

Benefits of making data from the EPA National Center for Computational Toxicology available for reuse

*Antony Williams¹, Kamel Mansouri², Valery Tkachenko³,
Kirill Blinov⁴ and Chris Grulke¹*

1. National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC
2. Scitovation, RTP, NC; 3. Science Data Software, LLC, MD, USA; 4. Molecule Apps LLC, 97330, Corvallis, OR


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

*August 2017
ACS Fall Meeting, Washington, DC*

- A **publicly accessible website** delivering access:
 - ~760,000 chemicals with related property data
 - Experimental and predicted physicochemical property data
 - Integration to “biological assay data” for 1000s of chemicals
 - Information regarding consumer products containing chemicals
 - Links to other agency websites and public data resources
 - “Literature” searches for chemicals using public resources
 - “Batch searching” for thousands of chemicals
 - **DOWNLOADABLE Open Data for reuse and repurposing**

Comptox Chemistry Dashboard


<https://comptox.epa.gov>

 EPA
United States
Environmental Protection
Agency

Home Advanced Search

Chemistry Dashboard

Aa Aa Aa



**~760,000 chemicals
>15 years of data**

Chemistry Dashboard

☐ Single component search ☐ Ignore isotopes



See what people are saying, read the dashboard comments!

Need more? Use advanced search.

741 Thousand Chemicals

Latest News

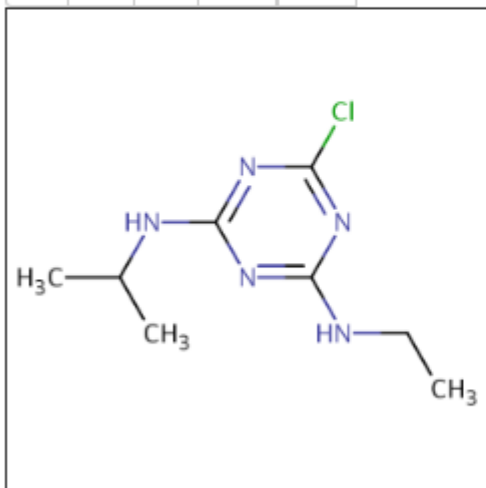
■ New CompTox Mobile app releases >700,000 chemicals on iPhone and iPad
January 20, 2017 at 8:40:18 AM

About Contact Privacy   Accessibility Help Downloads

Atrazine

1912-24-9 | DTXSID9020112 ⓘ

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




Wikipedia

Intrinsic Properties

Molecular Formula: C₈H₁₄ClN₅

Average Mass: 215.69 g/mol

Monoisotopic Mass: 215.093773 g/mol

 Find All Chemicals



Structural Identifiers

Record Information

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

Data Distribution

- Consuming and producing open data

Summary

LogP: Octanol-Water

Water Solubility

Density

Melting Point

Boiling Point

Surface Tension

Vapor Pressure

LogKoa: Octanol-Air

Henry's Law

Index of Refraction

Molar Refractivity

pKa Acidic Apparent

Download as: TSV Excel SDF

Property

LogP: Octanol-Water

Water Solubility

Density

Melting Point

Boiling Point

Surface Tension

Vapor Pressure

LogKoa: Octanol-Air

Henry's Law

Index of Refraction

Molar Refractivity

pKa Acidic Apparent

Molar Volume

Polarizability

☒ Select/Deselect All
☒ LogP: Octanol-Water
☒ Water Solubility
☒ Density
☒ Melting Point
☒ Boiling Point
☒ Surface Tension
☒ Vapor Pressure
☒ LogKoa: Octanol-Air
☒ Henry's Law
☒ Index of Refraction
☒ Molar Refractivity
☒ pKa Acidic Apparent
☒ Molar Volume
☒ Polarizability

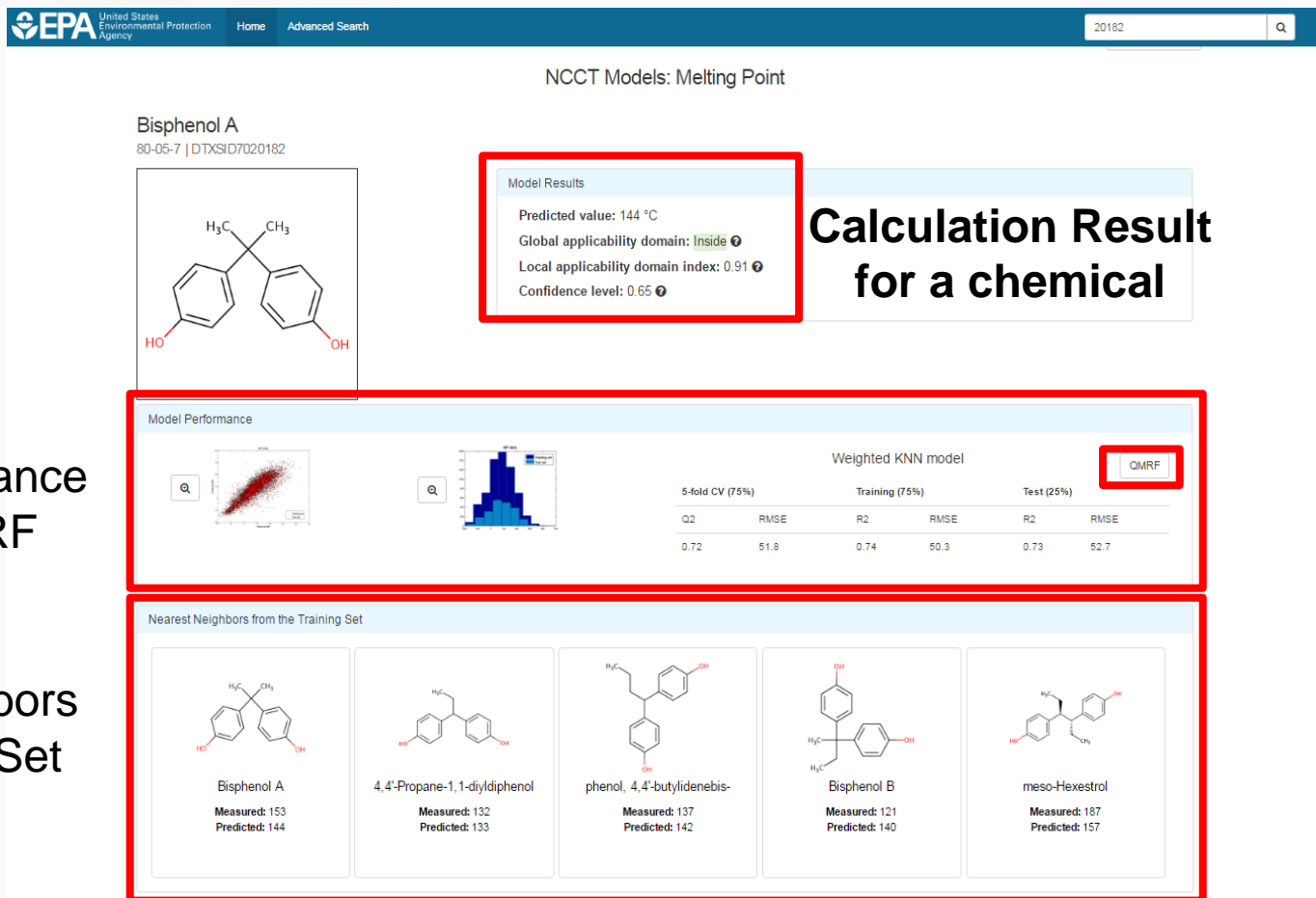
Download

	Median		Range		Unit
	Experimental	Predicted	Experimental	Predicted	
LogP: Octanol-Water	3.32	3.24	3.32	2.40 to 3.73	-
Water Solubility	5.26e-04	1.58e-03	5.26e-04	5.70e-04 to 3.68e-03	mol/L
Density		1.14	-	-	g/cm ³
Melting Point	56	144	153 to 158	132 to 157	°C
Boiling Point	100	349	200	334 to 364	°C
Surface Tension		46.0	-	-	dyn/cm
Vapor Pressure		2.52e-07	-	7.01e-08 to 5.34e-07	mmHg
LogKoa: Octanol-Air		8.38	-	-	-
Henry's Law		6.96e-07	-	-	atm-m ³ /mole
Index of Refraction		1.60	-	-	-
Molar Refractivity		68.2	-	-	cm ³
pKa Acidic Apparent	-	10.3 (1)	-	10.3	-
Molar Volume	-	200 (1)	-	200	cm ³
Polarizability	-	27.0 (1)	-	27.0	Å ³

Data Distribution

	A	B	C	D	E	F	G	H
1	Property	Average		Median		Range		Unit
2		Experimental	Predicted	Experimental	Predicted	Experimental	Predicted	
3	LogP: Octanol-Water	3.32 (1)	3.24 (4)	3.32	3.24	3.32	2.40 to 3.73	-
4	Water Solubility	5.26e-04 (1)	1.58e-03 (4)	0.000526	0.00158	0.000526	5.70e-04 to 3.68e-03	mol/L
5	Density	-	1.14 (1)	-	1.14	-	-	g/cm ³
6	Melting Point	155 (7)	144 (3)	156	144	153 to 158	132 to 157	°C
7	Boiling Point	200 (1)	349 (3)	200	349	200	334 to 364	°C
8	Surface Tension	-	46.0 (1)	-	46	-	-	dyn/cm
9	Vapor Pressure	-	2.52e-07 (3)	-	0.000000252	-	7.01e-08 to 5.34e-07	mmHg
10	LogKoa: Octanol-Air	-	8.38 (1)	-	8.38	-	-	-
11	Henry's Law	-	6.96e-07 (1)	-	0.000000696	-	-	atm-m ³ /mole
12	Index of Refraction	-	1.60 (1)	-	1.6	-	-	-
13	Molar Refractivity	-	68.2 (1)	-	68.2	-	-	cm ³
14	pKa Acidic Apparent	-	10.3 (1)	-	10.3	-	-	-
15	Molar Volume	-	200 (1)	-	200	-	-	cm ³
16	Polarizability	-	27.0 (1)	-	27	-	-	Å ³
17								

Modeling Details



Model Performance
with full QMRF

Nearest Neighbors
from Training Set

- Our approach to modeling:
 - Obtain **high quality** training sets
 - Apply appropriate modeling approaches
 - **Validate** performance of models
 - Define the applicability domain and model limitations
 - Use models to predict properties across our full datasets
 - Release as **Open Data and Open Models**

Consuming and Curating Public Data

Public data should be curated prior to modeling

Mol Block	CAS	NAME	Smiles
	000056-93-9	BENZYL TRIMETHYL AMMONIUM CHLORIDE	<chem>CN(C)(C)Cc1ccccc1.[Cl-]</chem>
	000068-05-3	TETRAETHYL AMMONIUM IODIDE	<chem>CC[N+](CC)(CC)CC.[I-]</chem>
	000071-91-0	TETRAETHYL AMMONIUM BROMIDE	<chem>CC[N+](CC)(CC)CC.[Br-]</chem>

Covalent Halogens

Mol Block	CAS	NAME	Smiles
	000076-43-7	FLUCYMESTERONE	<chem>CC12CCC3C(C1CC2=O)C(=O)CC4=C3C(=C(C=C4)F)C</chem>
	000077-99-6	1,1,1-TRIS(4-HYDROXYETHYL)PROPANE	<chem>CC(O)CC(C(CO)CC)(CO)CC</chem>
	000079-60-7	CORTISONE-4a-FLUORO	<chem>CC12CCC3C(C1CC2=O)C(=O)CC4=C3C(=C(C=C4)F)C</chem>
	000082-38-2	DISPERSE RED 9	<chem>CC1=CC=C(C=C1)C(C2=CC=CC=C2)(C3=CC=CC=C3)C4=CC=CC=C4N</chem>

Structure	Formula	FW	CAS	NAME	MP	ExpMP	ErrorMP
	C ₃ H ₅ O ₃	90.0779	000050-21-5	LACTIC ACID	1.6000000000000000e+001	2.2600000000000000e+001	5.8000000000000000e+000
	C ₃ H ₅ O ₃	90.0779	000079-33-4	L-LACTIC ACID	5.5000000000000000e+001	2.2600000000000000e+001	-3.0340000000000000e+001
	C ₃ H ₅ O ₃	90.0779	000590-02-3	4-HYDROXYPROPIONIC ACID	1.6000000000000000e+001	2.2600000000000000e+001	4.6000000000000000e+000
	C ₃ H ₅ O ₃	90.0779	010328-41-7	D-LACTIC ACID	5.5000000000000000e+001	2.2600000000000000e+001	-3.0140000000000000e+001

Identical Chemicals

Mismatches

- CAS Checksum: 12163 valid, 3646 **invalid** (>23%)
- **Invalid** names: 555
- **Invalid** SMILES 133
- Valence **errors**: 322 Molfile, 3782 SMILES (>24%)
- Duplicates check:
 - 31 DUPLICATE MOLFILES
 - 626 DUPLICATE SMILES
 - 531 DUPLICATE NAMES
- SMILES vs. Molfiles (structure check)
 - 1279 differ in stereochemistry (~8%)
 - 362 “Covalent Halogens”
 - 191 differ as tautomers
 - 436 are different compounds (~3%)



Journal

SAR and QSAR in Environmental Research >

Volume 27, 2016 - Issue 11: 17th International Conference on QSAR in Environmental and Health Sciences (QSAR 2016) - Part II. Guest Editors: C.G. Barber and G.J. Myatt

Enter keywords, authors, DOI etc.

258

Views

4

CrossRef citations

16

Altmetric

Articles

An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling\$

K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson & A. J. Williams ✉

Pages 911-937 | Received 03 Sep 2016, Accepted 24 Oct 2016, Published online: 25 Nov 2016

Download citation

<http://dx.doi.org/10.1080/1062936X.2016.1253611>



OPERA Models: <https://github.com/kmansouri/OPERA>

Downloadable Data File

<https://comptox.epa.gov/dashboard/downloads>

Downloads

[DSSTox Identifier to PubChem Identifier Mapping File](#)

Posted: 11/14/2016

The DSSTox to PubChem Identifiers mapping file is in TXT format and includes the PubChem SID, PubChem CID and DSSTox substance identifier (DTXSID).
[image](#)

[DSSTox identifiers mapped to CAS Numbers and Names File](#)

Posted: 11/14/2016

The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name.
[image](#)

[DSSTox Mapping File](#)

Posted: 12/14/2016

The DSSTox mapping file contains mappings between the DSSTox substance identifier (DTXSID) and the associated InChI String and InChI Key. The file is made available as a Tab Separated Value (TSV) file with each entry represented as shown:

DTXSID7020001 InChI=1S/C11H9N3/c12-10-6-5-8-7-3-1-2-4-9(7)13-11(8)14-10/h1-6H,(H3,12,13,14) FJTNLJLPLJDTRM-UHFFFAOYSA-N

[DSSTox Predicted Property Data](#)

Posted: 12/14/2016





A number of property prediction models were developed using curated data as described in the publication "[An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling](#)". These property prediction models include logP, water solubility, bioconcentration factor and many others. The files include DTXSIDs, names and the predicted properties where possible. The models cannot predict properties for all chemicals contained in the database (for example, inorganics, organometallics and elements cannot be handled).

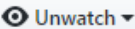

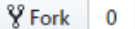
[DSSTox Synonyms File](#)


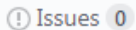

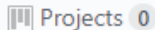


Posted: 12/14/2016

The DSSTox synonyms file is in SDF format and includes the DSSTox substance identifier (DTXSID). The preferred name, the CAS Registry Number and the list of associated synonyms for over 720,000 chemicals. In order to view an SDF file you will need to have access to the appropriate piece of software to open an SDF files. Examples include ChemAxon JChem, ACD/ChemFolder or ChemDraw.

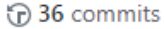
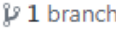



OPERA on GitHub

 This repository Search Pull requests Issues Marketplace Gist   


kmansouri / OPERA  1  1  0

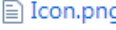
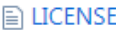

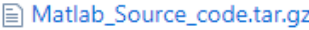



 Code  Issues 0  Pull requests 0  Projects 0  Wiki  Insights

Command line application providing QSAR models predictions as well as applicability domain and accuracy assessment for physicochemical properties and environmental fate endpoints.

 36 commits  1 branch  0 releases  1 contributor  MIT

Branch: master New pull request Create new file Upload files Find file Clone or download

 kmansouri committed on GitHub OPERA 1.2 Windows Latest commit 731deaf on May 19

	OPERA 1.2 icon	3 months ago
	Initial commit	9 months ago
	Added logo and icon	9 months ago
	OPERA 1.2 MATLAB source code	3 months ago
	OPERA 1.2 Linux	3 months ago
	OPERA 1.2 C++ Library	3 months ago
	OPERA 1.2 C Library	3 months ago

<https://github.com/kmansouri/OPERA.git>

Input:

- MATLAB .mat file, an ASCII file with only a matrix of variables
- SDF file or SMILES strings of QSAR-ready structures. In this case the program will calculate PaDEL 2D descriptors and make the predictions.
- The program will extract the molecules names from the input csv or SDF (or assign arbitrary names if not) as IDs for the predictions.



Output:


- Depending on the extension, the can be text file or csv with
 - A list of molecules IDs and predictions
 - Applicability domain
 - Accuracy of the prediction
 - Similarity index to the 5 nearest neighbors
 - The 5 nearest neighbors from the training set: Exp. value, Prediction, InChI key



Names and Identifiers

Found 100 synonyms

Legend: **Valid Synonyms** *Good Synonyms* *Other Synonyms*

 Copy all Synonyms

Atrazine

1,3,5-Triazine-2,4-diamine, 6-chloro-N-ethyl-N'-(1-methylethyl)-

1912-24-9 **Active CA 8-RN**

UNII-QJA9M5H4IM **FDA Registry Number**

1,3,5-Triazine-2,4-diamine, 6-chloro-N2-ethyl-N4-(1-methylethyl)-

1-Chloro-3-ethylamino-5-isopropylamino-2,4,6-triazine

2-Chloro-4-(ethylamino)-6-(2-propylamino)-s-triazine

2-Chloro-4-(ethylamino)-6-(isopropylamino)-s-triazine

2-Chloro-4-(ethylamino)-6-(isopropylamino)triazine

2-Chloro-4-ethylamineisopropylamine-s-triazine

2-Chloro-4-ethylamino-6-isopropylamino-1,3,5-triazine

2-Chloro-4-ethylamino-6-isopropylamino-s-triazine

2-Ethylamino-4-isopropylamino-6-chloro-s-triazine

6-Chloro-4-(ethylamino)-2-(isopropylamino)-s-triazine

6-Chloro-N-ethyl-N'-(1-methylethyl)-1,3,5-triazine-2,4-diamine

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

Downloadable Data File

<https://comptox.epa.gov/dashboard/downloads>

Downloads

[DSSTox Identifier to PubChem Identifier Mapping File](#)

Posted: 11/14/2016

The DSSTox to PubChem Identifiers mapping file is in TXT format and includes the PubChem SID, PubChem CID and DSSTox substance identifier (DTXSID).
image

[DSSTox identifiers mapped to CAS Numbers and Names File](#)

Posted: 11/14/2016

The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name.
image

[DSSTox MS Ready Mapping File](#)

Posted: 11/14/2016

The CompTox Chemistry Dashboard can be used by mass spectrometrists for the purpose of structure identification. A normal formula search would search the exact formula associated with any chemical, whether it include solvents of hydration, salts or multiple components. However, mass spectrometry detects ionized chemical structures and molecular formulae searches should be based on desalted, and desolvated structures with stereochemistry removed. We refer to these as "MS ready structures" and the MS-ready mappings are delivered as Excel Spreadsheets containing the Preferred Name, CAS-RN, DTXSID, Formula, Formula of the MS-ready structure and associated masses, SMILES and InChI Strings/Keys.
image














[DSSTox SDF File](#)

Posted: 12/14/2016














This zip file contains the entire chemical structure collection of over 700,000 chemicals from the DSSTox database contained in one large SDF file. The file contains the structure, The DSSTox Structure Identifier (DTXCID), The DSSTOX Substance Identifier (DTXSID listed as PubChem External Data Source), the associated Dashboard URL, associated synonyms and Quality Control Level details. In order to view an SDF file you will need to have access to the appropriate piece of software to open an SDF files. Examples include ChemAxon JChem, ACD/ChemFolder or ChemDraw.
image

Links to Other Resources











General

-  EPA Substance Re...
-  NIST Chemistry W...
-  Household Product...
-  PubChem
-  Chempider
-  CPCat
-  DrugBank
-  HMDB
-  Wikipedia
-  MSDS Lookup
-  ChEMBL
-  Chemical Vendors
-  Consumer Product...






Toxicology

-  ACToR
-  DrugPortal
-  CCRIS
-  ChemView
-  CTD
-  e
-  E
-  G
-  HSDB
-  ToxCast Dashboar...
-  LactMed
-  ACToR PDF Report
-  International Toxicit...

Publications

-  Toxline
-  Environmental Heal...
-  NIEHS
-  National Toxicology...
-  Google Books
-  Federal Register
-  Regulations.gov
-  Springer Materials
-  BioCaddie DataMed
-  RSC Publications

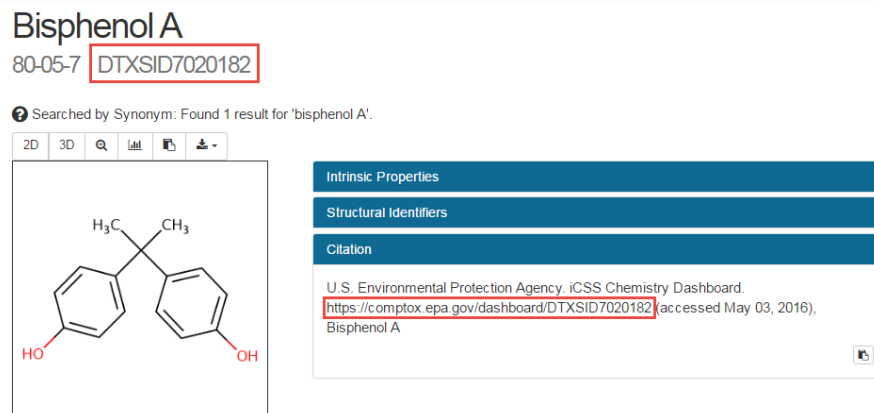
Analytical

-  National Environme...
-  MONA: MassBank ...
-  Tox21 Analytical Data
-  RSC Analytical Abs...
-  FOR-IDENT

The Office of the Federal Register (OFR) of the National Archives and Records Administration (NARA), and the U.S. Government Printing Office (GPO) jointly administer the FederalRegister.gov website.

In-Links into the Dashboard

- Linkages into the Dashboard are simple: using the associated identifiers



- For integration we can supply files of structures and identifiers mapped to DTXSIDs. Contact us...
- PubChem, EBI's UNICHEM, ChemSpider, etc.

In-linking Resources



EPA DSSTox

[PUBCHEM](#) > [DATA SOURCES](#) > [EPA DSSTOX](#)

Organization:	National Center for Computational Toxicology (NCCT) Office of Research and Development US Environmen
Category:	Research and Development, Governmental Organizations, Curation Efforts
URL:	https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database

ChemSpider

Search and share chemistry

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

Data source details

*** Name** EPA DSSTox
Primary user None
Contact name Antony Williams
Secondary users
None

Description

The foundation of chemical safety testing relies on chemistry information such as high-quality chemical structures and physicochemical properties. This information is used by scientists to predict the potential health risks of chemicals. The CompTox Dashboard is part of a suite of dashboards developed by EPA to help evaluate the safety of chemicals. It provides access to a variety of data and information on over 700,000 chemicals currently in use and of interest to environmental researchers. Within the CompTox Dashboard, users can access chemical structures, experimental and predicted physicochemical and toxicity data, and additional links to relevant websites and applications. It maps curated physicochemical property data associated with chemical substances to their corresponding chemical structures.

Data collections

None


Company name

National Center for Computational Toxicology (NCCT)

In-linking Resources

EMBL-EBI

Services Research Training About us

 UniChem

- Home / Search
- Web Services
- Connectivity Search
- Sources
 - General Info...
 - Background
 - Getting in touch
 - FAQ
 - Downloads
 - Connectivity Info
 - Other
- Analysis
 - Top Level Stats
 - Structures by Source

EBI > Databases > Small Molecules > UniChem

Data Sources.

UniChem currently contains data from the sources listed below. Follow the links on the short names for more detailed information on each source...

Show All entries

Apply filter: ...to whole table

src_id	Short name	Full name	Description	Process of Data Acquisition
1	chembl	ChEMBL	A database of bioactive drug-like small molecules and bioactivities abstracted from the scientific literature.	Standard InChIs and Keys provided on ftp site for each release.
2	drugbank	DrugBank	A database that combines drug (i.e. chemical, pharmacological and pharmaceutical) data with drug target (i.e. sequence, structure, and pathway) information.	Standard InChIs and Keys provided within sd file on ftp site for each release.
3	pdb	PDBe (Protein Data Bank Europe)	The European resource for the collection, organisation and dissemination of data on biological macromolecular structures, including structures of small molecule ligands for proteins.	Standard InChIs and Keys provided by direct querying of Oracle DB.
31	bindingdb	BindingDB	A public, web-accessible database of measured binding affinities, focusing chiefly on the interactions of proteins considered to be drug-targets with small, drug-like molecules	Standard InChIs and Keys available within a tsv file from download page.
32	comptox	EPA (Environmental Protection Agency) CompTox Dashboard	The foundation of chemical safety testing relies on chemistry information such as high-quality chemical structures and physicochemical properties. This information is used by scientists to predict the potential health risks of chemicals. The CompTox Dashboard is part of a suite of dashboards developed by EPA to help evaluate the safety of chemicals. It provides access to a variety of data and information on over 700,000 chemicals currently in use and of interest to environmental researchers. Within the CompTox Dashboard, users can access chemical structures, experimental and predicted physicochemical and toxicity data, and additional links to relevant websites and applications. It maps curated physicochemical property data associated with chemical substances to their corresponding chemical structures	Standard InChIs and Keys obtained from download page
33	lipidmaps	LipidMaps	LIPID Metabolites And Pathways Strategy (LIPID MAPS) is a multi-institutional effort created to identify and quantitate, using a systems biology approach and sophisticated mass spectrometers, all of the major, and many minor, lipid species in mammalian cells, as well as to quantitate the changes in these species in response to perturbation	Standard InChIs and Keys obtained from download page



Main page
Community portal
Project chat
Create a new item
Recent changes
Random item
Query Service
Nearby
Help
Donate


Print/export
Create a book
Download as PDF
Printable version

Tools
What links here
Related changes
Special pages
Permanent link
Page information

 English  Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Project page **Discussion**

Read [Edit](#) [View history](#)

Search Wikidata 

Wikidata:Property proposal/DTXSID

[< Wikidata:Property proposal](#)

DSSTOX substance identifier [\[edit \]](#)

Originally proposed at [Wikidata:Property proposal/Natural science](#)

 Done. [DSSTOX substance identifier \(P3117\)](#) ([Talk and documentation](#))

Description This DTXSID is the DSSTox substance identifier used in the Environmental Protection Agency CompTox Dashboard.

Data type [External identifier](#)

Domain [chemical substance \(Q79529\)](#)

Example [benzene \(Q2270\)](#) → [DTXSID3039242](#)

Planned use Adding DTXSIDs to @wikidata using a bot or possible Mix&Match, based on InChIKey matches and this CCZero data on Figshare:
https://figshare.com/articles/Mapping_file_of_InChIStrings_InChIKeys_and_DTXSIDs_for_the_EPA_CompTox_Dashboard/3578313/1

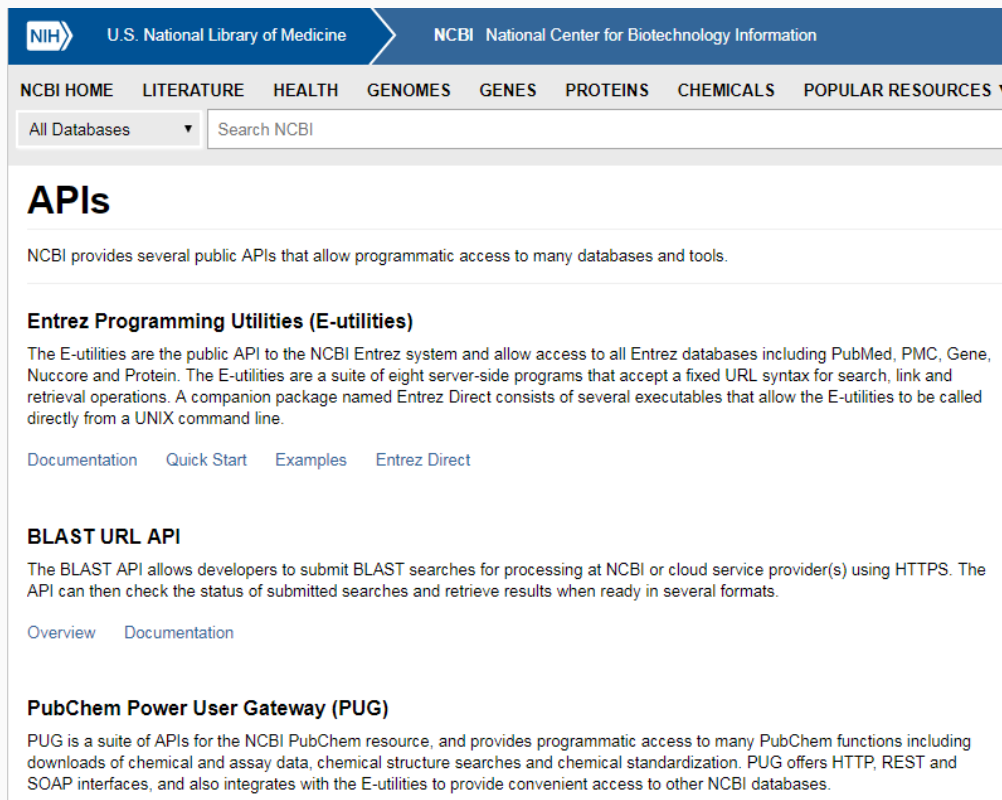
Formatter [https://comptox.epa.gov/dashboard/\\$1](https://comptox.epa.gov/dashboard/$1)
URL

Motivation

The [Environmental Protection Agency \(Q16919823\)](#)'s CompTox Dashboard is a fairly new website, but the DSSTox project exists for much longer ([Distributed structure-searchable toxicity \(DSSTox\) public database network: a proposal \(Q26701395\)](#)). The Dashboard aggregates over 720 thousand chemical substances and is open data. The CompTox dashboard also provides access to synonyms, experimental and predicted property data, product

Open APIs are of Value!

- Open APIs provide valuable integrations
 - PubChem API and widgets
 - NCBI Literature



The screenshot shows the NCBI website header with the NIH logo and navigation links for LITERATURE, HEALTH, GENOMES, GENES, PROTEINS, CHEMICALS, and POPULAR RESOURCES. Below the header is a search bar labeled "Search NCBI". The main content area is titled "APIs" and includes a sub-header "Entrez Programming Utilities (E-utilities)". The text describes the E-utilities as a public API to the NCBI Entrez system, providing access to various databases. Links for "Documentation", "Quick Start", "Examples", and "Entrez Direct" are provided. Below this is the "BLAST URL API" section, which describes the BLAST API for submitting searches and retrieving results. Links for "Overview" and "Documentation" are provided. The final section is the "PubChem Power User Gateway (PUG)", which describes the PUG as a suite of APIs for the NCBI PubChem resource, providing access to chemical and assay data. Links for "Overview" and "Documentation" are provided.

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information

NCBI HOME LITERATURE HEALTH GENOMES GENES PROTEINS CHEMICALS POPULAR RESOURCES ▼

All Databases Search NCBI

APIs

NCBI provides several public APIs that allow programmatic access to many databases and tools.

Entrez Programming Utilities (E-utilities)

The E-utilities are the public API to the NCBI Entrez system and allow access to all Entrez databases including PubMed, PMC, Gene, Nucleotide and Protein. The E-utilities are a suite of eight server-side programs that accept a fixed URL syntax for search, link and retrieval operations. A companion package named Entrez Direct consists of several executables that allow the E-utilities to be called directly from a UNIX command line.

[Documentation](#) [Quick Start](#) [Examples](#) [Entrez Direct](#)

BLAST URL API

The BLAST API allows developers to submit BLAST searches for processing at NCBI or cloud service provider(s) using HTTPS. The API can then check the status of submitted searches and retrieve results when ready in several formats.

[Overview](#) [Documentation](#)

PubChem Power User Gateway (PUG)

PUG is a suite of APIs for the NCBI PubChem resource, and provides programmatic access to many PubChem functions including downloads of chemical and assay data, chemical structure searches and chemical standardization. PUG offers HTTP, REST and SOAP interfaces, and also integrates with the E-utilities to provide convenient access to other NCBI databases.

PubChem Bioassay Data Widget

ToxCast

PubChem

PubChem Biological Activities

PUBCHEM > COMPOUND > ATRAZINE > BIOLOGICAL TEST RESULTS > BIOASSAY RESULTS >

BioAssay Results



Refine/Analyze

Download

All (1,491) Active(23) Inconclusive(35) Inactive(1,282) Unspecified(151)					
1 to 5 of 23					
1 2 3 ... 5					
Activity					
Activity	Activity Value [μM]	Substance SID	BioAssay AID	BioAssay Name	Target
Active		17390035	662	Cell signaling CRE-BLA (Fsk stim)	
Active		48413241	1189	DSSTox (CPDBAS) Carcinogenic Potency Database Summary SingleCellCall Results	
Active		48413241	1205	DSSTox (CPDBAS) Carcinogenic Potency Database Summary MultiCellCall Results	
Active		48413241	1208	DSSTox (CPDBAS) Carcinogenic Potency Database Summary Rat Bioassay Results	
Active		22407690	1663	MLPCN Platelet Activation -Dense Granule Release	

from PubChem

PubChem

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

PubChem Articles Widget

Google Scholar

PubMed Abstract Sifter

PubChem Articles

PubChem Patents


IRIS

PubChem Articles

[PUBCHEM](#) > [COMPOUND](#) > [ATRAZINE](#) > [LITERATURE](#) > [DEPOSITOR PROVIDED PUBMED CITATIONS](#) >



Depositor Provided PubMed Citations







 Download

1 to 5 of 2,556

1 2 3 ... 512

 Date 

PMID 	Date 	Title 	Journal 
26962873	2016-07-01	A High-Throughput Screening Strategy to Identify Protein-Protein Interaction Inhibitors That Block the Fanconi Anemia DNA Repair Pathway.	Journal of biomolecular screening
18378257	2008-05-09	Univariate method for background correction in liquid chromatography-Fourier transform infrared spectrometry.	Journal of chromatography. A
1430	1976-01-01	Hydrolysis of a chloro-s-triazine herbicide.	Journal of agricultural and food chemistry
5165	1976-03-01	N-nitrosamine formation from atrazine.	Bulletin of environmental contamination and toxicology
119497	1979-01-01	Sorption and desorption of atrazine by three bacterial species isolated from aquatic systems.	Archives of environmental contamination and toxicology

from PubChem





Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

PubChem Patents Widget

Google Scholar

PubMed Abstract Sifter

PubChem Articles

PubChem Patents

IRIS

PubChem Patents

PUBCHEM > COMPOUND > ATRAZINE > PATENTS > DEPOSITOR-SUPPLIED PATENT IDENTIFIERS >





Depositor-Supplied Patent Identifiers

Download

1 to 5 of 8,189

1 2 3 ... 1,638

Submitted Date

Patent ID 	Patent Title 	Submitted Date 	Granted Date 
EP0000002	TETRAHYDROFURANE DERIVATIVES, PROCESSES FOR THEIR PREPARATION AND THEIR USE AS HERBICIDES	1978-12-20	1981-08-26
EP0000681	AMINO-2 (OR -4) ALKYLTHIO-5 PYRIMIDINES, PROCESSES FOR THEIR PREPARATION, THEIR USE AS HERBICIDES AND COMPOSITION CONTAINING THEM	1979-02-07	1981-08-26
EP0000852	TREATMENT OF HERBICID CONTAINING CULTURES WITH DERIVATIVES OF DICHLOROACETAMIDE OR OF TRICHLOROACETAMIDE, NOVEL DERIVATIVES OF DICHLOROACETAMIDE OR TRICHLOROACETAMIDE AND PROCESS FOR THEIR PREPARATION	1979-02-21	1982-08-25
EP0001519	WATER-DISPERSIBLE HERBICIDAL COMPOSITIONS AND USE THEREOF	1979-04-18	1982-01-06
EP0002881	5-Isoxazolyurea derivatives, their preparation, their use and compositions containing them.	1979-07-11	

from PubChem

PubChem

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

PubMed Literature Searching

[Google Scholar](#)
Abstract Sifter
[PubChem Articles](#)
[PubChem Patents](#)
[IRIS](#)

Select Term:

Select a Query Term ▼

[Retrieve Articles](#) [0 Articles](#)

Add additional query terms to filter abstracts:

[Search and Count](#)

Edit the Query Before Retrieving Articles

"1912-24-9" OR "Atrazine" OR "Atrazine"

Select Term:

Select a Query Term ▼

Select a Query Term

Hazard
Fate and Transport
Metabolism/PK/PD
Chemical Properties
Exposure
Mixtures
Male Reproduction
Androgen Disruption
Female Reproduction
GeneTox
Embryo and embryonic development
Child (infant through adolescent)
Dust and Exposure

[Chemical Properties](#)[Synonyms](#)[External Links](#)[Product Composition](#)[Bioassays](#)[Exposure](#)[Analytical](#)[Literature](#)[Comments](#)

Integrated Literature Searching

Select Term:

Hazard

Edit the Query Before Retrieving Articles

("1912-24-9" OR "Atrazine" OR "Atrazine") AND (NOAEL or NOEL OR LOEL or Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level" [tiab] OR "cancer slope factor"[tiab])

Retrieve Articles

28 Articles (out of 28)

Add additional query terms to filter abstracts:

estrogen

endocrine

Search and Count

est...	end...	Ter...	Total	PMID	Pub..	Title
3	1	0	4	10613397	1999	Estrous cycle patterns of Sprague-Dawley rats during acute and chronic atrazine administration.
2	1	0	3	15655806	2005	A risk assessment of atrazine use in California: human health and ecological aspects.
1	1	0	2	15590006	2004	Impact of endocrine toxicants on survival, development, and reproduction of the estuarine copepod Eurytemora affinis...
0	1	0	1	8829248	1996	Effect of atrazine on ovarian function in the rat.
0	0	0	0	28089211	2017	Toxicity effects of an environmental realistic herbicide mixture on the seagrass Zostera noltei.
Record: [First] [Previous] 1 of 28 [Next] [Last]						

Title: Effect of atrazine on ovarian function in the rat.

Abstract: The effect of the chlorotriazine herbicide, atrazine, on ovarian function was studied in Long-Evans hooded (LE-hooded) and Sprague-Dawley (SD) rats. Atrazine was administered by gavage for 21 d to females displaying regular 4-d estrous cycles. In both strains, 75 mg/kg/d disrupted the 4-d ovarian cycle; however, no distinct alteration (i.e., irregular cycles but not persistent estrus or diestrus) was apparent at this dose. At 150 mg/kg/d, atrazine induced repetitive pseudopregnancies in females of both strains. The highest dose tested (300 mg/kg/d) also induced repetitive pseudopregnancies in the SD females, while the ovaries of the LE-hooded female appeared regressed and the smear cytology was indicative of the anestrus condition. Although a NOAEL was not established, the doses employed in this experiment were in excess of those used in chronic feeding studies in which an early onset of mammary gland tumors was noted. These data demonstrate that atrazine can disrupt ovarian function and bring about major changes in the **endocrine** profile of the female.

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

Integrated Literature Searching

Select Term:

Hazard

Retrieve Articles

28 Articles (0)

Add additional query terms to filter

estrogen

Edit the Query Before Retrieving Articles

("1912-24-9" OR "Atrazine" OR "Atrazine") AND (NOAEL or NOEL OR LOEL or Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

est...	end...	Ter...	Total	PMID	Pub..	Title
3	1	0	4	10613397	1999	Estrous cycle patterns of Sprague-Dawley rats during acute and chronic atrazine administration.
2	1	0	3	15655806	2005	A risk assessment of atrazine use in California: human health and ecological aspects.
1	1	0	2	15590006	2004	Impact of endocrine toxicants on survival, development, and reproduction of the estuarine copepod Eurytemora affinis...
0	1	0	1	8829248	1996	Effect of atrazine on ovarian function in the rat.
0	0	0	0	28089211	2017	Toxicity effects of an environmental realistic herbicide mixture on the seagrass Zostera noltei.
Record: [First] [Previous] 1 of 28 [Next] [Last]						

Title: Effect of atrazine on ovarian function in the rat.

Abstract: The effect of the chlorotriazine herbicide, atrazine, on ovarian function was studied in Long-Evans hooded (LE-hooded) and Sprague-Dawley (SD) rats. Atrazine was administered by gavage for 21 d to females displaying regular 4-d estrous cycles. In both strains, 75 mg/kg/d disrupted the 4-d ovarian cycle; however, no distinct alteration (i.e., irregular cycles but not persistent estrus or diestrus) was apparent at this dose. At 150 mg/kg/d, atrazine induced repetitive pseudopregnancies in females of both strains. The highest dose tested (300 mg/kg/d) also induced repetitive pseudopregnancies in the SD females, while the ovaries of the LE-hooded female appeared regressed and the smear cytology was indicative of the anestrus condition. Although a NOAEL was not established, the doses employed in this experiment were in excess of those used in chronic feeding studies in which an early onset of mammary gland tumors was noted. These data demonstrate that atrazine can disrupt ovarian function and bring about major changes in the **endocrine** profile of the female.

Chemical Properties

Synonyms

External Links

Product Composition

Bioassays

Exposure

Analytical

Literature

Comments

Downloadable Data File

<https://comptox.epa.gov/dashboard/downloads>

Downloads

[DSSTox Identifier to PubChem Identifier Mapping File](#)

Posted: 11/14/2016

The DSSTox to PubChem Identifiers mapping file is in TXT format and includes the PubChem SID, PubChem CID and DSSTox substance identifier (DTXSID).
image

[DSSTox identifiers mapped to CAS Numbers and Names File](#)

Posted: 11/14/2016

The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name.
image

[DSSTox MS Ready Mapping File](#)

Posted: 11/14/2016

The CompTox Chemistry Dashboard can be used by mass spectrometrists for the purpose of structure identification. A normal formula search would search the exact formula associated with any chemical, whether it include solvents of hydration, salts or multiple components. However, mass spectrometry detects ionized chemical structures and molecular formulae searches should be based on desalted, and desolvated structures with stereochemistry removed. We refer to these as "MS ready structures" and the MS-ready mappings are delivered as Excel Spreadsheets containing the Preferred Name, CAS-RN, DTXSID, Formula, Formula of the MS-ready structure and associated masses, SMILES and InChI Strings/Keys.
image

[DSSTox SDF File](#)

Posted: 12/14/2016

This zip file contains the entire chemical structure collection of over 700,000 chemicals from the DSSTox database contained in one large SDF file. The file contains the structure, The DSSTox Structure Identifier (DTXCID), The DSSTOX Substance Identifier (DTXSID listed as PubChem External Data Source), the associated Dashboard URL, associated synonyms and Quality Control Level details. In order to view an SDF file you will need to have access to the appropriate piece of software to open an SDF files. Examples include ChemAxon JChem, ACD/ChemFolder or ChemDraw.
image

~720,000 structures

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

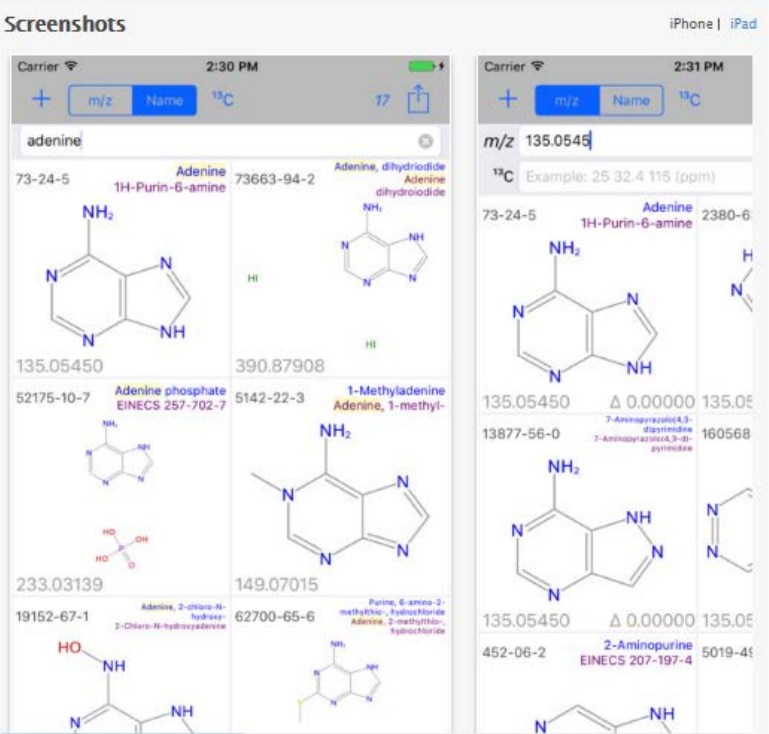
Description

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



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

Name Searching: "Bisp"

AT&T 8:59 AM 30% 25

+ m/z Name ¹³C

bisp

<p>42310-84-9 Sulfur fluoride oxide Bispentafluorosulfur oxide</p> <p>269.92309</p>	<p>57677-95-9 Bis(3,3,4,4,5,5,6,6,7,7,8,8,8-tetradecafluorooctyl) hydrogen phosphate bisperfluorooctyl phosphate</p> <p>789.98233</p>	<p>80-05-7 Bisphenol A BPA</p> <p>228.11503</p>	<p>24447-78-7 Bisphenol A bis(2-hydroxyethyl ether) diacrylate EINECS 246-253-2</p> <p>424.18859</p>
<p>24448-20-2 Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate 2,2-Bis(4-(2-(methacryloxyethoxy)phenyl)propane</p> <p>452.21989</p>	<p>116-37-0 Bisphenol A bis(2-hydroxypropyl ether) 2,2-Bis(p-(2-hydroxy-2-methylethoxy)phenyl)propane</p> <p>344.19876</p>	<p>5945-33-5 Phosphoric acid, P,P'-([1-methylethylidene]di-4,1-phenylene) P,P',P'-tetraphenyl ester Bisphenol A bis(diphenyl phosphate)</p> <p>692.17289</p>	<p>1675-54-3 Bisphenol A diglycidyl ether Oxirane, 2,2'-([1-methylethylidene]bis(4,1-phenyleneoxy)methylene)bis-</p> <p>340.16746</p>
<p>1565-94-2 Bisphenol A glycidyl methacrylate Bisphenol A glycidyl methacrylate (2-Propenoic acid, 2-methyl-, (1-methylethylidene)bis(4,1-phenyleneoxy)(2-hydroxy-3,1-propanediyl) ester)</p> <p>106100-55-4 Bisphenol A propoxylate diglycidyl ether</p> <p>105650-05-3 Bisphenol A propoxylate glycerolate diacrylate</p>	<p>2444-90-8 Phenol, 4,4'-([1-methylethylidene]bis-, disodium salt Bisphenol A sodium salt</p>		

Mass Searching

AT&T 9:02 AM 30%

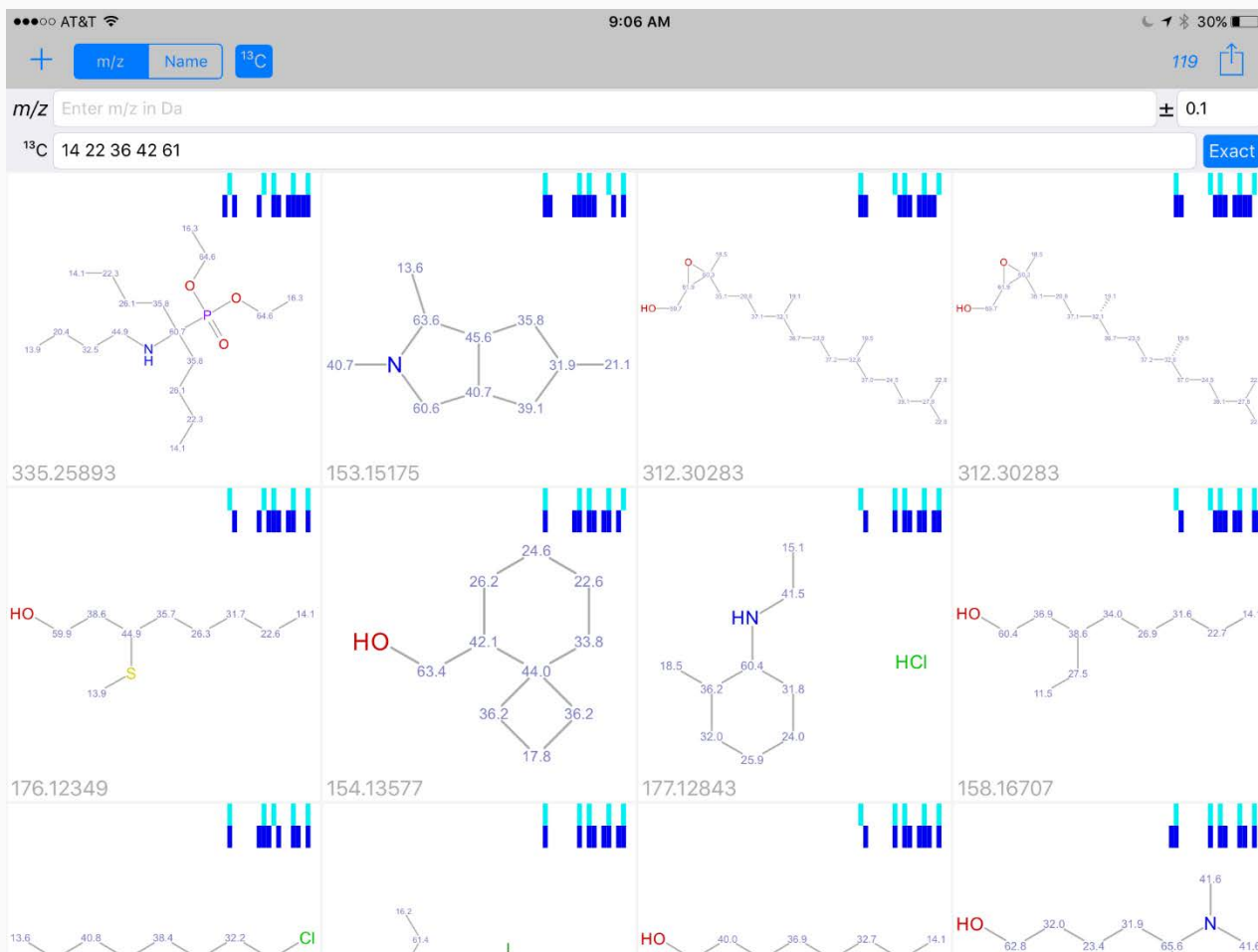
+ m/z Name ¹³C 82

m/z 300.156 ± 0.005

¹³C Example: 25 32.4 115 (ppm) [Exact](#)

<p>42978-66-5</p> <p>Tripropylene glycol diacrylate 2-Propenoic acid, (1-methyl-1,2-ethanediyloxy)methyl-2,1-ethanediyloxy 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 5-17-06-00498</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, 2,2'2''-[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-oxanon-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



- The CompTox Chemistry Dashboard provides access to data for ~760,000 chemicals
- Data downloads allows for reuse in other systems and integration of resources to support research
 - Mobile applications
 - Integration between resources – PubChem, ChEMBL, ChemSpider, etc.
- Open OPERA Models and curated training and test data freely available
- OUP Open API and prediction web services are imminent

Acknowledgments



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>