NMR : 400 MHz in \$CDCl_3\$](#)

[IR : nujol mull](#)

[IR : KBr disc](#)

[Chemical Information:](#)

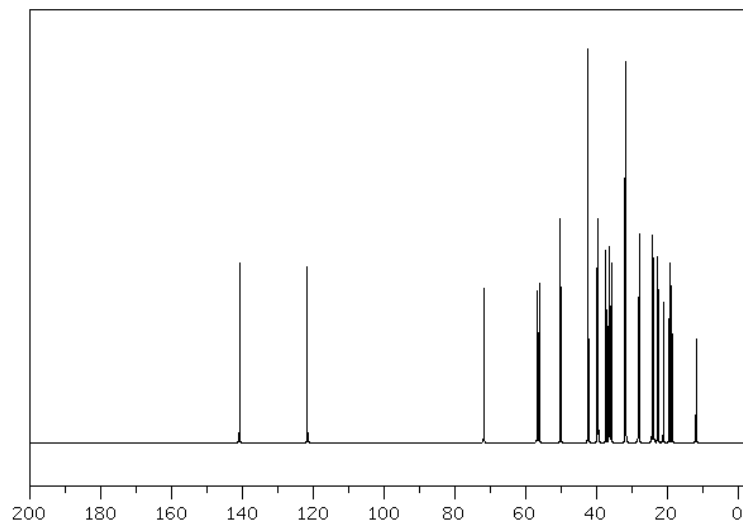
SDBS- ^{13}C NMR SDBS No. 887CDS-05-861

$C_{27}H_{46}O$

cholesterol

25.16 MHz

0.141 g : 1.5 ml $CDCl_3$



ChemSpider

Search and share chemistry

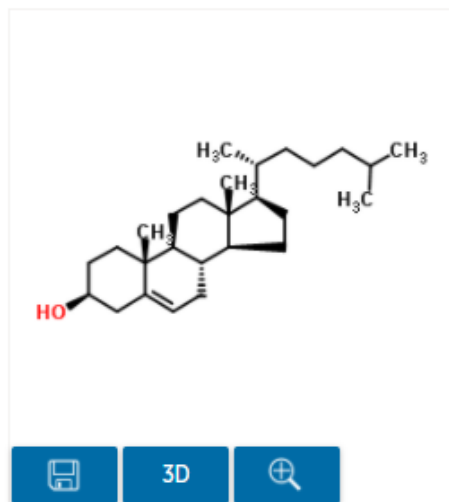
[Simple](#) [Structure](#) [Advanced](#) [History](#)

Found 1 result

Search term: **cholesterol** (Found by approved synonym)



COMMENT ON
THIS RECORD



Cholesterol

Molecular Formula $C_{27}H_{46}O$

Average mass 386.654 Da

Monoisotopic mass 386.354858 Da

ChemSpider ID 5775

- 8 of 8 defined stereocentres

Featured data source



The Merck
Index *Online*
has more data on
this compound

Names and identifiers

Properties

Searches

Spectra

Vendors

Articles

More ▾

✓ CNMR

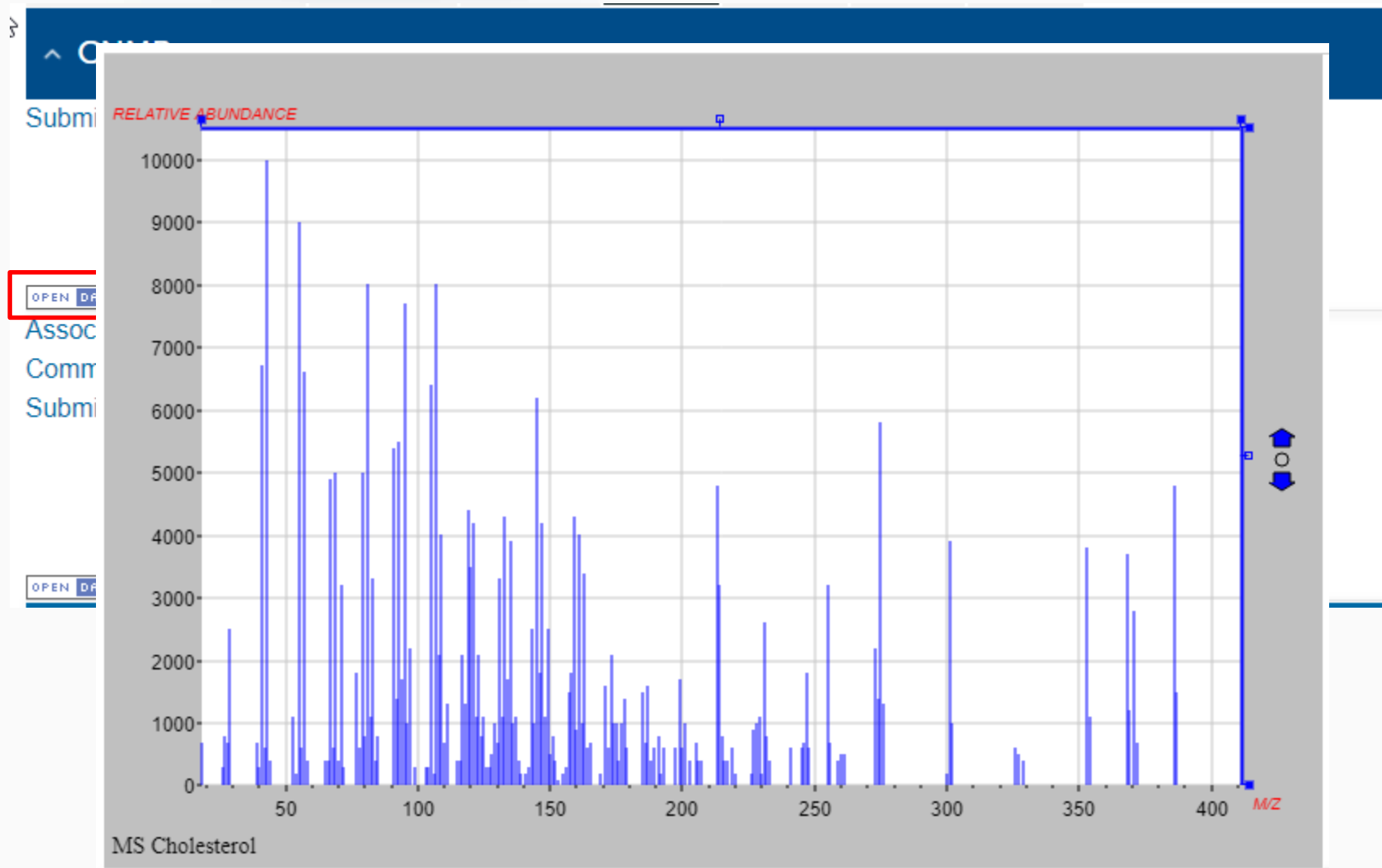
✓ Electron Impact

✓ HNMR

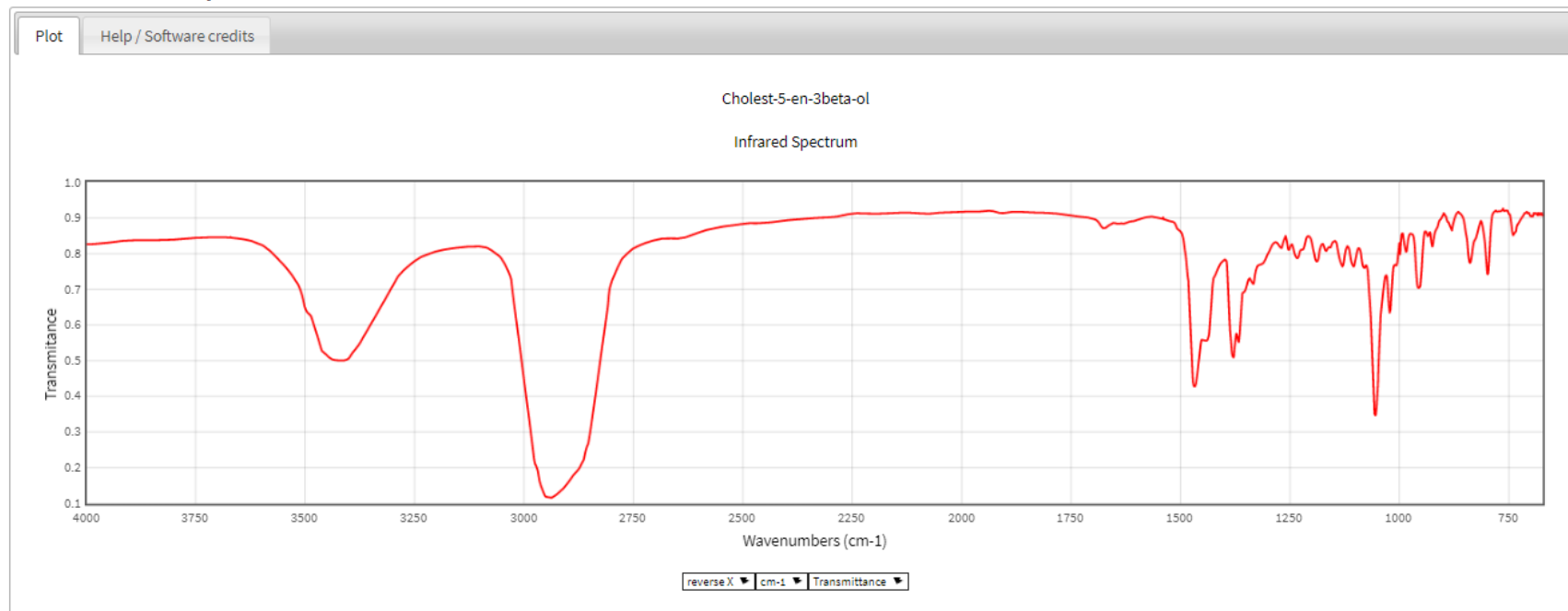
✓ Infrared

✓ Raman

✓ UV-Vis



Condensed Phase Spectrum

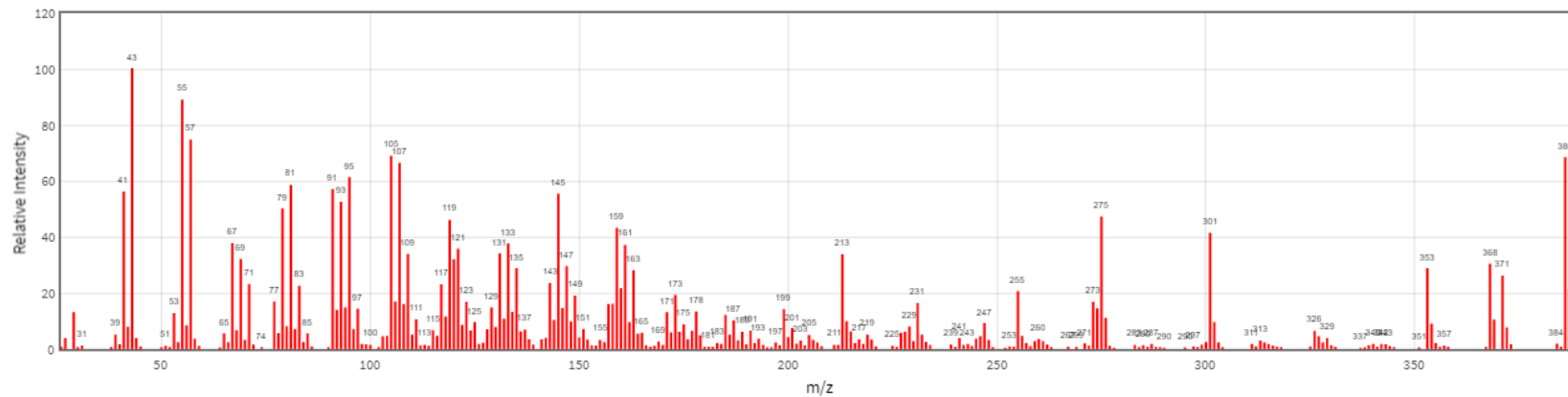


Spectrum

Plot Help / Software credits

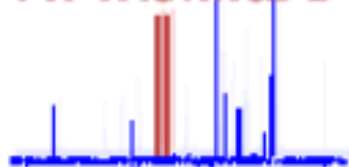
Cholesterol

Mass Spectrum



- Focused databases
 - Compiled focused databases of Open Data are preferable
 - Spectral data for structure elucidation – Open Mass Spec Data especially useful (Emma's talk!)
 - Data can be brought in-house and integrated
 - Algorithms can be derived – e.g. NMR shift prediction

NMRShiftDB



Current usage is:

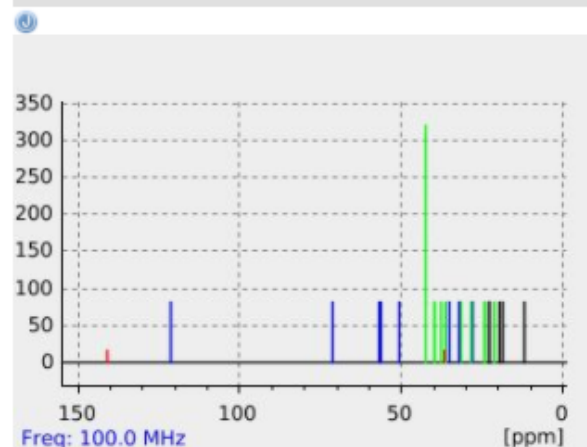
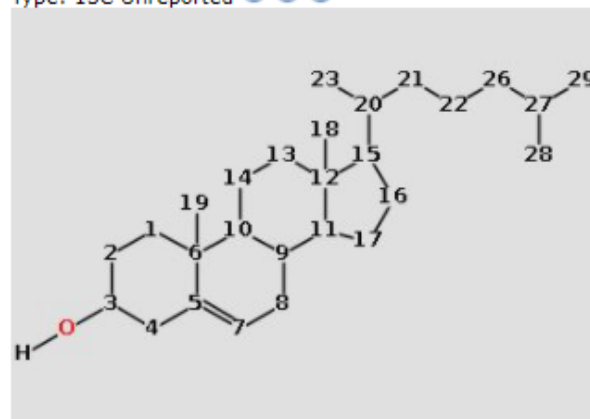
Registered Users: 1216

Structures: 43445

Spectra: Measured 52114, calculated 549

[Spectral Data](#) [Additional Data](#) [Download](#)

Type: ¹³C Unreported



Atom No. #	Mult.(coupling const.)	Meas. Shift
1	T	37.5
2	T	31.6
3	D	71.3
4	T	42.4
5	S	141.2
6	S	36.5
7	D	121.3
8	T	32.0
9	D	32.3
10	D	50.5
11	D	56.9
12	T	42.4
13	T	40.0
14	T	21.2
15	D	56.5
16	T	28.3
17	T	24.3
18	Q	12.0
19	Q	19.4
20	D	35.4
21	T	36.4
22	T	24.1
23	Q	18.8
26	T	39.6
27	D	28.0
28	Q	22.8
29	Q	22.5

Spectral Data **Additional Data** **Download**

Molecule	10016303
Chemical name(s)	
Chemical formula	C ₂₇ H ₄₆ O
Molecular weight	386.653
Number of double bond equivalents (DBEs)	5.0
Number of all rings, size of smallest set of smallest rings	10, 4
Canonical name(s)	<ul style="list-style-type: none">• OC1CC2=CCC3C(CCC4(C)C(CCC34)C(... (truncated) (SMILES)• cholest-5-en-3-ol (IUPAC from ACD/Name)• cholest-5-en-3-ol (Index from ACD/Name)• InChI=1S/C27H46O/c1-18(2)7-6-8... (truncated) (InChI)• HVYWMOMLDIMFJA-UHFFFAOYSA-N (InChI Key)• InChI=1/C27H46O/c1-18(2)7-6-8-... (truncated) (InChI with fixed H layer)
CAS-Number	
Additional information	Deposition in PubChem; Links from Unichem: ChEMBL ; eMolecules ; eMolecules ; IBM strategic IP insight platform and the National Institutes of Health ; SureChEMBL ; PubChem Compounds ; McuLe ; ACToR ; MolPort
Molecule keywords	
13C Spectrum	10038299 Rating: 10
Type	13C
Measurement conditions	
Temperature [K]	Unreported
Solvent	Unreported
Field Strength [MHz]	Unreported
Assignment Method	Unreported
Literature	S. Berger; S. Braun; H.-O. Kalinowski: 13-C-NMR-Spektroskopie, New York: Thieme Verlag 1984.
Additional comments	610.MOL; Multiplicities generated automatically from H count
Additional information	
Spectrum categories	

It is the policy of nmrshiftdb2 to have its content freely available. This means that spectra in nmrshiftdb2 can be accessed and used without restrictions. Access to single items is possible via the software and we also make available the bunch data (without the personal data). Note that, according to our opinion, NMR spectra do not represent "literary and artistic works" as mentioned in Art. 2 (1) of the Berne Convention, the author of a spectrum is, so to say, nature, and that therefore there is no copyright in NMR data. nmrshiftdb2 as a database is a protected work in the sense of article 5 of the WIPO Copyright Convention, authored by its contributors and reviewers. It is available under the [**nmrshiftdb2 Data License**](#).

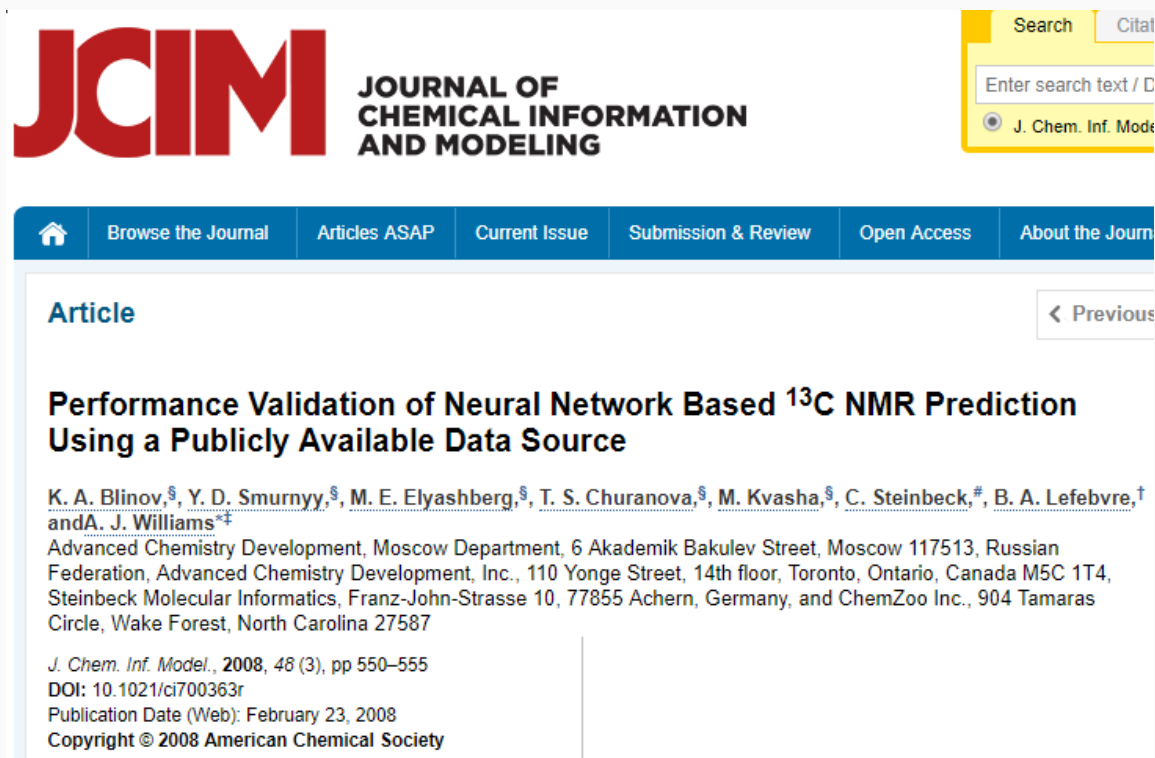
The nmrshiftdb2 data are available as:

Table 1. Data downloads

-	With 2D coordinates
sdf	<u>nmrshiftdb2.sd</u>
sdf with spectra	<u>nmrshiftdb2withsignals.sd</u>
cml	<u>nmrshiftdb2.xml</u>

Questions about nmrshiftdb2 should be directed to the [**administrator**](#).

- Open Databases offer more value
 - Bring the data in-house, integrate, link
 - Ingest and train algorithms



The screenshot shows the J. Chem. Inf. Model. journal article page. The header features the JCIM logo and the journal title. A search bar is located in the top right corner. Below the header is a navigation bar with links to 'Browse the Journal', 'Articles ASAP', 'Current Issue', 'Submission & Review', 'Open Access', and 'About the Journal'. The main content area displays the article title 'Performance Validation of Neural Network Based ¹³C NMR Prediction Using a Publicly Available Data Source' and the authors 'K. A. Blinov, Y. D. Smurnyy, M. E. Elyashberg, T. S. Churanova, M. Kvasha, C. Steinbeck, B. A. Lefebvre, and A. J. Williams'. The article is published in J. Chem. Inf. Model., 2008, 48 (3), pp 550–555. The DOI is 10.1021/ci700363r. The publication date is February 23, 2008. The copyright is © 2008 American Chemical Society.

JCIM JOURNAL OF
CHEMICAL INFORMATION
AND MODELING

Search Cital
Enter search text / D
J. Chem. Inf. Mode

Home Browse the Journal Articles ASAP Current Issue Submission & Review Open Access About the Journal

Article < Previous

Performance Validation of Neural Network Based ¹³C NMR Prediction Using a Publicly Available Data Source

K. A. Blinov,[§] Y. D. Smurnyy,[§] M. E. Elyashberg,[§] T. S. Churanova,[§] M. Kvasha,[§] C. Steinbeck,[#] B. A. Lefebvre,[†] and A. J. Williams^{*‡}

Advanced Chemistry Development, Moscow Department, 6 Akademik Bakulev Street, Moscow 117513, Russian Federation, Advanced Chemistry Development, Inc., 110 Yonge Street, 14th floor, Toronto, Ontario, Canada M5C 1T4, Steinbeck Molecular Informatics, Franz-John-Strasse 10, 77855 Achern, Germany, and ChemZoo Inc., 904 Tamaras Circle, Wake Forest, North Carolina 27587

J. Chem. Inf. Model., 2008, 48 (3), pp 550–555
DOI: 10.1021/ci700363r
Publication Date (Web): February 23, 2008
Copyright © 2008 American Chemical Society

Welcome to the CSEARCH / NMRPREDICT-Server

I have ...	I want ...	The solution to your problem
A structure Proposal	A predicted ^{13}C -NMR Spectrum	Use the CSEARCH-Robot-Referee don't enter any lines
A Structure Proposal and a ^{13}C -NMR Peaklist	A Structure Verification	Use the CSEARCH-Robot-Referee Enter as many lines as necessary Optionally assign as many lines as possible
A Structure Proposal and a ^{13}C -NMR Peaklist	A Structure Verification	Use the Wiley Data Checker based on CSEARCH technology Enter your structure and the peaklist and all will be done for you
A ^{13}C -NMR Peaklist	A Structure Proposal	Use the Spectral Similarity Search You can use this feature directly from Bruker's TOPSPIN
A ^{13}C -NMR Peaklist and hundreds/thousands of Structure Proposals	A Hitlist sorted by Similarity	Use the CSEARCH-Decision Engine



High Resolution Mass Spectral Database

Quick Search Results

[Home](#) | [Quick Search](#) | [Peak Search](#) | [Record Index](#) | [Statistics](#) | [Imprint/Data privacy](#) | MassBank ID:

Search Parameters :

Compound Name: **cholesterol**

Instrument Type: **CE-ESI-TOF ,
ESI-ITTOF ,
LC-ESI-IT ,
LC-ESI-Q ,
LC-ESI-QQ ,
UPLC-ESI-QTOF**

**ESI-FTICR ,
ESI-QTOF ,
LC-ESI-ITFT ,
LC-ESI-QFT ,
LC-ESI-QTOF ,**

**ESI-ITFT
HPLC-ESI-TOF
LC-ESI-ITTOF
LC-ESI-QIT
LC-ESI-TOF**

MS Type: **All**


Ion Mode: **Positive**

[Edit / Resubmit Query](#)

Results : 5 Hit. (1 - 5 Displayed)

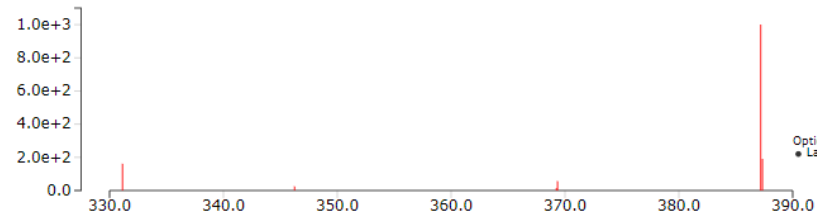
[First](#) [Prev](#) **1** [Next](#) [Last](#) (Total **1** Page)

[▼ Results End](#)

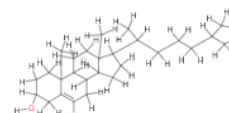
<input type="checkbox"/>	Name	Formula / Structure	ExactMass	ID
<input type="checkbox"/>	<input checked="" type="checkbox"/> Cholesterol	C27H46O	386.35486	
	5 spectra			
<input type="checkbox"/>	LC-ESI-QTOF, MS2, CE:10 eV, [M+H]⁺			CO000101
<input type="checkbox"/>	LC-ESI-QTOF, MS2, CE:20 eV, [M+H]⁺			CO000102
<input type="checkbox"/>	LC-ESI-QTOF, MS2, CE:30 eV, [M+H]⁺			CO000103
<input type="checkbox"/>	LC-ESI-QTOF, MS2, CE:40 eV, [M+H]⁺			CO000104
<input type="checkbox"/>	LC-ESI-QTOF, MS2, CE:50 eV, [M+H]⁺			CO000105


[Home](#) | [Quick Search](#) | [Peak Search](#) | [Record Index](#) | [Statistics](#) | [Imprint/Data privacy](#) MassBank ID:

Mass Spectrum



Chemical Structure

[illegible]



Advanced Mass Spectral Database

7,547 (+76)
compounds

11,684 (+79)
trees

2,661,175 (+12,168)
spectra

Cholest-4-en-3-one

[enlarge](#)

Systematic / IUPAC Name: (8S,9S,10R,13R,14S,17R)-10,13-dimethyl-17-[(2R)-6-methylheptan-2-yl]-1,2,6,7,8,9,11,12,14,15,16,17-dodecahydrocyclopenta[a]phenanthren-3-one

ID: Reference231

Other Names: 4-Cholesten-3-one;
Cholestenone;
3-Oxo-4-cholestene;
3-Keto-4-cholestene;
(+)-4-Cholesten-3-one ; [more](#)

Formula: C₂₇H₄₄O

Class: Endogenous Metabolites

Spectral Data

Cholest-4-en-3-one mass spectral data can be found in a separate interface. The data are manually curated and of the highest quality.

[View Mass Spectra](#)


Available MS Data

Used Instruments	Orbitrap Elite
No. of Spectral Trees	1
No. of Spectra	107
Tandem Spectra	MS ¹ , MS ² , MS ³
Ionization Methods	ESI
Analyzers	FT

- Integration:
 - Use simple URL linking for navigation
 - Provide simple services for real time prediction

Example Integrations

CompTox Chemistry Dashboard

 United States
Environmental Protection
Agency

HomeAdvanced SearchBatch SearchListsPredictionsDownloads

Chemistry Dashboard

761 Thousand Chemicals

☐ Identifier substring search

[See what people are saying, read the dashboard comments!](#)


Latest News

[Read more news](#)

A Major Update to the Dashboard Releases on March 7th 2018

March 7th, 2018 at 12:00:48 PM

◀ A major update to the dashboard has been released prior to the Society of Toxicology and American Chemical Society Spring meetings. This release brings together six months of effort in adding and curating data, major updates to the batch searching functionality and access to real time predictions for both physiochemical and toxicity endpoints. A [list of release notes](#) is available for your review. We look forward to your feedback. ▶



Discover.
[About/Disclaimer](#)

Connect.
[ACToR](#)

Ask.
[Contact](#)

Link-Based Access

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

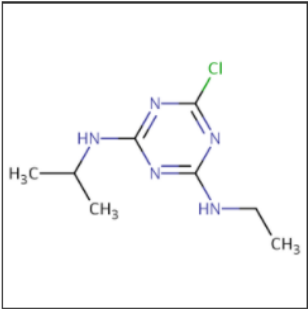
Chemistry Dashboard

Submit Comment Share Copy Aa Aa Aa

Atrazine

1912-24-9 | DTXSID9020112

Ⓢ Searched by Approved Name: Found 1 result for 'atrazine'.



Wikipedia

Atrazine is an herbicide of the triazine class. Atrazine is used to prevent pre- and postemergence broadleaf weeds in crops such as maize (corn) and sugarcane and on turf, such as golf courses and residential lawns. It is one of the most widely used herbicides in US and Australian agriculture. It was banned in the European Union in October 2003, when the EU found groundwater levels exceeding the limits set by regulators, and Syngenta could neither show that this could be prevented nor that...[Read more](#)

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

Executive Summary (Beta)

Chemical Properties

Env. Fate/Transport

Hazard

ADME (Beta)

Exposure

Bioassays

Similar Compounds

Related Substances

Synonyms

Literature

Links

Comments

General

Toxicology

Publications

Analytical

Predictions

Link Access

Executive Summary (Beta) | Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exp | **Links** | Comments

General

- EPA Substance Registry Ser...
- Household Products Database
- PubChem
- Chempider
- CPCat
- Wikipedia
- MSDS Lookup
- ChEMBL
- Chemical Vendors
- CalEPA Office of Environmen...
- NIOSH Chemical Safety Cards
- ToxPlanet
- Wikidata
- ChemHat: Hazards and Alter...
- Wolfram Alpha

Toxicology

- ACToR
- DrugPortal
- CCRIS
- ChemView
- CTD
- eChemPortal
- Gene-Tox
- HSDB
- ToxCast Dashboard 2
- LactMed
- International Toxicity Estim...
- ACToR PDF Report
- CREST

Publications

- Toxline
- Environmental Health Perspe
- NIEHS
- National Toxicology Program
- Google Books
- Google Scholar
- Google Patents
- PubMed
- RSC Publications
- BioCaddie DataMed
- Springer Materials
- Federal Register
- Regulations.gov
- Bielefeld Academic Search E...
- CORE Literature Search

Prediction

- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction

ABOUT

PREDICT 1H NMR

PREDICT 13C NMR

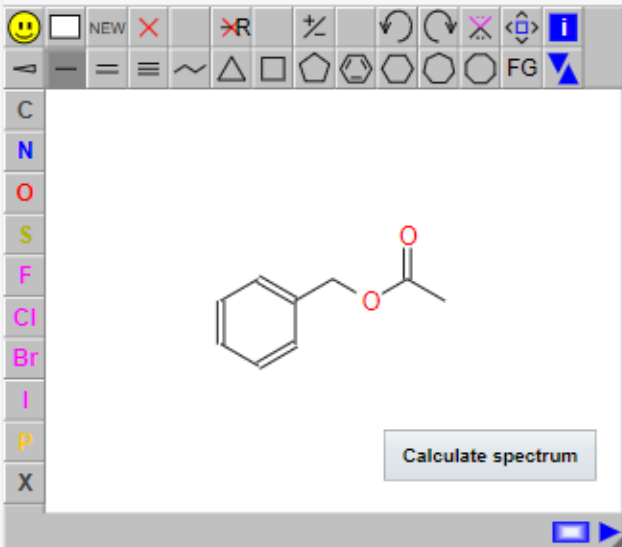
PREDICT 2D

Molfile or SMILES

Paste or drop a molfile or SMILES

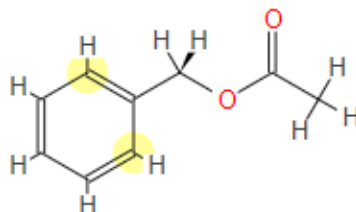
Help

Draw a chemical structure to predict



Draw a chemical structure and click on "Calculate spectrum".
You may also DRAG / DROP a molfile !
You will get an interactive NMR spectrum.

Chemical structure with hydrogen exploded



δ (ppm)

128.9

20.8

127.9

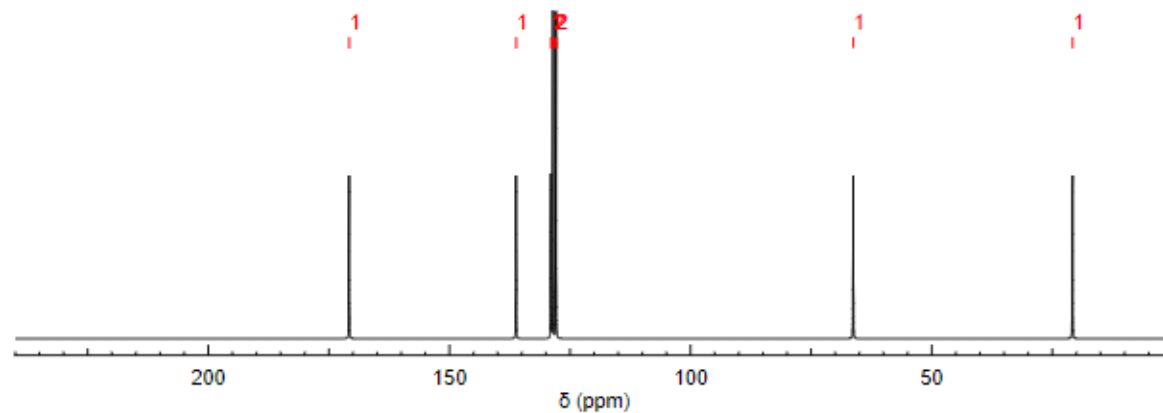
127.9

66.2

128.5

Drop or paste a JCamp file

13C NMR spectra



Open Data For Bulk Predictions

**CompTox
Mobile**

View in iTunes

This app is designed for both iPhone and iPad

Free

Category: Productivity
Released: Jan 16, 2017
Version: 1.0
Size: 267 MB
Language: English
Seller: Kirill Blinov
© 2017 Molecule Apps, 2017 EPA
Rated 4+

Compatibility: Requires iOS 6.0 or later. Compatible with iPhone, iPad, and iPod touch.

Customer Ratings

We have not received enough ratings to display an average for the current version of this application.

More by Kirill Blinov

NMR

View in Mac App Store

Description

Find chemical structure instantly by exact mass (m/z), ¹³C NMR chemical shifts, structure name or CAS Registry Number in a database of about 720,000 EPA CompTox structures.

[Kirill Blinov Web Site](#) [CompTox Mobile Support](#) [...More](#)

Screenshots

iPhone | iPad

Carrier 2:30 PM

+ m/z Name ¹³C 17

adenine

73-24-5 Adenine 1H-Purin-6-amine <chem>Nc1ncnc2[nH]cnc12</chem> 135.05450	73663-94-2 Adenine, dihydriodide <chem>Nc1ncnc2[nH]cnc12.[I-]</chem> 390.87908
52175-10-7 Adenine phosphate EINECS 257-702-7 <chem>Nc1ncnc2[nH]cnc12OP(=O)(O)O</chem> 233.03139	5142-22-3 1-Methyladenine Adenine, 1-methyl- <chem>Cn1cnc2[nH]cnc12</chem> 149.07015
19152-67-1 Adenine, 2-chloro-N-hydroxy-2-chloro-N-hydroxyadenine <chem>ClC1=NC(=O)NC(=O)N1</chem> 233.03139	62700-65-6 Purine, 6-amino-2-methylthio-, hydrochloride Adenine, 2-methylthio-, hydrochloride <chem>CSc1ncnc2[nH]cnc12</chem> 233.03139

Carrier 2:31 PM

+ m/z Name ¹³C

m/z 135.0545

¹³C Example: 25 32.4 115 (ppm)

73-24-5 Adenine 1H-Purin-6-amine <chem>Nc1ncnc2[nH]cnc12</chem> 135.05450 Δ 0.00000	2380-6 Adenine <chem>Nc1ncnc2[nH]cnc12</chem> 135.05450
13877-56-0 7-Aminopurazole(4,3-dipyridine) 7-Aminopurazole(4,3-dipyridine) <chem>Nc1nc2c(ncn2)ccn1</chem> 160568	160568 Adenine <chem>Nc1ncnc2[nH]cnc12</chem> 135.05450
135.05450 Δ 0.00000 2-Aminopurine EINECS 207-197-4 <chem>Nc1ncnc2[nH]cnc12</chem> 452-06-2	5019-4 Adenine <chem>Nc1ncnc2[nH]cnc12</chem> 135.05450

- Open Data for apps
 - Structures
 - CAS Registry Numbers
 - Names
 - Formulae
 - Mass
- iOS app including predicted C13 NMR

Mass Searching

AT&T 9:02 AM 30%

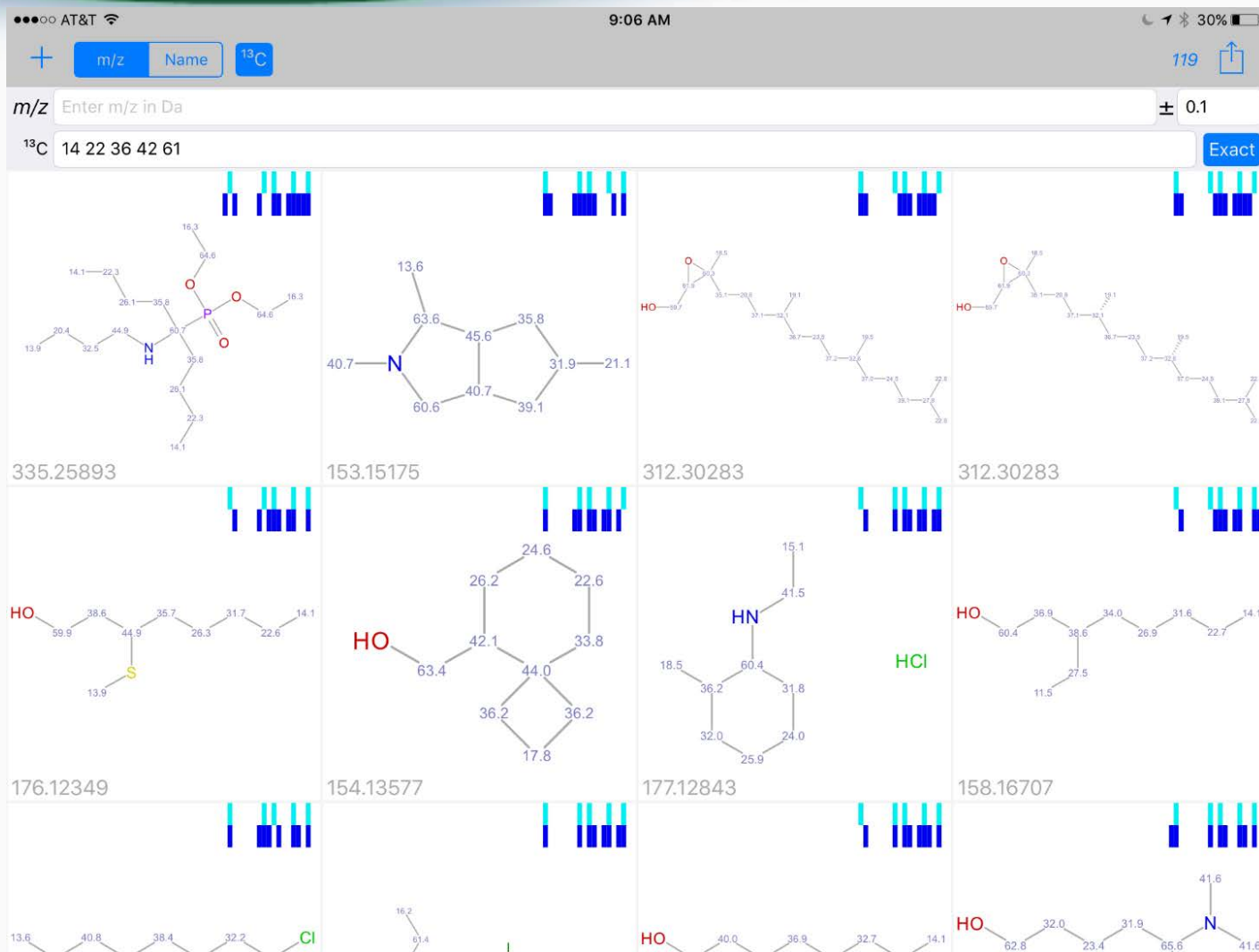
+ m/z Name ¹³C

m/z 300.156 ± 0.005

¹³C Example: 25 32.4 115 (ppm) Exact

<p>42978-66-5</p> <p>Tripolyene glycol diacrylate 2-Propenoic acid, (1-methyl-1,2-ethanediyl)bis[oxy(methyl-2,1-ethanediyl)] ester</p> <p>300.15729 Δ -0.00129</p>	<p>71348-43-1</p> <p>Dioxaspiro[5.5]undecane-3,3-dicarboxylic acid, diethyl ester</p> <p>300.15729 Δ -0.00129</p>	<p>76319-15-8</p> <p>Cyclopenta(c)pyran-4,7-dimethanol, 1,4a,5,6,7,7a-hexahydro-1,6-dihydroxy-, 1-isovalerate</p> <p>300.15729 Δ -0.00129</p>	<p>90694-04-5</p> <p>2,2,5,5,8,8-hexamethylhexahydrobenzo[1,2-d:3,4-d':5,6-d'']tris[1,3]dioxole</p> <p>300.15729 Δ -0.00129</p>
<p>79410-21-2</p> <p>AG-H-18608</p> <p>300.15729 Δ -0.00129</p>	<p>62180-73-8</p> <p>2-[2,2-dimethyl-3-(2-prop-2-enyloxyethoxy)propoxy]ethyl Prop-2-enoate</p> <p>300.15729 Δ -0.00129</p>	<p>152065-61-7</p> <p>Ethanol, 2,2'-[1,3,5-benzenetriyltris(methyleneoxy)] tris-</p> <p>300.15729 Δ -0.00129</p>	<p>77393-39-6</p> <p>Trioxacyclododecane-2,6,10-trione, 3,3,7,7,11,11-hexamethyl-</p> <p>300.15729 Δ -0.00129</p>
<p>56275-48-0</p> <p>1-Ethoxy-1-oxonon-2-en-2-yl ethyl ethanedioate</p> <p>300.15729 Δ -0.00129</p>	<p>91743-83-8</p> <p>Diethyl 3,9-dioxoundecanedioate</p> <p>300.15729 Δ -0.00129</p>	<p>112079-61-5</p> <p>4,4',4''-(1,3,5-Trioxane-2,4,6-triyl)tributanol</p> <p>300.15729 Δ -0.00129</p>	<p>63530-28-9</p> <p>Dimethyl 2-(1,2-dihydroxycyclohexyl)cyclopentane-1,1-dicarboxylate</p> <p>300.15729 Δ -0.00129</p>

Mass and CNMR Searching



- Structures – Molfile, SDF file, InChIs (standard and non-standard)
- NMR – JCAMP and all its variants
- MS – mzML, MSP (and all its variants), MassBank



NMReDATA, a standard to report the NMR assignment and parameters of organic compounds

Marion Pupier^a, Jean-Marc Nuzillard^b, Julien Wist^c, Nils E. Schlörer^d, Stefan Kuhn^d, Mate Erdelyi^e, Christoph Steinbeck^f, Antony J. Williams^g, Craig Butts^h, Tim D. W. Claridgeⁱ, Bozhana Mikhova^j, Wolfgang Robien^k, Hesam Dashti^l, Hamid R. Eghbalnia^l, Christophe Farès^m, Kessler Pavelⁿ, Fabrice Moriaud^o, Mikhail Elyashberg^p, Dimitris Argyropoulos^q, Manuel Pérez^r, Patrick Giraudeau^{s,t}, Roberto R. Gil^u, Paul Trevorrow^v, Damien Jeannerat^{a,}*

- The abundance of online data continues to grow
- There are “integrated data”, there are databases, there are online tools, there are mobile apps
- Data Quality is critical and OPENness is enabling
 - Open Data
 - Open Standards
 - Open Source
- The rest of the day will expand on these efforts...

Antony Williams

US EPA Office of Research and Development

National Center for Computational Toxicology (NCCT)

Williams.Antony@epa.gov

ORCID: <https://orcid.org/0000-0002-2668-4821>