

The US-EPA CompTox Chemicals Dashboard – a key player in the domain of Open Science, Cheminformatics, and Online Databases supporting Non-Targeted Screening

***Antony J. Williams, Elin Ulrich, Joachim Pleil,
Alex Chao, Charles Lowe, Andrew McEachran*
and Jon Sobus***

Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

**Agilent Inc.*

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


...


Pittcon, March 2020

- Quick overview of the dashboard
- Specific data of interest to this audience (it's not just Computational Toxicology)
- Support for Mass Spectrometry
- Data ***quality*** in the public domain
- Work in progress – prototypes


BASIC Search

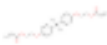
Chemicals **Product/Use Categories** **Assay/Gene**

 Bisphenol

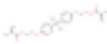


Bisphenol A
DTXSID7020182

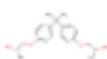




Bisphenol A bis(2-hydroxyethyl ether) diacrylate
DTXSID6066991

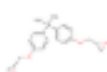


Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate
DTXSID1066992

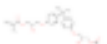


Bisphenol A bis(2-hydroxypropyl) ether
DTXSID8051592

Bisphenol A carbonate polymer
DTXSID6027840



Bisphenol A diglycidyl ether
DTXSID6024624



Bisphenol A glycidyl methacrylate
DTXSID7044841

Detailed Chemical Pages

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES


SYNONYMS

▶ LITERATURE

LINKS

COMMENTS

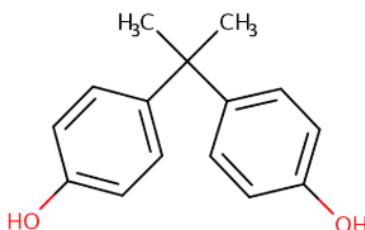
[Home](#) [Advanced Search](#) [Batch Search](#) [Lists](#) [Predictions](#) [Downloads](#) [Copy](#) [Share](#) [Submit Comment](#)



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.



Wikipedia

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4\text{OH})_2$ belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water (0.344 wt % at 83 °C). BPA is a starting material for the synthesis of plastics, primarily certain polycarbonates

[Read more](#)

Intrinsic Properties

Molecular Formula: $\text{C}_{15}\text{H}_{16}\text{O}_2$ [Mol File](#) [Find All Chemicals](#)

Average Mass: 228.291 g/mol [Isotope Mass Distribution](#)

Monoisotopic Mass: 228.11503 g/mol

Structural Identifiers

Linked Substances

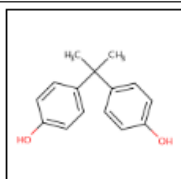
Presence in Lists

Record Information

Quality Control Notes

4

Properties, Fate and Transport



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Property

Summary


Summary

Download

Columns

Property	Experimental average	Predicted average	Experimental median	Predicted median
LogKow: Octanol-Water	3.32 (1)	3.30		3.39
Melting Point	155 (7)	140	156	144
Boiling Point	200 (1)	360		355
Water Solubility	8.55e-4 (3)	8.78e-4	5.26e-4	7.56e-4
Vapor Pressure	-	6.83e-7		1.51e-7
Flash Point	-	190		190
Surface Tension	-	46.0		
Index of Refraction	-	1.60		
Molar Refractivity	-	68.2		

Sources of Exposure to Chemicals

United States
Environmental Protection
Agency

HomeAdvanced SearchBatch SearchLists ▼PredictionsDownloads

Copy ▼Share ▼Submit CommentSearch all data

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

Download ▼

Columns ▼10 ▼

Product or Use Categorization

manufacturing, metals

adhesive

Product and Use Categories (PUCs) ⓘ

Product or Use Categorization	Categorization type	Number of Unique Products
manufacturing, metals	CPCat Cassette	17
adhesive	CPCat Cassette	17
	CPCat Cassette	16
	CPCat Cassette	12
	CPCat Cassette	11
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	7
	CPCat Cassette	6

First<<<12345678910>>>Last

EXPOSURE

PRODUCT & USE CATEGORIES

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

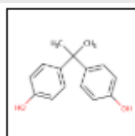
TOXICS RELEASE INVENTORY

MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

Identifiers to Support Searches



Bisphenol A

80-05-7 | DTXSID7020182

Searched by Approved Name.

Synonyms

Download

25

Search query

Synonym	Quality
Bisphenol A	Valid
4,4'-(Propane-2,2-diyl)diphenol	Valid
Phenol, 4,4'-(1-methylethylidene)bis-	Valid
80-05-7 Active CAS RN	Valid
BPA	Valid
4,4'-Propane-2,2-diyl diphenol	Valid
Phenol, 4,4'-(1-methylethylidene)bis-	Valid
4-06-00-06717 Beilstein Registry Number	Beilstein
(4,4'-Dihydroxydiphenyl)dimethylmethane	Good
2,2-Bis(4'-hydroxyphenyl) propane	Good
2,2'-Bis(4-hydroxyphenyl)propane	Good
2,2-BIS-(4-HYDROXY-PHENYL)-PROPANE	Good
2,2-Bis(4'-hydroxyphenyl)propane	Good
2,2-Bis(p-hydroxyphenyl)propane	Good
2,2-Di(4-Hydroxyphenyl) Propane	Good

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

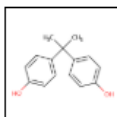
RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

COMMENTS



Bisphenol A

80-05-7 | DTXSID7020182

Searched by Approved Name.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES


SYNONYMS

▶ LITERATURE


LINKS

COMMENTS

General

 EPA Substance Registry Service

 Household Products Database

 Chemical Entities of Biological Interest (ChEBI)

 PubChem

 ChempSpider

 CPCat

 DrugBank


 HMDB

 Wikipedia

 MSDS Lookup


 ChEMBL

 Chemical Vendors

 CalEPA Office of Environmental Health Hazard Assessment

 NIOSH Chemical Safety Cards

 ToxPlanet


 ACS Reagent Chemicals

 Wikidata

 ChemHat: Hazards and Alternatives Toolbox

 Wolfram Alpha

 ScrubChem

 ECHA Brief Profile

Toxicology

 ACToR

 DrugPortal

 CCRIS

 ChemView

 CTD

 eChemPortal


 Gene-Tox


 HSDB


 ToxCast Dashboard 2

 LactMed

 International Toxicity Estimates for Risk

 ATSDR Toxic Substances Portal


 Superfund Chemical Data matrix

 NIOSH IDLH Values

 ACToR PDF Report

 Toxics Release Inventory

 CREST

 National Air Toxics Assessment

Publications

 Toxline

 Environmental Health Perspectives


 NIEHS

 National Toxicology Program

 Google Books


 Google Scholar

 Google Patents


 PPRTVWEB

 PubMed

 IRIS Assessments

 EPA HERO

 NIOSH Skin Notation Profiles

 NIOSH Pocket Guide


 RSC Publications

 BioCaddie DataMed

 Springer Materials


 Federal Register


 Regulations.gov

 Bielefeld Academic Search Engine

 CORE Literature Search


Analytical

 FOR-IDENT

 NEMI: National Environmental Methods Index

 RSC Analytical Abstracts

 Tox21 Analytical Data

 MONA: MassBank North America

 mzCloud

 NIST NIST IR Spectrum

 NIST NIST MS Spectrum

Prediction

 2D NMR HSQC/HMBC Prediction

 Carbon-13 NMR Prediction

 Proton NMR Prediction

 ChemRTP Predictor

 LSERD

Analytical



[RSC Analytical Abstracts](#)



[Tox21 Analytical Data](#)



[MONA: MassBank North America](#)



[mzCloud](#)



[NIST IR Spectrum](#)



[NIST MS Spectrum](#)



[MassBank](#)



[NEMI: National Environmental Methods Index](#)



[NIST Antoine Constants](#)





[IR Spectra on PubChem](#)




[NIST Kovats Index values](#)


Analytical

 FOR-IDENT

 NEMI: National Environmental
Methods Index

 RSC Analytical Abstracts

 Tox21 Analytical Data

 MONA: MassBank North
America

 mzCloud

 NIST IR Spectrum

 NIST MS Spectrum

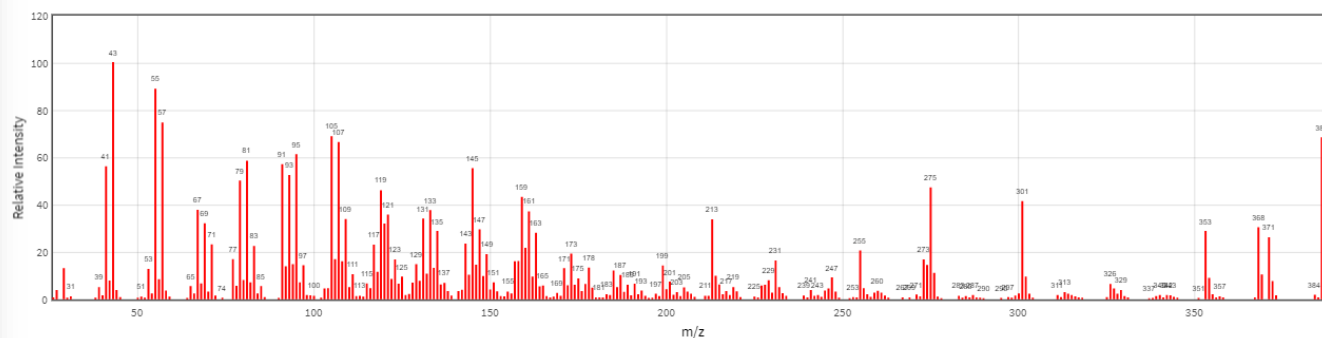
Spectrum

Plot

Help / Software credits

Cholesterol

Mass Spectrum



MassBank of North America

<https://mona.fiehnlab.ucdavis.edu>

Analytical

- FOR-IDENT
- NEMI: National Environmental Methods Index
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America**
- mzCloud
- NIST NIST IR Spectrum
- NIST NIST MS Spectrum

MoNA - MassBank of North America

Spectra Downloads Upload Help


Search...

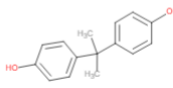
Display Generated Query

10 records/page

Bisphenol A

Score: ★★★★★





Originally submitted to the MassBank High Quality Mass Spectral Database

instrument	LTQ Orbitrap XL Thermo Sc...
instrument type	LC-ESI-ITFT
ms level	MS2
ionization	ESI
collision energy	30 % (nominal)
retention time	14.0 min
precursor m/z	229.1223
precursor type	[M+H] ⁺
ionization mode	positive
accession	EA016309

MassBank LC-MS Display Full Record

Batch Searching

Aggregate data for a list of chemicals



Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059



Opioid occurrence in environmental water samples—A review

Marina Celia Campos-Mañas ^a, Imma Ferrer ^b  , E.Michael Thurman ^b, Ana Agüera ^a

 [Show more](#)

<https://doi.org/10.1016/j.teac.2018.e00059>

[Get rights and content](#)

Batch Search Names

Buprenorphine
 Codeine
 Dextromethorphan
 Dihydrocodeine
 Dihydromorphine
 Ethylmorphine
 Fentanyl
 Heroin
 Hydrocodone
 Hydromorphone
 Ketamine
 Meperidine
 Methadone
 Morphine
 Morphinone
 Naloxone
 Naltriben
 Oxycodone
 Oxymorphone
 Propoxyphene
 Sufentanil
 Tramadol

Step 1 Step 2 Step 3 Step 4 Step 5 Step 6

Step Five: Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

- ☒ Identifiers
 - ☒ Chemical Name
 - ☐ CASRN
 - ☐ InChIKey
 - ☐ DSSTox Substance ID
 - ☐ DSSTox Compound ID
 - ☐ InChIKey Skeleton
 - ☐ MS-Ready Formula(e)
 - ☐ Exact Formula(e)
 - ☐ Monoisotopic Mass

Enter Identifiers to Search (searches should be limited to <5000 identifiers)







Buprenorphine
 Codeine
 Dextromethorphan
 Dihydrocodeine
 Dihydromorphine
 Ethylmorphine
 Fentanyl
 Heroin
 Hydrocodone
 Hydromorphone

Excel
Download






INPUT	FOUND_BY	DTXSID
Buprenorphine	Approved Name	DTXSID2022705
Codeine	Approved Name	DTXSID2020341
Dextromethorphan	Approved Name	DTXSID3022908
Dihydrocodeine	Approved Name	DTXSID5022936
Dihydromorphine	Approved Name	DTXSID7048908
Ethylmorphine	Approved Name	DTXSID1046760
Fentanyl	Approved Name	DTXSID9023049
Heroin	Synonym	DTXSID6046761
Hydrocodone	Approved Name	DTXSID8023131
Hydromorphone	Approved Name	DTXSID8023133
Ketamine	Approved Name	DTXSID8023187
Meperidine	Approved Name	DTXSID9023253
Methadone	Approved Name	DTXSID7023273
Morphine	Approved Name	DTXSID9023336

Add Other Data of Interest






Chemical Identifiers

- ☒ DTXSID 
- ☒ Chemical Name 
- ☐ DTXCID 
- ☒ CAS-RN 
- ☒ InChIKey 
- ☐ IUPAC Name 

Structures

- ☐ Mol File 
- ☐ SMILES 
- ☐ InChI String 
- ☒ MS-Ready SMILES 
- ☐ QSAR-Ready SMILES 

Intrinsic And Predicted Properties

- ☒ Molecular Formula 
- ☐ Average Mass 
- ☒ Monoisotopic Mass 
- ☐ TEST Model Predictions 
- ☐ OPERA Model Predictions 

INPUT	DTXSID	CASRN	MOLECULAR_FORMULA	MONOISOTOPIC	MS_READY_SMILES
Buprenorphine	DTXSID202	52485-79-7	C29H41NO4	467.3035588	[H]C12CC3=C4C
Codeine	DTXSID202	76-57-3	C18H21NO3	299.1521435	[H]C12CC3=C4C
Dextromethamphetamine	DTXSID302	125-71-3	C18H25NO	271.1936144	[H]C12CC3=C(C=
Dihydrocodone	DTXSID502	125-28-0	C18H23NO3	301.1677936	[H]C12CC3=C4C
Dihydromorphine	DTXSID704	509-60-4	C17H21NO3	287.1521435	[H]C12CC3=C4C
Ethylmorphine	DTXSID104	76-58-4	C19H23NO3	313.1677936	[H]C12CC3=C4C
Fentanyl	DTXSID902	437-38-7	C22H28N2O	336.2201635	CCC(=O)N(C1CC
Heroin	DTXSID604	561-27-3	C21H23NO5	369.1576228	[H]C12CC3=C4C
Hydrocodone	DTXSID802	125-29-1	C18H21NO3	299.1521435	[H]C12CC3=C4C
Hydromorphone	DTXSID802	466-99-9	C17H19NO3	285.1364935	[H]C12CC3=C4C
Ketamine	DTXSID802	6740-88-1	C13H16ClNO	237.0920418	CNC1(CCCCC1=
Meperidine	DTXSID902	57-42-1	C15H21NO2	247.1572289	CCOC(=O)C1(CC
Methadone	DTXSID702	76-99-3	C21H27NO	309.2092645	CCC(=O)C(CC(C)
Morphine	DTXSID902	57-27-2	C17H19NO3	285.1364935	[H]C12CC3=C4C
Morphinone	DTXSID501	467-02-7	C17H17NO3	283.1208434	[H]C12CC3=C4C
Naloxone	DTXSID802	465-65-6	C19H21NO4	327.1470582	[H]C12CC3=C4C
Naltrexone	-	-	-	-	-
Oxycodone	DTXSID502	76-42-6	C18H21NO4	315.1470582	[H]C12CC3=C4C
Oxymorphone	DTXSID502	76-41-5	C17H19NO4	301.1314081	[H]C12CC3=C4C
Propoxyphene	DTXSID102	469-62-5	C22H29NO2	339.2198292	CCC(=O)OC(CC1
Sufentanil	DTXSID602	56030-54-7	C22H30N2O2S	386.2027994	CCC(=O)N(C1=C
Tramadol	DTXSID908	27203-92-5	C16H25NO2	263.188529	COC1=CC=CC(=

Chemical Lists of Interest...

225 Chemical Lists (and growing)

Home Advanced Search Batch Search Lists ▾ Predictions Downloads

Lists of Chemicals
List of Assays

Download ▾ Columns ▾ Copy Filtered Lists URL

List Acronym ▾	List Name ▾	Last Updated ▾	Number of Chemicals ▾	List Description ▾
HDXEXCH	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - Under HDX Conditions	2018-11-07	592	Observed species (deuterated and undeuterated) from the HDXNOEX list under hydrogen deuterium exchange conditions (Ruttkies, Schymanski et al. in prep.)
HDXNOEX	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - No Exchange	2018-11-07	765	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule high resolution mass spectrometry (Ruttkies, Schymanski et al. in prep.)
MASSBANKEUSP	MASSPECDB: MassBank.EU Collection: Special Cases	2017-07-16	263	The MassBank.EU list contains curated chemicals (Schymanski/Williams) associated with the literature/tentative/unknown/SI spectra available on MassBank.EU that are not available as part of the full MassBank collection of reference standard spectra.
MASSBANKREF	MASSPECDB: MassBank Reference Spectra Collection	2017-07-13	1267	This MassBank list contains chemicals associated with the full MassBank collection of reference standard spectra available on MassBank.EU, MassBank.JP and MassBank of North America as well as the Open Data collection, curated by Williams/Schymanski.
MYCOTOXINS	MASSPECDB: Mycotoxins from MassBank.EU	2017-08-02	88	This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to MassBank.EU by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada

“Volatilome” Human Breath

LIST: VOLATILOME: Human Breath

Search VOLATILOME Chemicals

☐ Identifier substring search

List Details

Description: This list is a subset of compounds detected in human breath and reported in the peer-reviewed literature and identified in experimental work at US-EPA. The bulk of the collection is extracted from the article "The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva" by de Lacy Costello et al in J. Breath Res. 8 (2014) 034001 ([DOI:10.1088/1752-7155/8/3/034001](https://doi.org/10.1088/1752-7155/8/3/034001)), from the article "On-line analysis of exhaled breath", by Bruderer et al in Chemical Reviews ([DOI:10.1021/acs.chemrev.9b00005](https://doi.org/10.1021/acs.chemrev.9b00005)) as well as an increasing number of chemicals identified in our own laboratory studies.

Number of Chemicals: 1075

1075 chemicals

Select all

Download

Send to Batch Search

Default



CASRN



DTXSID

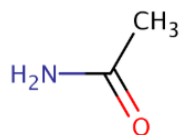


DTXSID

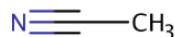


Hide chemicals that are:

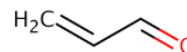
Filter by Name or CASRN



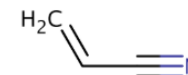
Acetamide
CASRN:60-35-5
DTXSID:DTXSID7020005



Acetonitrile
CASRN:75-05-8
DTXSID:DTXSID7020009



Acrolein
CASRN:107-02-8
DTXSID:DTXSID5020023



Acrylonitrile
CASRN:107-13-1
DTXSID:DTXSID5020029

Tire Crumb Rubber (298)

Related Topics: [Safer Chemicals Research](#)

[CONTACT US](#)

SHARE



July 2019 Report: Tire Crumb Rubber Characterization

Key Takeaways:

- EPA is releasing a new report that addresses exposure (that is, chemicals and how people come in contact with these) to tire crumb rubber on synthetic turf fields. **This report is not a risk assessment**, nor can the information be used to identify a level above which health effects could occur.
- In general, the findings for human exposure appear to be limited.
- Only Part 1 is being released for public comment and risk assessment.
- Part 1 of this report presents the findings of the literature search.
- The scope of this study was to identify potential chemical constituents in tire crumb rubber.

Tire Crumb Rubber

Search TIRECRUMB Chemicals

☐ Identifier substring search

List Details

Description: This chemical list is based on data contained within the [Federal Research Action Plan \(FRAP\) on Recycled Tire Crumb Used on Playing Fields and Playgrounds](#). The chemical list is obtained from the [Toxicity reference information spreadsheet](#) compiled for the potential tire crumb rubber chemical constituents identified in the State-of-Science Literature Review/Gaps Analysis, White Paper Summary of Results. Eleven sources of publicly available toxicity reference information were searched. It is important to recognize that not all potential chemical constituents identified through the literature search were confirmed through measurements made under the Federal Research Action Plan.

Number of Chemicals: 298

Select all

Download

Send to Batch Search

Default



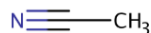
CASRN

DTXSID

298 chemicals

Hide chemicals that are:

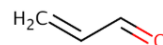
Filter by Name or CASRN



Acetonitrile

CASRN:75-05-8

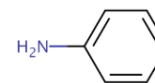
DTXSID:DTXSID7020009



Acrolein

CASRN:107-02-8

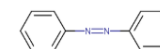
DTXSID:DTXSID5020023



Aniline

CASRN:62-53-3

DTXSID:DTXSID8020090



Azobenzene

CASRN:103-33-3

DTXSID:DTXSID8020123

Hydraulic Fracturing (1640)

EPA's Study of Hydraulic Fracturing and Its Potential Impact on Drinking Water Resources

[Contact Us](#)

Hydraulic Fracturing Study
Home

Final Assessment

EPA Published Research

Fact Sheets

Questions & Answers about
the final assessment

Multi-agency collaboration
on unconventional oil and
gas research

EPA Hydraulic Fracturing -
Agency Main Page

Hydraulic Fracturing For Oil And Gas: Impacts From The Hydraulic

WATER|EPA; Chemicals associated with hydraulic fracturing

☐ Identifier substring search

List Details

Description: Chemicals used in hydraulic fracturing fluids and/or identified in produced water from 2005-2013, corresponding to chemicals listed in Appendix H of EPA's Hydraulic Fracking Drinking Water Assessment Final Report (Dec 2016). Citation: U.S. EPA, Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). U.S. Environmental Protection Agency, Washington, D.C. EPA/600/R-16/236F, 2016. <https://www.epa.gov/hfstudy>

*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances (as of Jan 2018), both were included in the current EPAHFR chemical listing for completeness; additionally, 34 previously unmapped chemicals in Table H-5 are now registered in DSSTox (all but 2 assigned CASRN) and, thus, have been added to the current EPAHFR listing.

Number of Chemicals: 1640

Select all

Download

Send to Batch Search

Default

↑

CASRN

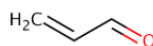
DTXSID

1640 chemicals

Hide chemicals that are:

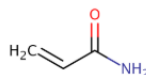
Filter by Name or CASRN

⌵



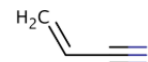
Acrolein

CASRN:107-02-8
DTXSID:DTXSID5020023



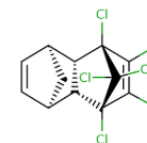
Acrylamide

CASRN:79-06-1
DTXSID:DTXSID5020027



Acrylonitrile

CASRN:107-13-1
DTXSID:DTXSID5020029



Aldrin

CASRN:309-00-2
DTXSID:DTXSID8020040

Opioids and Metabolites (160)

DRUGS: Opioids and related metabolites

☐ Identifier substring search

List Details

Description: This list of opioids and related metabolites is assembled primarily from public resources (e.g. Wikipedia, databases and literature articles) and is under ongoing curation and expansion.
Number of Chemicals: 180

Select all

Download

Send to Batch Search

Default

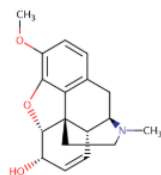
CASRN

DTXSID

180 chemicals

Hide chemicals that are:

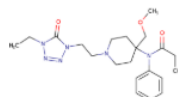
Filter by Name or CASRN



Codeine

CASRN:76-57-3

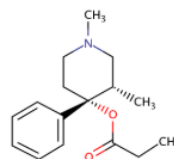
DTXSID:DTXSID2020341



Alfentanil

CASRN:71195-58-9

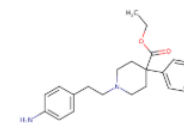
DTXSID:DTXSID9022570



Alphaprodine

CASRN:77-20-3

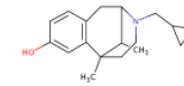
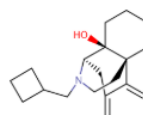
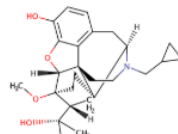
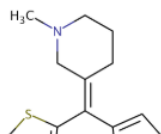
DTXSID:DTXSID4022575



Anileridine

CASRN:144-14-9

DTXSID:DTXSID8022610



Disinfection By-Products

LIST: Disinfection By-Products

Search DBPRODUCTS Chemicals

☐ Identifier substring search

List Details

Description: Disinfection by-products (DBPs) result from chemical reactions between organic and inorganic matter in water with chemical treatment agents during the water disinfection process. DBPs are present in most drinking water supplies that have been subject to chlorination, chloramination, ozonation, or treatment with chlorine dioxide.

Number of Chemicals: 87

87 chemicals

Select all

Download

Send to Batch Search

Default

↑

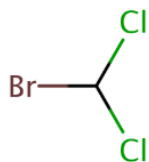
CASRN

DTXSID

Hide chemicals that are:

Filter by Name or CASRN

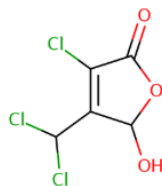
☰



Bromodichloromethane

CASRN:75-27-4

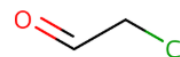
DTXSID:DTXSID1020198



3-Chloro-4-(dichloromethyl)-5-hydroxy-...

CASRN:77439-76-0

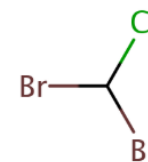
DTXSID:DTXSID6020276



Chloroacetaldehyde

CASRN:107-20-0

DTXSID:DTXSID4020292



Chlorodibromomethane

CASRN:124-48-1

DTXSID:DTXSID1020300

PFAS lists of Chemicals

Select List

Download

Columns

PFAS

Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	PFAS[EPA: List of 75 Test Samples (Set 1)]	2018-06-29	74	PFAS list corresponds to 75 samples (Set 1) submitted for initial testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFAS75S2	PFAS[EPA: List of 75 Test Samples (Set 2)]	2019-02-21	75	PFAS list corresponds to a second set of 75 samples (Set 2) submitted for testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFASCAT	PFAS[EPA Structure-based Categories]	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	PFAS[EPA: Chemical Inventory Insoluble in DMSO]	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	PFAS[EPA: ToxCast Chemical Inventory]	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	PFAS[EPA: Cross-Agency Research List]	2017-11-16	199	EPAPFASRL is a manually curated listing of mainly straight-chain and branched PFAS (Per- & Poly-fluorinated alkyl substances) compiled from various internal, literature and public sources by EPA researchers and program office representatives.
PFASKEMI	PFAS: List from the Swedish Chemicals Agency (KEMI) Report	2017-02-09	2416	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	PFAS Master List of PFAS Substances	2018-07-26	5061	PFASMASTER is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.
PFASOECD	PFAS: Listed in OECD Global Database	2018-05-16	4729	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing more than 4700 new PFAS
PFASTRIER	PFAS Community-Compiled List (Trier et al, 2015)	2017-07-16	597	PFASTRIER community-compiled public listing of PFAS (Trier et al, 2015)

MASSPECDB: Mycotoxins from MassBank.EU

☐ Identifier substring search

List Details

Description: This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to [MassBank.EU](#) by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada. This list is also a part of the [MASSBANKREE](#) list and the [NORMAN Suspect Exchange](#) and will be expanded as new contributions arrive.

Number of Chemicals: 88

88 chemicals

Select all

Download

Send to Batch Search

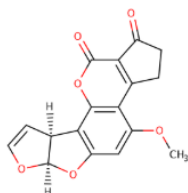
Default

CASRN

DTXSID

Hide chemicals that are:

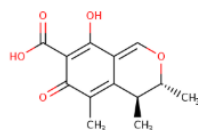
Filter by Name or CASRN



Aflatoxin B1

CASRN:1162-65-8

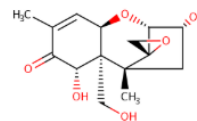
DTXSID:DTXSID9020035



Citrinin

CASRN:518-75-2

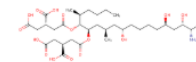
DTXSID:DTXSID8020333



Vomitoxin

CASRN:51481-10-8

DTXSID:DTXSID3020382



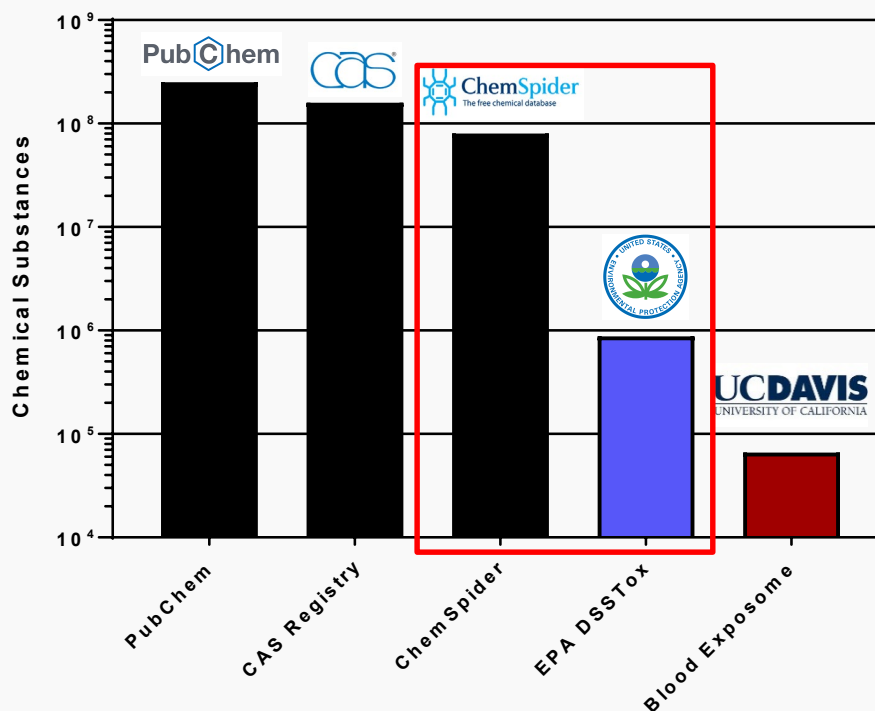
Fumonisin B1

CASRN:116355-83-0

DTXSID:DTXSID6020644

BIG databases are GREAT!

- Thanks to all of the public database efforts
- So much benefit from what's been done
- There are hundreds of them at this point...



Data Quality is important

- Data quality in free web-based databases!



Drug Discovery Today
Volume 17, Issues 13–14, July 2012, Pages 685–701



Drug Discovery Today
Volume 16, Issues 17–18, September 2011, Pages 747–750

Review
Keynote

Towards a gold standard:
quality in public domain
databases and approaches
the

Antony J. Williams
Show
<https://doi.org/10.1186/s13321-015-0057-7>

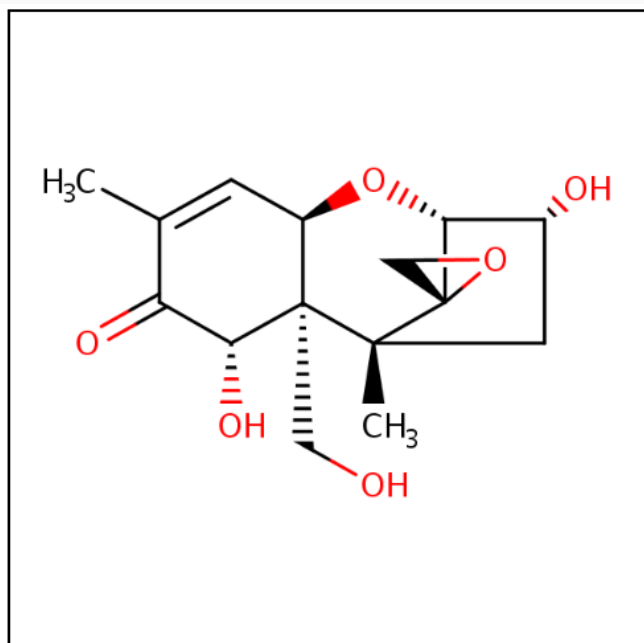
Editorial

**Machines first, humans second: on the importance
of algorithmic interpretation of open chemistry
data**

Alex M Clark , Antony J Williams and Sean Ekins

Journal of Cheminformatics 2015 7:9
<https://doi.org/10.1186/s13321-015-0057-7> | © Clark et al.; licensee Springer. 2015
Received: 24 November 2014 | Accepted: 23 February 2015 | Published: 22 March 2015

and content



Structural Identifiers

IUPAC Name: (3alpha,7alpha)-3,7,15-Trihydroxy-12,13-epoxytrichothec-9-en-8-one

SMILES: CC1=C[C@H]2O[C@@H]3[C@H](O)C[C@@](C)([C@]33CO3)[C@@]2(CO)[C@H]1O

InChI String: InChI=1S/C15H20O6/c1-7-3-9-14(5-16,11(19)10(7)18)13(2)4-8(17)12(20)3/t8-,9-,11-,12-,13-,14-,15+/m1/s1

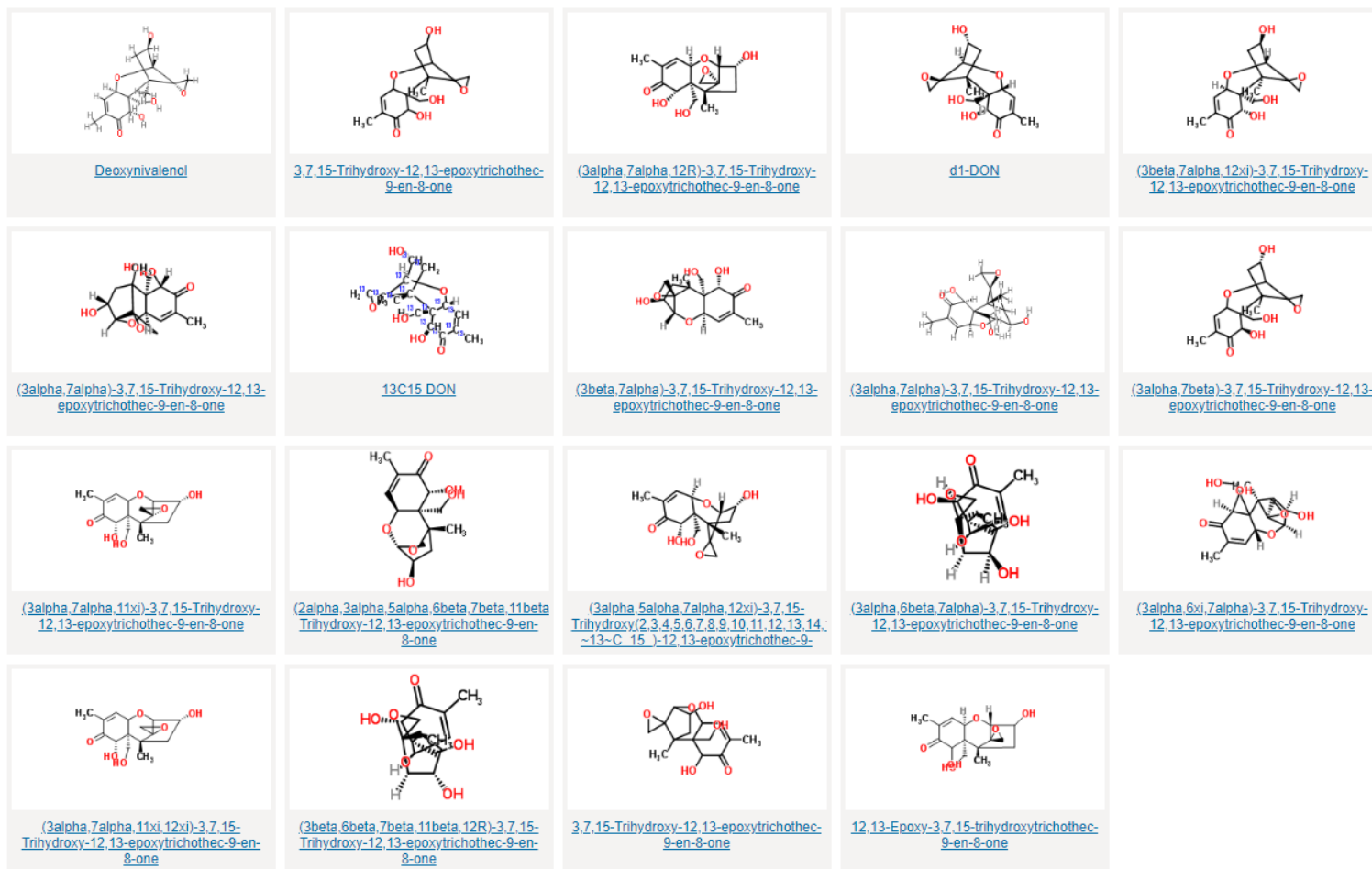
InChIKey: LINOMUASTDIRTM-QGRHZQQGSA-N

Search Google for:

Copy All

- 19 “Vomitoxins” – 3 isotopically labeled

Search term: **LINOMUASTDIRTM** (Found by InChIKey (skeleton match))



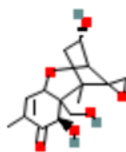
- 33 unique InChI Keys

Compounds
(33)

Substances
(10)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that pages is not searched. [Read More...](#)

33 results Filters SORT BY Relevance



Vomitoxin; Trichothec-9-En-8-One, 12,13-Epoxy-3,7,15-Trihydroxy-, (3.α.,7.α.)-; 12,13-Epoxy-3.α.,7.α.,15-Trihydroxy-9-Trichothecen-8-One; LINOMUASTDIRTM-LMJBVPRVSA-N; Spiro[2,5-Methano-1-Benzoxepin-10,2'-Oxirane], Trichothec-9-En-8-One Deriv.

Compound CID: 6432495

MF: $C_{15}H_{20}O_6$ MW: 296.31g/mol

InChIKey: LINOMUASTDIRTM-LMJBVPRVSA-N

IUPAC Name: (3R,10S)-3,10-dihydroxy-2-(hydroxymethyl)-1,5-dimethylspiro[8-oxatricyclo[7.2.1.0^{2,7}]dodec-5-ene-12,2'-oxirane]-4-one

Create Date: 2006-04-28

Not Vomitoxin

- Other databases grow quickly...a lot of “virtual chemistry” and “make on demand” compounds. Vomitoxin has 7 ZINC stereoforms.

PUBCHEM_CID	Compound_Name	Compound_Synonym	InChIKey
98043267	(1R,2S,3R,7R,9S,10R,12S)	ZINC100006545	LINOMUASTDIRTM-DOZBXCHUSA-N
98051113	(1R,2R,3S,7S,9S,10R,12S)	ZINC100066010	LINOMUASTDIRTM-KCWNRFLPSA-N
100853641	(1R,2R,3S,7S,9S,10R,12R)	ZINC229762267	LINOMUASTDIRTM-OMTHLLQNSA-N
98043268	(1R,2S,3R,7S,9S,10R,12S)	ZINC100006546	LINOMUASTDIRTM-UBTIPYQWSA-N
95566296	(1R,2S,3R,7R,9R,10R,12S)	ZINC71789640	LINOMUASTDIRTM-WYQUPHEGSA-N
100853642	(1R,2R,3S,7R,9S,10R,12R)	ZINC229762273	LINOMUASTDIRTM-XFRIDARHSA-N
95566297	(1R,2S,3R,7S,9R,10R,12S)	ZINC71789642	LINOMUASTDIRTM-XGQZSAOASA-N

- The Dashboard database grows slowly (next release is +20k chemicals in 6 months)

“MS-ready” structures

McEachran et al. *J Cheminform* (2018) 10:45
<https://doi.org/10.1186/s13321-018-0299-2>

Journal of Cheminformatics

METHODOLOGY

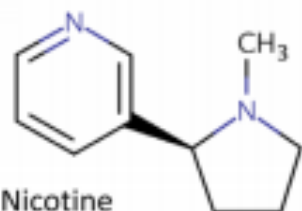
Open Access

**“MS-Ready” structures for non-targeted
high-resolution mass spectrometry screening
studies**



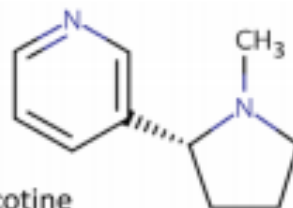
Andrew D. McEachran^{1,2*}, Kamel Mansouri^{1,2,3}, Chris Grulke², Emma L. Schymanski⁴, Christoph Ruttkies⁵
and Antony J. Williams^{2*}

- All structure-based chemical substances are algorithmically processed to
 - Split multicomponent chemicals into individual structures
 - Desalt and neutralize individual structures
 - Remove stereochemical bonds from all chemicals
- MS-Ready structures are then mapped to original substances to provide a path between chemicals detected by mass spectrometry to original substances



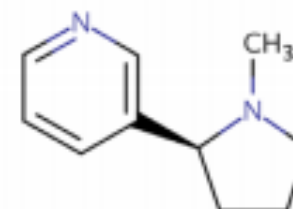
Nicotine

CN1CCC[C@H]1C1=CN=CC=C1
 DTXSID1020930 | SNICXCGAKADSCV
 54-11-5 | **162.1157** | 0.929 | **72**
 Tox: **yes** | Expo: **yes** | Bioassay: **yes**



D-Nicotine

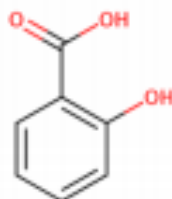
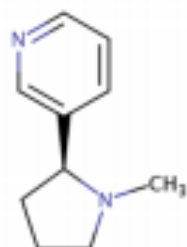
CN1CCC[C@H]1C1=CN=CC=C1
 DTXSID004635 | SNICXCGAKADSCV
 25162-00-9 | **162.1157** | 0.929 | **20**
 Tox: **no** | Expo: **yes** | Bioassay: **yes**



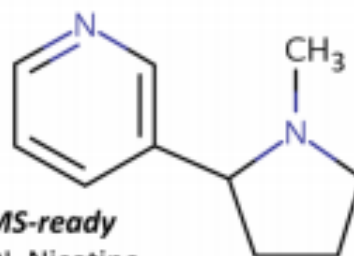
HCl

Nicotine hydrochloride

Cl.CN1CCC[C@H]1C1=CN=CC=C1
 DTXSID602093 | HDJBTCJAJIMNXEW
 2820-51-1 | **198.0924** | 0.929 | **9**
 Tox: **no** | Expo: **yes** | Bioassay: **yes**


 Benzoic acid, 2-hydroxy-, compd. with
 3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)

OC(=O)C1=CC=C(C=C1)C(=O)C1=CC=C(C=C1)C1=CN=CC=C1
 DTXSID5075319 | AIBWPBUAKCMKNS
 29790-52-1 | **300.1474** | 0.929 | **6**
 Tox: **no** | Expo: **yes** | Bioassay: **no**



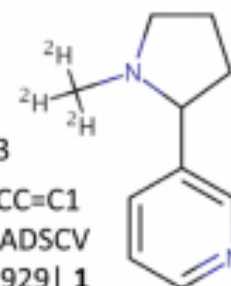
MS-ready

DL-Nicotine

CN1CCCC1C1=CN=CC=C1
 DTXSID3048154 | SNICXCGAKADSCV
 22083-74-5 | **162.1157** | 0.953 | **9**
 Tox: **yes** | Expo: **no** | Bioassay: **yes**

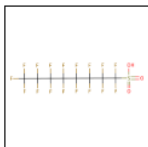
DL-Nicotine-d3

[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1
 DTXSID80442666 | SNICXCGAKADSCV
 69980-24-1 | **165.1345** | 0.929 | **1**
 Tox: **no** | Expo: **no** | Bioassay: **no**



LEGEND: Name, SMILES
 DTXSID | InChIKey 1st Block
 CAS | **Monoiso.** Mass | logP | **Sources**
 Data on: **Toxicity** | **Exposure** | **Bioassays**

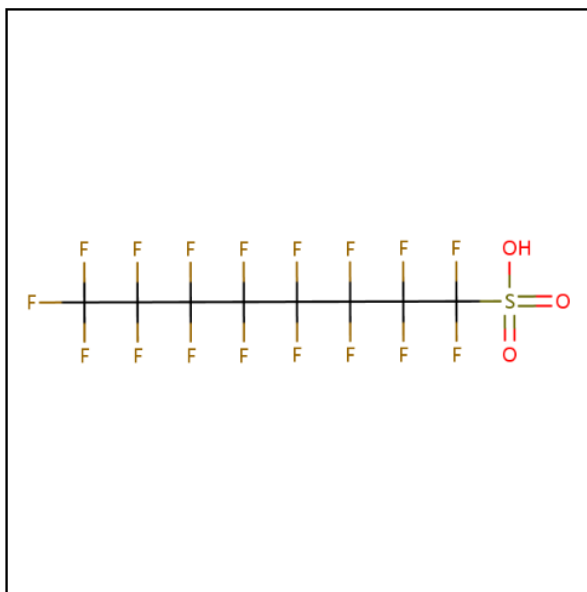
MS-Ready Mappings from Details Page



Perfluorooctanesulfonic acid

1763-23-1 | DTXSID3031864

Searched by Synonym from Valid Source.



Wikipedia

Perfluorooctanesulfonic acid (conjugate base **perfluorooctanesulfonate**) (**PFOS**) is an anthropogenic fluorosurfactant and global pollutant. PFOS was the key ingredient in Scotchgard, a fabric protector made by 3M, and numerous stain repellents. It was added to Annex B of the Stockholm Convention on Persistent Organic Pollutants in May 2009. PFOS can be synthesized in industrial production or result from the degradation of precursors. PFOS levels that have been detected in wildlife

...
[Read more](#)

Quality Control Notes

Intrinsic Properties

Structural Identifiers

Linked Substances

Same Connectivity: [4 records](#) (based on first layer of InChI)

Mixtures, Components and Neutralized Forms: [9 records](#) (based on QSAR ready mappings and with the compound as a component of a mixture)

MS-Ready Mappings: [DTXCID1011864: 18 records;](#)

Similar Compounds: [83 records](#) (based on Tanimoto coefficient >0.8)

[ed search/index](#)

MS-Ready Mappings Set of 20 substances for “PFOS”

EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads




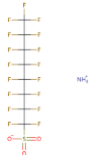
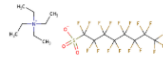

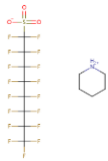

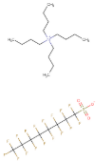
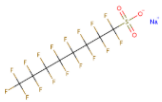

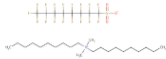
Share Search all data

MS-Ready Mappings of Perfluorooctanesulfonic acid (Isotopes pre-filtered)

18 of 20 chemicals visible

Select all Download Send to Batch Search Default CASRN DTXSID

Isotopes Filter by Name or CASRN

 <p>Perfluorooctanesulfonic acid CASRN:1763-23-1 DTXSID:DTXSID3031864</p>	 <p>Lithium perfluorooctanesulfonate CASRN:29457-72-5 DTXSID:DTXSID2032421</p>	 <p>Potassium perfluorooctanesulfonate CASRN:2795-39-3 DTXSID:DTXSID8037706</p>	 <p>Ammonium perfluorooctanesulfonate CASRN:29081-56-9 DTXSID:DTXSID9067435</p>	 <p>Tetraethylammonium perfluorooctanesulfonate CASRN:56773-42-3 DTXSID:DTXSID5069128</p>	 <p>Bis(2-hydroxyethyl)ammonium perfluorooctanesulfonate CASRN:70225-14-8 DTXSID:DTXSID2072049</p>
 <p>Piperidinium perfluorooctanesulfonate CASRN:71463-74-6 DTXSID:DTXSID0072352</p>	 <p>Perfluorooctanesulfonate CASRN:45298-90-6 DTXSID:DTXSID80108992</p>	 <p>Tetraethylammonium perfluorooctanesulfonate CASRN:111873-33-7 DTXSID:DTXSID40584995</p>	 <p>Sodium perfluorooctanesulfonate CASRN:4021-47-0 DTXSID:DTXSID50635462</p>	 <p>Magnesium bis(perfluorooctanesulfonate) CASRN:91036-71-4 DTXSID:DTXSID80881314</p>	 <p>N-Decyl-N,N-dimethyl-1-decanaminium perfluorooctanesulfonate CASRN:251099-16-8 DTXSID:DTXSID00882964</p>

Mass and Formula Searching

Advanced Searches

Mass Search

Mass Search

± Min/Max

Adduct

Neutral



All Adducts



Choose adduct from dropdown

191.131

Da

±

5

Da

ppm

Search 

Advanced Searches

Mass Search

Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

329 of 329 chemicals visible

Select all

Download

Send to Batch Search

Mass Difference

DTXSID

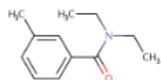
CASRN

TOXCAST

Mass Diff

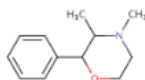
Multicomponent Chemicals

Filter by Name or CASRN



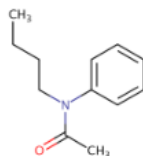
DEET

DTXSID: DTXSID2021995
CASRN: 134-62-3
TOXCAST: 12/768
Mass Diff: 0.000014



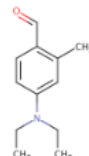
Phendimetrazine

DTXSID: DTXSID1023447
CASRN: 634-03-7
TOXCAST: -
Mass Diff: 0.000014



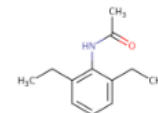
N-Butylacetanilide

DTXSID: DTXSID2042197
CASRN: 91-49-6
TOXCAST: -
Mass Diff: 0.000014



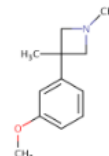
Benzaldehyde, 4-(diethylamino)-2-methyl-

DTXSID: DTXSID4059041
CASRN: 92-14-8
TOXCAST: -
Mass Diff: 0.000014



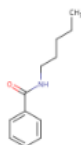
Acetanilide, 2',6'-diethyl-

DTXSID: DTXSID90168148
CASRN: 16665-89-7
TOXCAST: -
Mass Diff: 0.000014



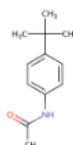
Azetidine, 1,3-dimethyl-3-(m-methoxyphenyl)-

DTXSID: DTXSID40173560
CASRN: 19832-26-9
TOXCAST: -
Mass Diff: 0.000014



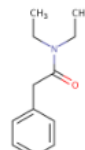
Benzamide, N-pentyl-

DTXSID: DTXSID20174196
CASRN: 20308-43-4
TOXCAST: -
Mass Diff: 0.000014



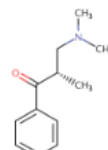
p-t-Butylacetanilide

DTXSID: DTXSID00174238
CASRN: 20330-45-4
TOXCAST: -
Mass Diff: 0.000014



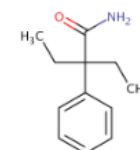
N,N-Diethylphenylacetamide

DTXSID: DTXSID00179048
CASRN: 2431-96-1
TOXCAST: -
Mass Diff: 0.000014



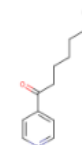
3-(Dimethylamino)-2-methylpropionophenone

DTXSID: DTXSID60180796
CASRN: 26171-50-6
TOXCAST: -
Mass Diff: 0.000014



Butyramide, 2-ethyl-2-phenyl-

DTXSID: DTXSID60184653
CASRN: 30568-39-9
TOXCAST: -
Mass Diff: 0.000014



1-Heptanone, 1-(4-pyridyl)-

DTXSID: DTXSID40186594
CASRN: 32941-30-3
TOXCAST: -
Mass Diff: 0.000014

MS-Ready Structures for Formula Search

Molecular Formula Search

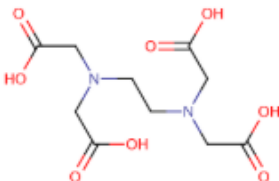

☒ MS Ready Formula  ☐ Exact Formula 

Formula

Please use the format of the following example: C₆H₈O₂ or C₆H(8-10)O(0-2)



Search 

- **EXACT Formula:** C₁₀H₁₆N₂O₈: 3 Hits

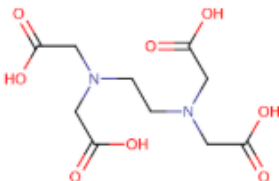
MS Ready Formula  **Exact Formula** 

Formula

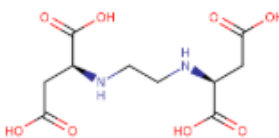
C₁₀H₁₆N₂O₈

Select all Download Send to Batch Search Default  DTXSID X PubChem X CPDAT X 

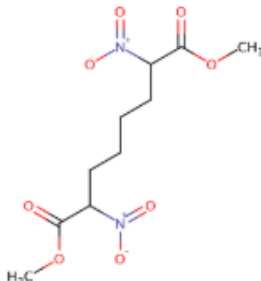
3 of 3 chemi



Ethylenediaminetetraacetic acid
DTXSID: DTXSID6022977
PubChem: 158
CPDAT: 387



N,N'-Ethylenedi-L-aspartic acid
DTXSID: DTXSID1051852
PubChem: 25
CPDAT: 8



Dimethyl 2,7-dinitrooctanedioate
DTXSID: DTXSID20498864
PubChem: 5
CPDAT: 0

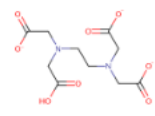
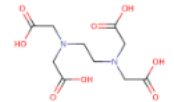
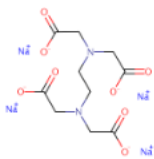
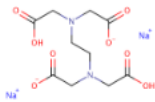
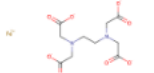
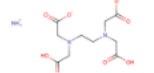
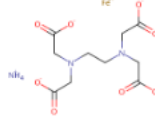
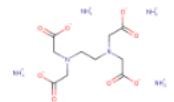
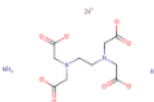
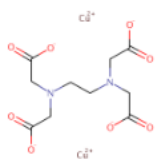
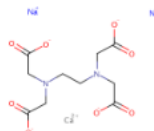
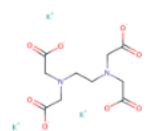
MS-Ready Mappings

- **Same Input Formula: C₁₀H₁₆N₂O₈**
- **MS Ready Formula Search: 125 Chemicals**

125 chemicals

Select all Download Send to Batch Search Default DTXSID PubChem CPDAT

Hide chemicals that are: Filter by Name or CASR

 <p>Trisodium ethylenediaminetetraacetate DTXSID: DTXSID7020556 PubChem: 33 CPDAT: 82</p>	 <p>Ethylenediaminetetraacetic acid DTXSID: DTXSID6022977 PubChem: 158 CPDAT: 387</p>	 <p>Ethylenediaminetetraacetic acid tetrasod... DTXSID: DTXSID3026350 PubChem: 57 CPDAT: 1227</p>	 <p>Ethylenediaminetetraacetic acid, disodiu... DTXSID: DTXSID9027073 PubChem: 56 CPDAT: 1359</p>	 <p>Ethylenediaminetetraacetic acid ferric so... DTXSID: DTXSID5027774 PubChem: 53 CPDAT: 62</p>	 <p>Diammonium dihydrogen ethylenediami... DTXSID: DTXSID9027813 PubChem: 12 CPDAT: 17</p>
 <p>Ferrate(1-), [[N,N'-1,2-ethanediy]bis[N-]]... DTXSID: DTXSID9027815 PubChem: 24 CPDAT: 20</p>	 <p>Tetraammonium ethylenediaminetetraac... DTXSID: DTXSID8027820 PubChem: 11 CPDAT: 12</p>	 <p>Zincate(2-), [[N,N'-1,2-ethanediy]bis[N-]]... DTXSID: DTXSID8028343 PubChem: 5 CPDAT: 10</p>	 <p>EDTA, copper salt DTXSID: DTXSID0034564 PubChem: 8 CPDAT: 10</p>	 <p>Calcium disodium ethylenediaminetetra... DTXSID: DTXSID2036409 PubChem: 42 CPDAT: 29</p>	 <p>Tetrapotassium ethylenediaminetetra... DTXSID: DTXSID3036442 PubChem: 25 CPDAT: 36</p>

- 125 chemicals returned in total
 - 8 of the 125 are **single component** chemicals
 - 3 of the 8 are **isotope-labeled**
 - 3 are **neutral compounds** and 2 are **charged**
- Multiple components, stereo, isotopes and charge all collapsed and mapped through MS-Ready

Batch Searching mass and formula

- Singleton searches are useful but we work with **thousands** of masses and formulae!
- Typical questions
 - What is the list of chemicals for the formula $C_xH_yO_z$
 - What is the list of chemicals for a mass +/- error
 - Can I get chemical lists in Excel files? In SDF files?
 - Can I include properties in the download file?

Batch Searching Formula/Mass

Batch Search


Step 1 Step 2 Step 3 Step 4 Step 5 Step 6


Step Five: Choose Data Fields to Download


Please enter one identifier per line ✕


Select Input Type(s)


☐ Identifiers


☐ Chemical Name 


☐ CASRN 


☐ InChIKey 


☐ DSSTox Substance ID 

☐ DSSTox Compound ID 

☐ InChIKey Skeleton 

☐ MS-Ready Formula(e) 

☐ Exact Formula(e) 

☒ Monoisotopic Mass 

+/- ppm

Enter Identifiers to Search (searches should be limited to <5000 identifiers)

41.0265

56.02621

53.0265

58.0418

93.0578

113.9639

151.8754

69.9377

77.9872

This search is based on what we refer to as "Mass Spec (MS) Ready" structures. All chemicals within the database are treated in a manner such that all are desalted, mixtures are separated, and stereochemistry is removed as Mass Spectrometry detects the major components of a salt or mixture and is insensitive to stereochemistry. As an example, a search for the monoisotopic mass of phenol will return phenol, sodium phenolate and calcium phenoxide. See the publication for more details: <https://doi.org/10.1186/s13321-018-0299-2>.

Searching batches using MS-Ready Formula (or mass) searching

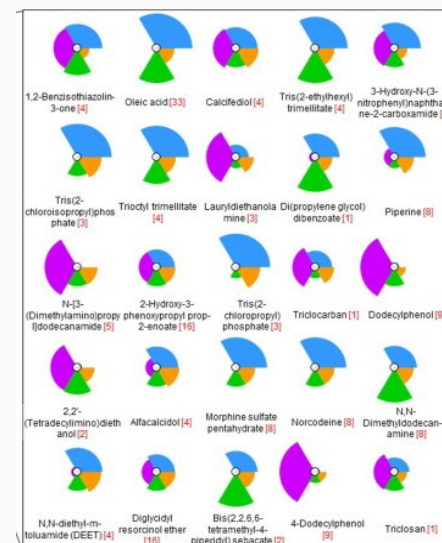
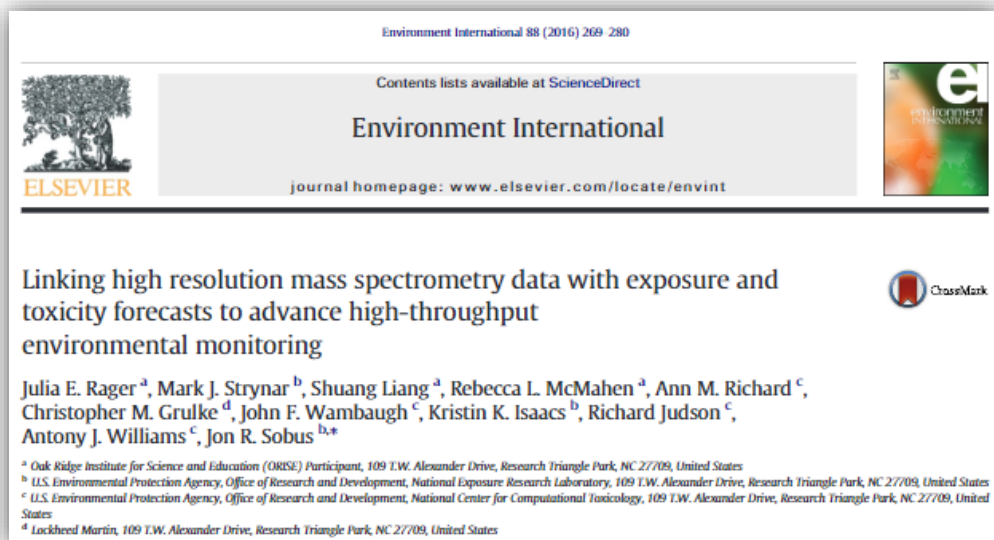
	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3	338.116398	19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23ClN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S	406.213757997	35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35ClN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM 71432748	C18H35ClN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O	176.094963014	40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13ClN2O	212.0716407	11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxy pyrimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026	13140-86-8	N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3	290.137890456	68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3	290.137890456	51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19ClN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2-	C14H18N4O3	290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quinaz	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxyphenyl	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid-6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6025567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	50

Batch Search in specific lists

<input type="checkbox"/>	INPUT	DTXSID	MASSBANKREF	NEMILIST	WRTMSD	NORMANPRI	SUSDAT
<input type="checkbox"/>	Buprenorph	DTXSID202	-	-	Y	-	Y
<input type="checkbox"/>	Codeine	DTXSID202	Y	Y	Y	Y	Y
<input type="checkbox"/>	Dextrometh	DTXSID302	Y	Y	Y	-	Y
<input type="checkbox"/>	Dihydrocod	DTXSID502	Y	-	Y	Y	Y
<input type="checkbox"/>	Dihydromor	DTXSID704	-	-	-	-	Y
<input type="checkbox"/>	Ethylmorph	DTXSID104	-	-	Y	-	Y
<input type="checkbox"/>	Fentanyl	DTXSID902	Y	-	Y	-	Y
<input checked="" type="checkbox"/>	Heroin	DTXSID604	Y	-	Y	Y	Y
<input checked="" type="checkbox"/>	Hydrocodor	DTXSID802	Y	Y	Y	Y	Y
<input type="checkbox"/>	Hydromorph	DTXSID802	-	-	Y	-	Y
<input type="checkbox"/>	Ketamine	DTXSID802	Y	-	Y	-	Y
<input checked="" type="checkbox"/>	Meperidine	DTXSID902	Y	-	Y	-	Y
<input type="checkbox"/>	Methadone	DTXSID702	Y	Y	Y	-	Y
<input checked="" type="checkbox"/>	Morphine	DTXSID902	Y	Y	Y	Y	Y
<input type="checkbox"/>	Morphinone	DTXSID501	-	-	-	-	Y
<input checked="" type="checkbox"/>	Naloxone	DTXSID802	-	-	Y	-	Y
<input type="checkbox"/>	Naltriben	-	-	-	-	-	-
<input type="checkbox"/>	Oxycodone	DTXSID502	Y	Y	Y	Y	Y
<input type="checkbox"/>	Oxymorpho	DTXSID502	-	-	Y	-	Y
<input type="checkbox"/>	Propoxyph	DTXSID102	Y	Y	Y	-	Y
<input type="checkbox"/>	Sufentanil	DTXSID602	-	-	Y	-	Y
<input type="checkbox"/>	Tramadol	DTXSID908	Y	Y	Y	Y	Y

Benefits of bringing it all together

- The true dashboard benefit is integration
- Rank potential candidates for toxicity using available data – hazard, exposure, *in vitro*



Candidate ranking using metadata



© American Society for Mass Spectrometry, 2011

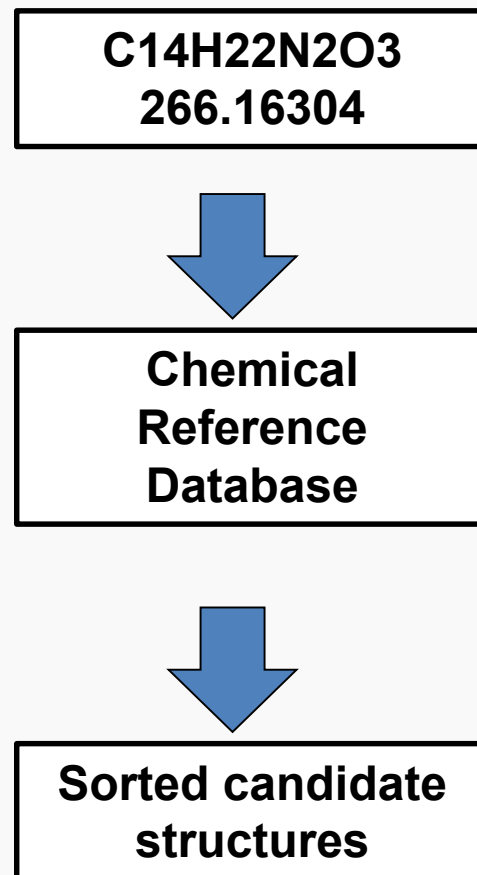
J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

RESEARCH ARTICLE



**Identification of “Known Unknowns” Utilizing
Accurate Mass Data and ChemSpider**

Data Source Ranking of “*known unknowns*”

- A mass and/or formula search is for an ***unknown*** chemical but it is a ***known*** chemical contained within a reference database
- **Most likely** candidate chemicals have the **most** associated data sources, **most** associated literature articles or both



Data Streams for Ranking

- CompTox Dashboard Data Sources
-  Data Source Count
-  Reference Count
- Toxcast *in vitro* bioactivity
- Presence in CPDat database
- OPERA PhysChem Properties
- Other possibilities – predicted media occurrence, frequency of InChIs online

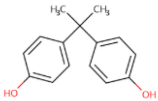
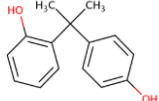
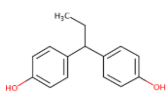
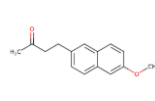
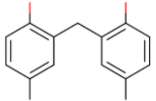
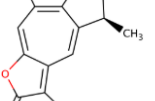
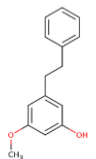
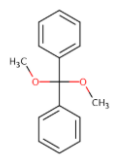
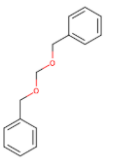
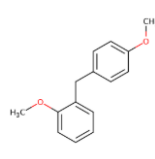
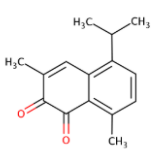
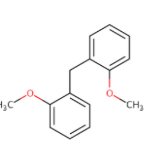
Search 228.115 +/- 5.0 ppm

234 single component chemicals

Search Results
Searched by Mass: 228.115 +/- 5.0 ppm.

234 of 247 chemicals visible

Select all Download Send to Batch Search Mass Difference CASRN DTXSID Mass Diff Multicomponent Chemicals Filter by Name or CASRN

 <p>Bisphenol A CASRN:80-05-7 DTXSID:DTXSID7020182 Mass Diff:0.00003</p>	 <p>Phenol, 2-[1-(4-hydroxyphenyl)-1-methyl-2-propenyl]-4-hydroxy CASRN:837-08-1 DTXSID:DTXSID7042275 Mass Diff:0.00003</p>	 <p>4,4'-Propane-1,1-diylidiphenol CASRN:1576-13-2 DTXSID:DTXSID3044594 Mass Diff:0.00003</p>	 <p>Nabumetone CASRN:42924-53-8 DTXSID:DTXSID4045472 Mass Diff:0.00003</p>	 <p>Phenol, 2,2'-methylenbis[4-methyl-5-hydroxy CASRN:3236-63-3 DTXSID:DTXSID0062923 Mass Diff:0.00003</p>	 <p>Gweicurculactone CASRN:123914-43-2 DTXSID:DTXSID50154143 Mass Diff:0.00003</p>
 <p>Phenol, 3-methoxy-5-(2-phenylethyl)- CASRN:17635-59-5</p>	 <p>1,1'-(Dimethoxymethylene)bisbenzene CASRN:2235-01-0</p>	 <p>Preventol D2 CASRN:2749-70-4</p>	 <p>1-Methoxy-2-((4-methoxyphenyl)methyl)-4-methoxybenzene CASRN:30567-87-4</p>	 <p>1,2-Naphthalenedione, 3,8-dimethyl-5-(2-methylpropyl)- CASRN:5574-34-5</p>	 <p>Methane, bis-(o-methoxyphenyl)- CASRN:5819-93-2</p>

Search 228.115 +/- 5.0 ppm

234 single component chemicals

CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts
80-05-7	Level 1	326	170	161	3850
42924-53-8	Level 2	14	45	138	342
87619-52-1	Level 5	0	2		0
87607-32-7	Level 5	0	2		0

The original ChemSpider work

Compound class	Number in class	Average rank	Number of compounds in each position rank-ordered				
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Steroid hormones	7	1.0	7				
Perfluorochemicals	6	1.2	5	1			
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 ^a	1.3	2	1			

Is a bigger database better?

- ChemSpider was 26 million chemicals for the original work
- Much BIGGER today
- Is bigger better??
- Are there other metadata to use for ranking?



81 Million
chemical structures

³
Anal Bioanal Chem (2017) 409:1729–1735
DOI 10.1007/s00216-016-0139-z



RAPID COMMUNICATION

Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard

Andrew D. McEachran¹ · Jon R. Sobus² · Antony J. Williams³

- When dashboard contained 720k chemicals
- Only **3%** of ChemSpider size
- What was the comparison in performance?

SAME dataset for comparison

Compound class	Number in class	Average rank	Number of compounds in each position rank-ordered				
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Steroid hormones	7	1.0	7				
Perfluorochemicals	6	1.2	5	1			
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 ^a	1.3	2	1			

EXACTLY THE SAME DATASET

How did performance compare?

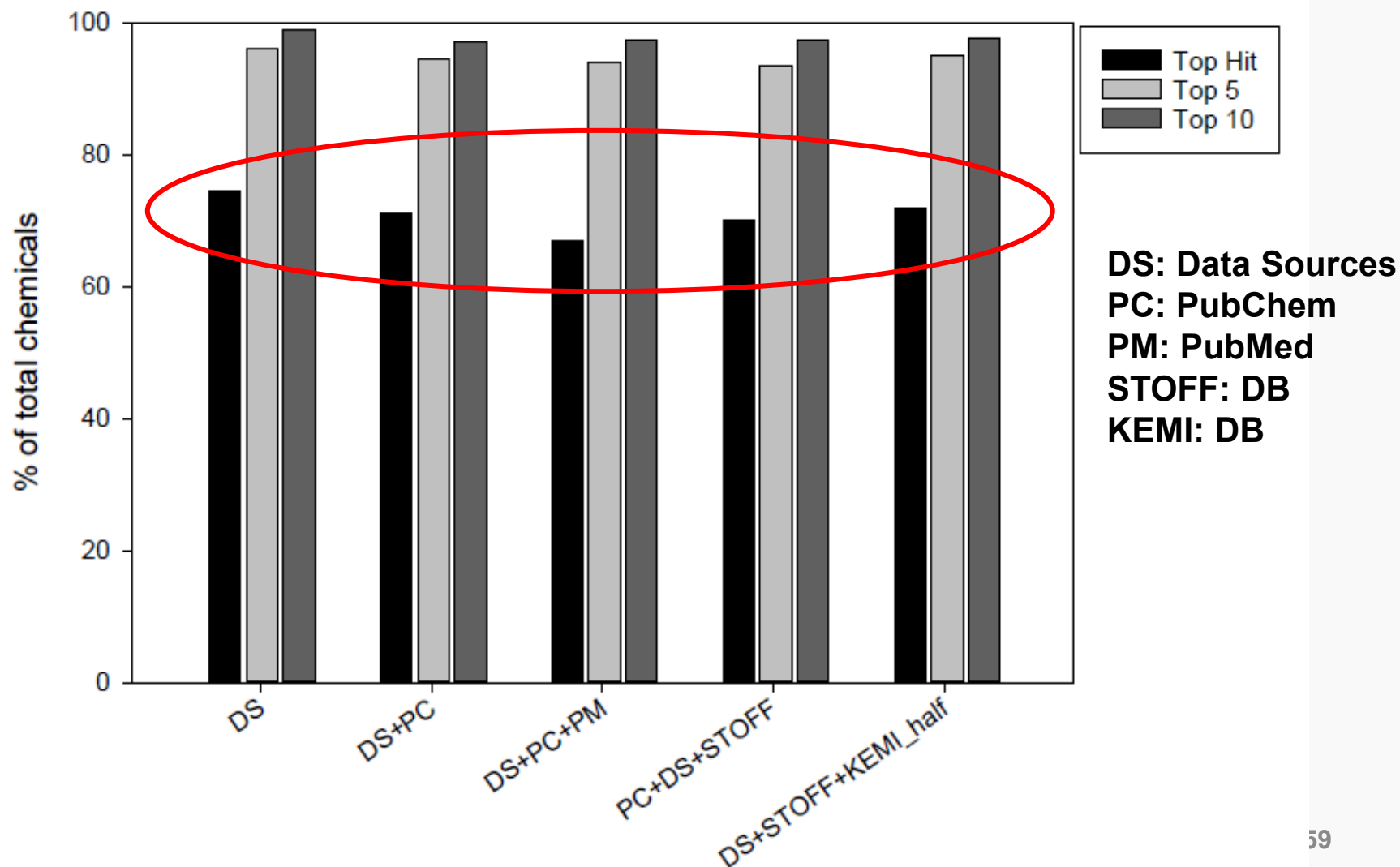
	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2 ^a	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

^a Average rank in ChemSpider shown here does not include an outlier where the rank was 201, when added the average rank position is 3.5

**For the same 162 chemicals,
Dashboard outperforms
ChemSpider for both Mass and
Formula Ranking**

Identification ranks for 1783 chemicals using multiple data streams

**Data Sources alone
rank ~75% of the
chemicals as Top Hit**



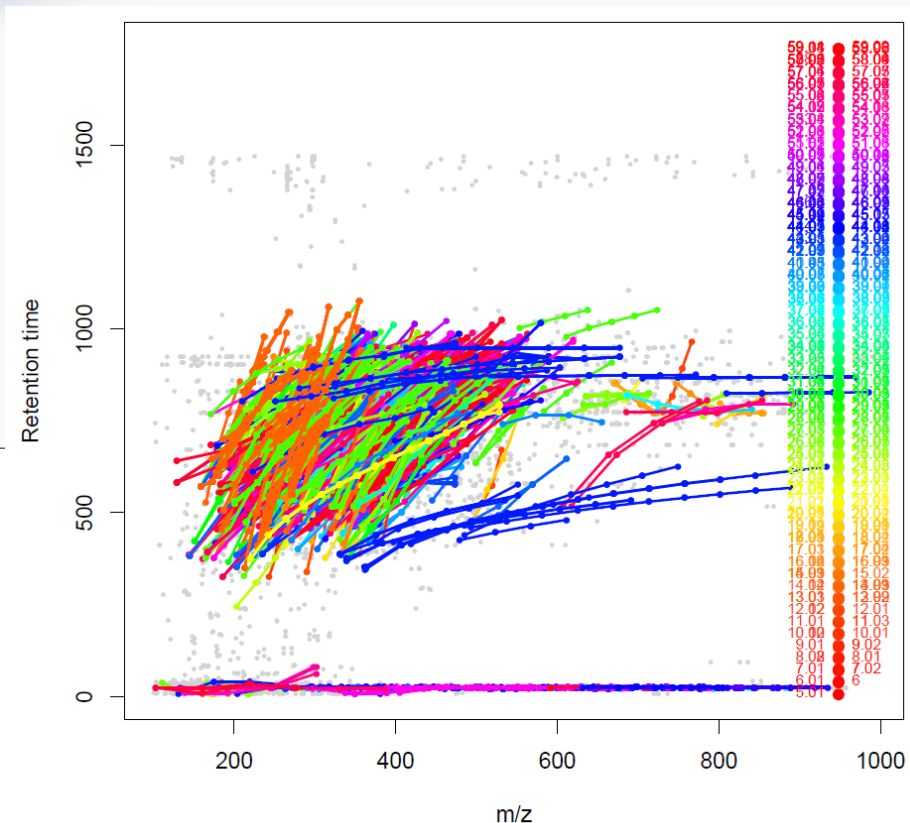
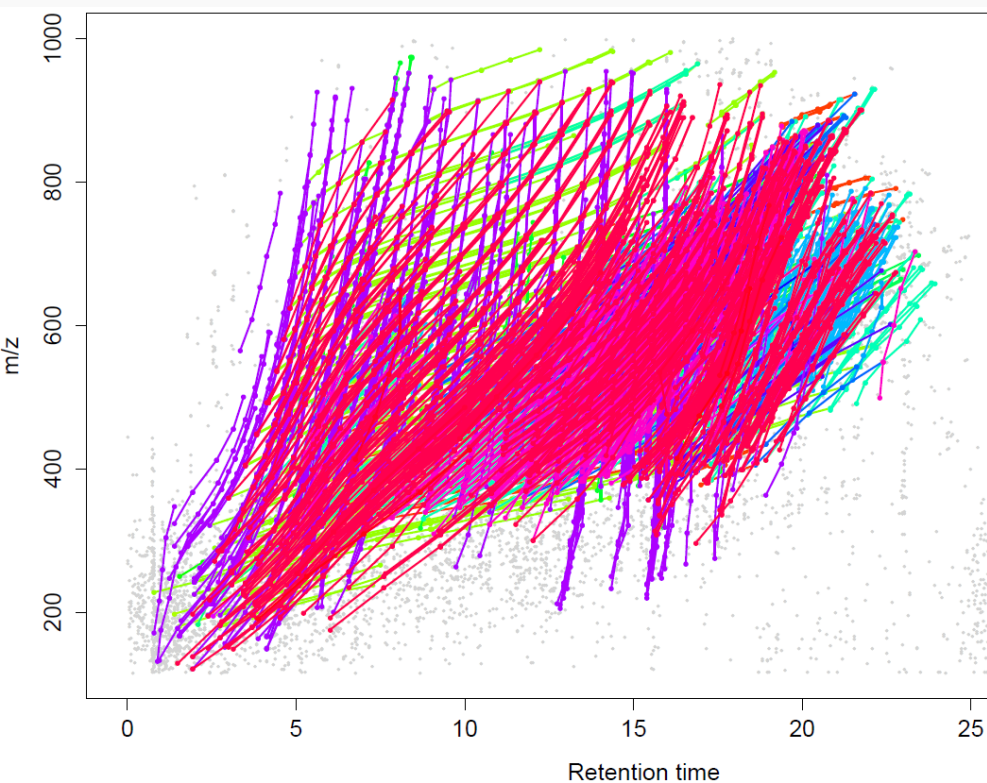
“UVCB” Chemicals

Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB Substance) on the TSCA Inventory

This paper is a compendium of information related to the broad class of chemical substances referred to as UVCBs for the Toxic Substances Control Act (TSCA) Chemical Substance Inventory. These chemical substances cannot be represented by unique structures and molecular formulas.

UVCBs challenge in non-target analysis

- Complex mixtures (UVCBs) are a **huge** and **very challenging** part of the unknowns in many environmental samples



Homologue screening plots from
Swiss Wastewater (Schymanski *et al*
2014, left) and Novi Sad (right)

solutions

Public TSCA Inventory on Dashboard

31,460 Chemicals (1/24/2020)

EPA|TSCA: TSCA Inventory, active non-confidential portion

☐ Identifier substring search

List Details

Description: Section 8 (b) of the Toxic Substances Control Act (TSCA) requires EPA to compile, keep current and publish a list of each chemical substance that is manufactured or processed, including imports, in the United States for uses under TSCA. Information about what types of substances are on the TSCA inventory can be found here. The Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, requires EPA to designate chemical substances on the TSCA Chemical Substance Inventory as either "active" or "inactive" in U.S. commerce. To accomplish this, EPA finalized a rule requiring industry reporting of chemicals manufactured (including imported) or processed in the U.S.. This reporting is used to identify which chemical substances on the TSCA Inventory are active in U.S. commerce and help inform the prioritization of chemicals for risk evaluation. The list contained in the dashboard includes the active TSCA inventory based on notifications through Feb. 7th 2018 and substances reported from Feb 8, 2018 – March 30, 2018 that have been unambiguously mapped to DSSTox using CASRN and chemical names. The curation of the non-confidential portion of active TSCA inventory is an ongoing process involving trained chemists to validate the correctness of DSSTox structural and identifier data. The content of the list will change over time as the non-confidential active TSCA inventory is updated and more substances are curated. (Updated January 5th 2020)

Number of Chemicals: 31460

2250 of 31460 chemicals loaded

Select all

Download

Send to Batch Search

Default

↑

CASRN

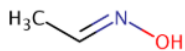
DTXSID

▼

Hide chemicals that are:

Filter by Name or CASRN

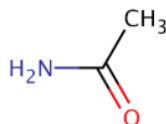
≡



Acetaldehyde oxime

CASRN:107-29-9

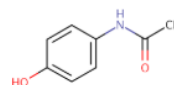
DTXSID:DTXSID2020004



Acetamide

CASRN:60-35-5

DTXSID:DTXSID7020005



Acetaminophen

CASRN:103-90-2

DTXSID:DTXSID2020006



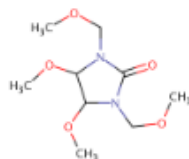
Acetonitrile

CASRN:75-05-8

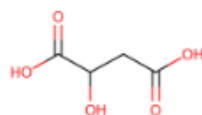
DTXSID:DTXSID7020009

Many Chemicals are “Complex”

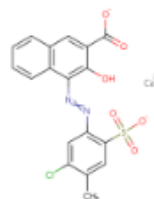
>14000 chemicals are UVCBs



2-Imidazolidinone, 4,5-dimethoxy-1,3-bis(methoxymethyl)-
DTXSID: DTXSID0027569
PubChem: 24
CASRN: 4356-80-9



Malic acid
DTXSID: DTXSID0027640
PubChem: 273
CASRN: 6915-15-7



C.I. Pigment Red 48, calcium salt (1:1)
DTXSID: DTXSID0027642
PubChem: 0
CASRN: 7023-81-2

0 related chemical
structures with this
substance

Lard, oil
DTXSID: DTXSID0027690
PubChem: 0
CASRN: 8016-28-2

0 related chemical
structures with this
substance

Tall-oil pitch
DTXSID: DTXSID0027692
PubChem: 0
CASRN: 8016-81-7

0 related chemical
structures with this
substance

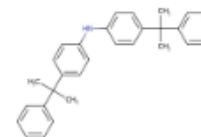
Palm kernel oil
DTXSID: DTXSID0027694
PubChem: 0
CASRN: 8023-79-8

0 related chemical
structures with this
substance

Tallow, hydrogenated
DTXSID: DTXSID0027696
PubChem: 0
CASRN: 8030-12-4

0 related chemical
structures with this
substance

Quaternary ammonium compounds, tri...
DTXSID: DTXSID0027698
PubChem: 0
CASRN: 8030-78-2



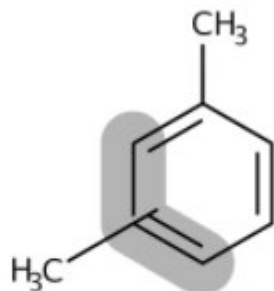
4-(2-Phenylpropan-2-yl)-N-[4-(2-phenylpropan-2-yl)phenyl]benzenesulfonamide
DTXSID: DTXSID0027721
PubChem: 50
CASRN: 10081-67-1

1 related chemical
structure with this
substance

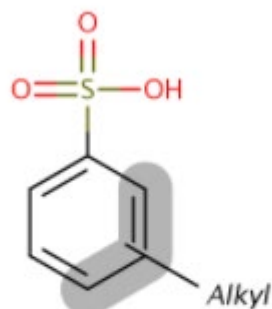
Isomethyltetrahydrophthalic anhydride
DTXSID: DTXSID0027729
PubChem: 0
CASRN: 11070-44-3

“Markush Structures”

https://en.wikipedia.org/wiki/Markush_structure



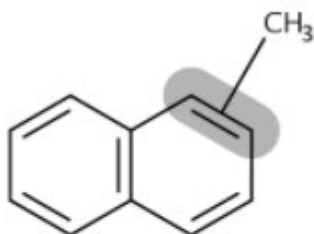
Xylenes
1330-20-7



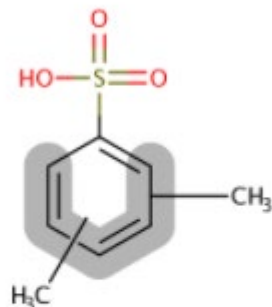
(C10-C16) Alkylbenzenesulfonic acid
68584-22-5



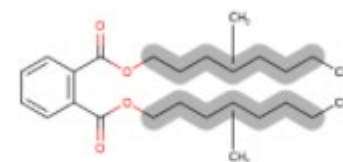
n-Nonylphenol
25154-52-3



Methyl-naphthalene
1321-94-4

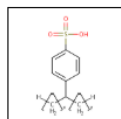


Sodium xylenesulfonate
1300-72-7



Diisononyl phthalate
28553-12-0

How to represent complexity?



Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

Searched by DSSTox Substance Id.

15 of 25 chemicals selected

Deselect all

Download

Send to Batch Search

Relationship

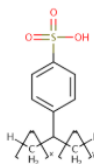
CASRN

DTXSID

Unselected

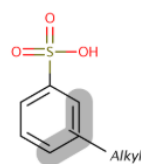
Filter by Name or CASRN

Searched Chemical



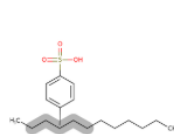
Alkylbenzenesulfonate, linear
CASRN:42615-29-2
DTXSID:DTXSID3020041

Component



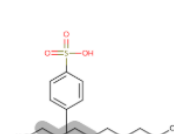
(C10-C16) Alkylbenzenesulfonic acid
CASRN:68584-22-5
DTXSID:DTXSID2028723

Component



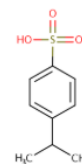
C12-linear alkyl benzene sulfonate
CASRN:NOCAS_891641
DTXSID:DTXSID90891641

Component



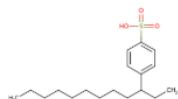
C10-linear alkylbenzenesulfonate
CASRN:NOCAS_891689
DTXSID:DTXSID70891689

Markush Child



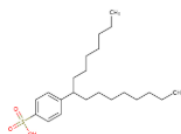
4-Isopropylbenzenesulfonic acid
CASRN:16066-35-6
DTXSID:DTXSID1044932

Markush Child



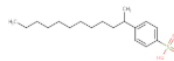
4-(3-Dodecyl)benzenesulfonic acid
CASRN:18777-54-3
DTXSID:DTXSID7058670

Markush Child



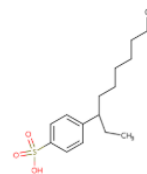
4-(1-Heptylnonyl)benzenesulfonic acid
CASRN:80233-94-9
DTXSID:DTXSID40273953

Markush Child



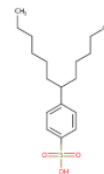
2-Phenyldodecane-p-sulfonate
CASRN:18777-53-2
DTXSID:DTXSID40274021

Markush Child



4-(Decan-3-yl)benzene-1-sulfonic acid
CASRN:65186-00-7
DTXSID:DTXSID20859618

Markush Child



4-(Dodecan-6-yl)benzene-1-sulfonic acid
CASRN:23003-92-1
DTXSID:DTXSID30860093

Work in Progress

- Registering and curating numerous lists
 - NIST library of chemicals –clean up especially around stereochemical representation
 - United States Geological Survey chemicals in water
 - Scientific Working Group for the Analysis of Seized Drugs
 - Synthetic Cannabinoids
 - Blood Exposome Database

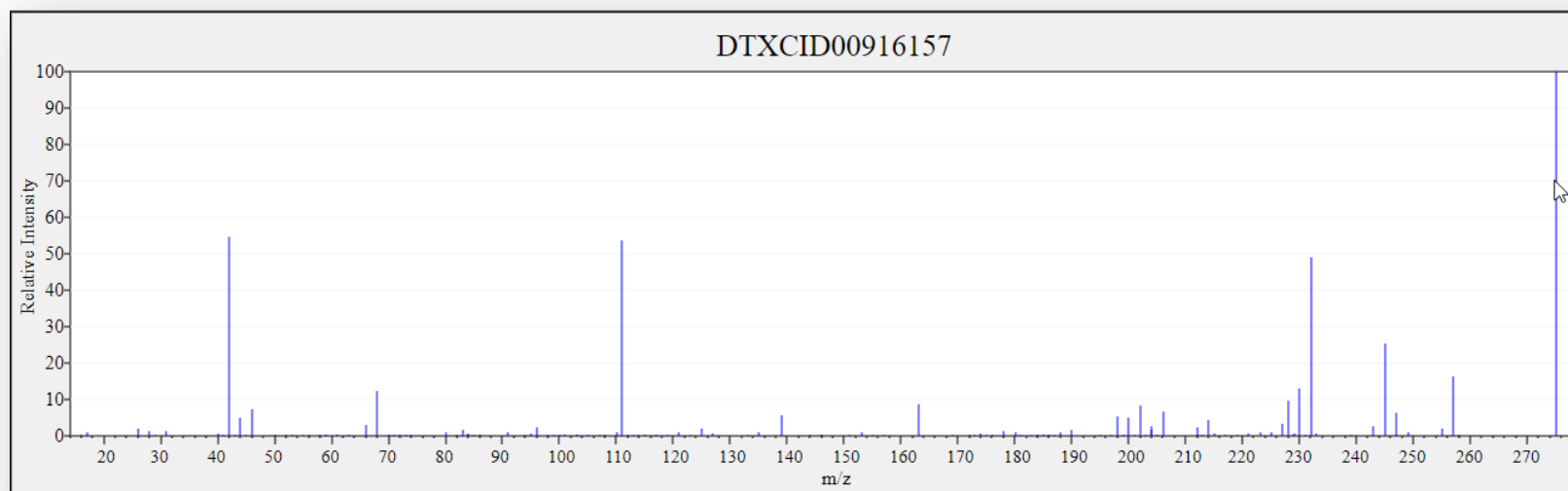
- CFM-ID
 - Viewing and Downloading pre-predicted spectra
 - Search spectra against the database
- Structure/substructure/similarity search
- Access to API and web services
- Integration to EPA “Chemical Transformation Simulator”

Predicted Mass Spectra

<http://cfmid.wishartlab.com/>



- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra



Non Target Analysis Prototype

Mass Search

±

Min/Max

321.136493476

Da

±

0.0000002

Da

ppm

Molecular Formula Search

Molecular Formula

Mass or Formula must be entered before searching spectrum

Ionization Type

ESI+ ▼

ESI+

ESI-

EI

Spectra Input

Single Energy

Multiple

304.1332052 11.6199475
198.0913404 7.306439699
123.0440559 6.538348292
196.0756904 5.269463115
216.1019051 4.700461978
300.1080005 4.800442384

Peak Match Window:

0.02

Da

ppm

Search

Search Expt. vs. Predicted Spectra



United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

Share

Spectra Input

Single Ener

304.1332052 11.61
198.0913404 7.30
123.0440559 6.53
196.0756904 5.28
216.1019051 4.70
200.1089005 4.80

Peak Match

Search

TSVCSVExcel

Chemical Structure ID

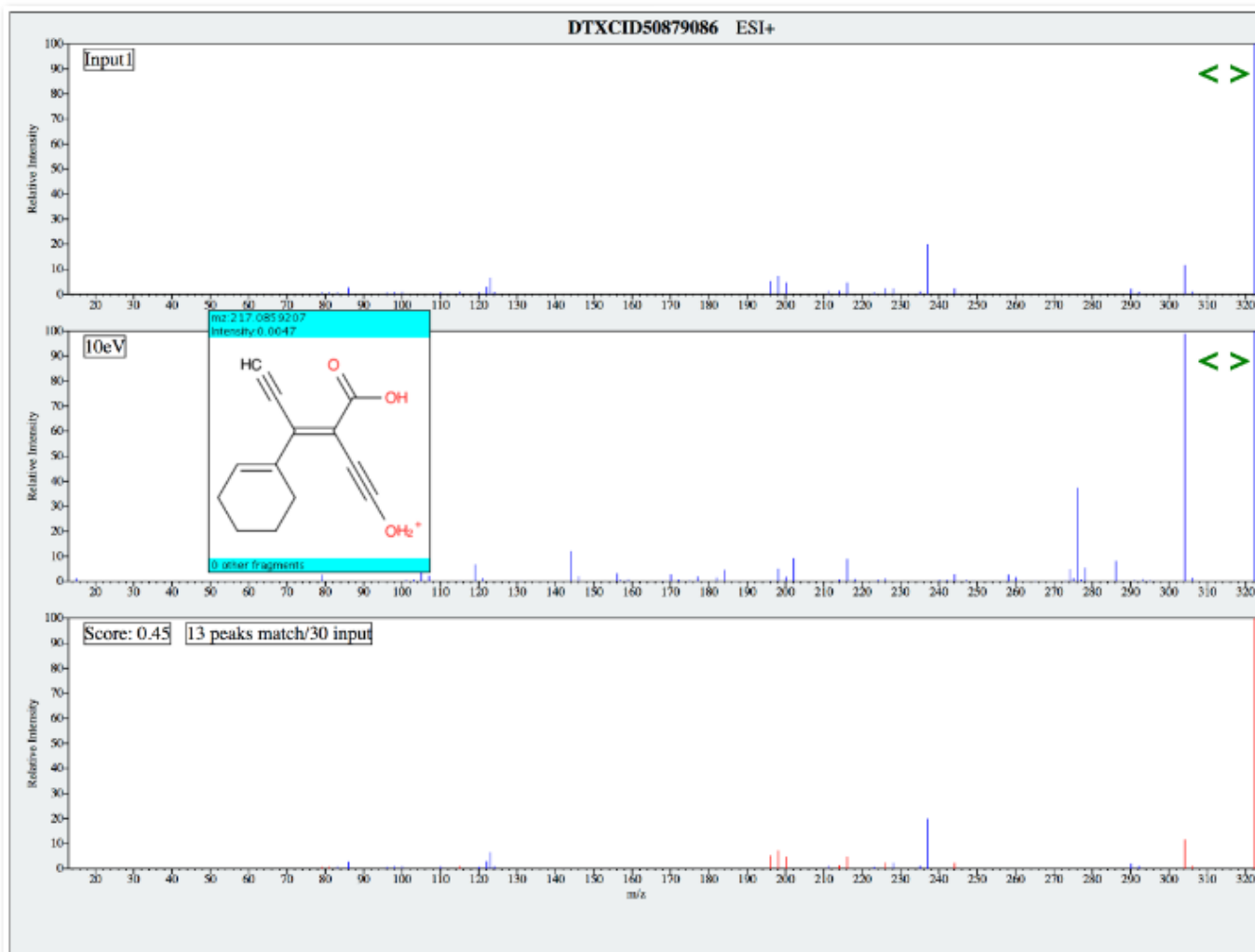
DTXCID101048191
DTXCID101181567
DTXCID50879086
DTXCID60686349
DTXCID00830900
DTXCID10971176
DTXCID60301242
DTXCID40703048
DTXCID60349982
DTXCID10316649

Showing 1 to 10 of 38 entries

Chemical Structure ID	Score (10eV)
DTXCID101048191	0.22
DTXCID101181567	0.19
DTXCID50879086	0.17
DTXCID60686349	0.14
DTXCID00830900	0.13
DTXCID10971176	0.12
DTXCID60301242	0.12
DTXCID40703048	0.11
DTXCID60349982	0.11
DTXCID10316649	0.09

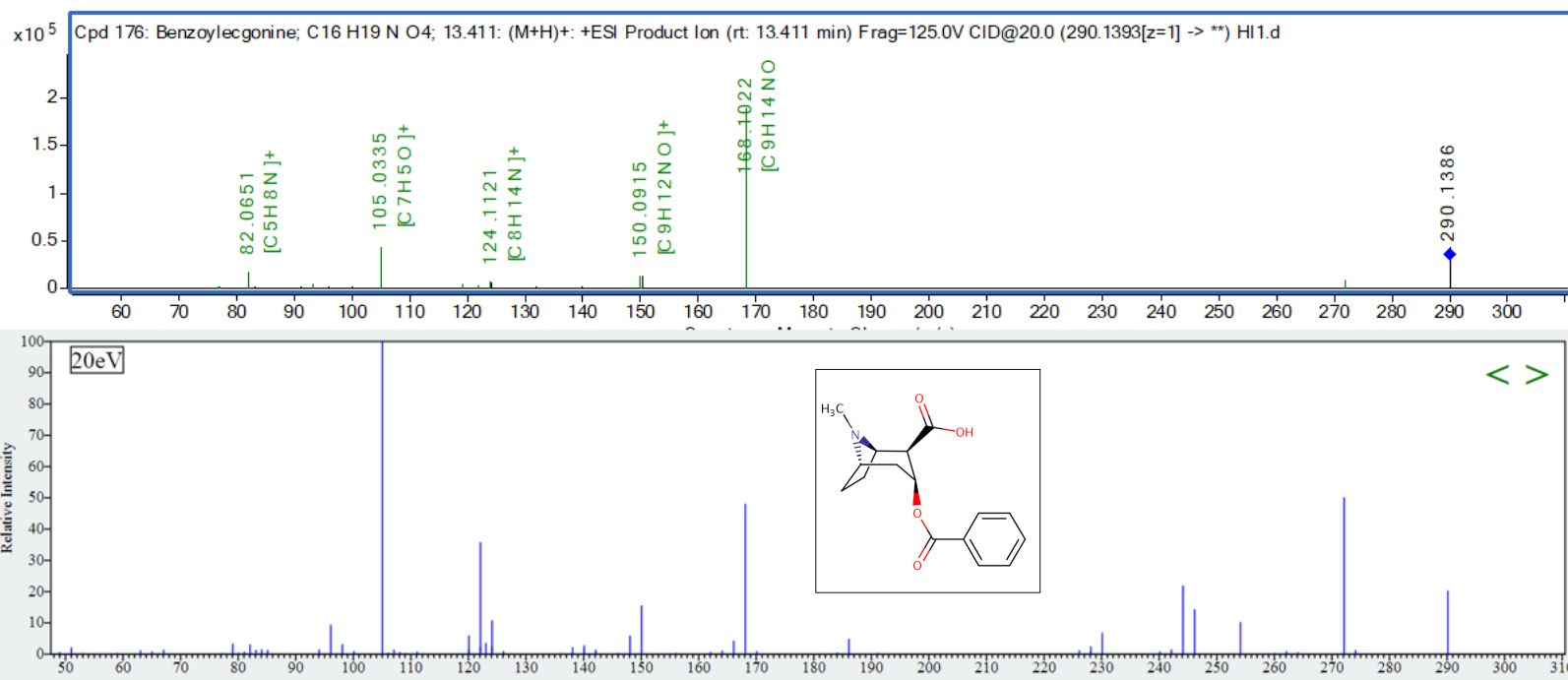
1234Next

Spectral Viewer Comparison



Example match



MS/MS of benzoylecgonine




Predicted Data Already Public *Publication and Data Files*

Data Descriptor | OPEN | Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

Andrew D. McEachran , Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams 

Scientific Data **6**, Article number: 141 (2019) | [Download Citation](#) 

CFM-ID Paper Data

[Dataset](#) posted on 01.03.2019, 08:38 by EPA's National Center for Computational Toxicology

88
views

17
downloads

0
citations

This upload is a zip containing the following files:

Predicted EI-MS Spectra of CompTox Chemicals Dashboard Structures:

Predicted EI-MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (<https://doi.org/10.1021/acs.analchem.6b01622>). These data are provided in .dat ASCII format.

Predicted MS/MS Spectra in ESI-positive mode of CompTox Chemicals Dashboard Structures:

Predicted MS/MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (<https://doi.org/10.1007/s11306-014-0676-4>) in ESI-positive mode. These data are provided in .dat ASCII format.

Predicted MS/MS Spectra in ESI-negative mode of CompTox Chemicals Dashboard Structures:

Predicted MS/MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (<https://doi.org/10.1007/s11306-014-0676-4>) in ESI-negative mode. These data are provided in .dat ASCII format.



CATEGORIES

• Toxicology

KEYWORD(S)

Computational Toxicology

DSSTox Chemical Database

Chemicals Dashboard

Non-targeted analysis

CFM-ID

LICENCE



EXPORT

RefWorks

BibTeX

Ref. manager

Endnote

https://epa.figshare.com/articles/CFM-ID_Paper_Data/7776212/1

Analytical and Bioanalytical Chemistry

RESEARCH PAPER

In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples

Alex Chao^{1,2} • Hussein Al-Ghoul^{1,2} • Andrew D. McEachran^{1,3} • Ilya Balabin⁴ • Tom Transue⁴ • Tommy Cathey⁴ • Jarod N. Grossman^{2,3} • Randolph Singh^{1,5} • Elin M. Ulrich² • Antony J. Williams⁶ • Jon R. Sobus²

Received: 4 October 2019 / Revised: 27 November 2019 / Accepted: 11 December 2019

© The Author(s) 2019

Prototype Development

AADashboard

atrazine

Search



Select properties to predict

H

C

N

O

S

P

F

Cl

T.E.S.T. 18

OPERA

Search

Exact

Substructure

Similarity

Molecular Formula

Molecular Weight

Filter by elements (enter comma separated list e.g. C,F,H) include

Search result 2540

Show

☐ Isotopically Labeled

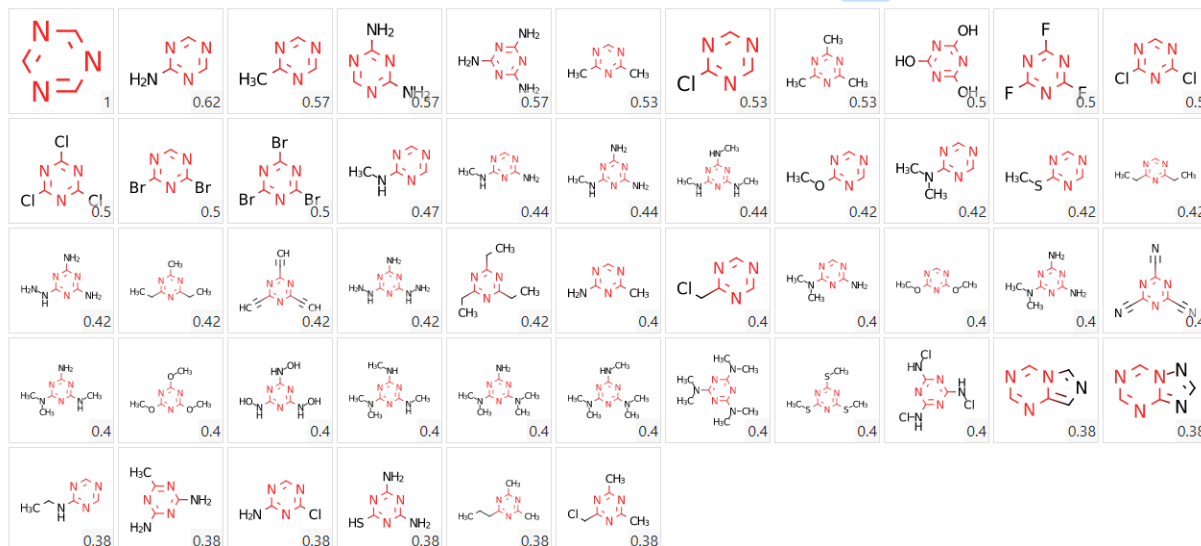
☐ Charged

☐ Salts or Mixtures

Sort

Similarity

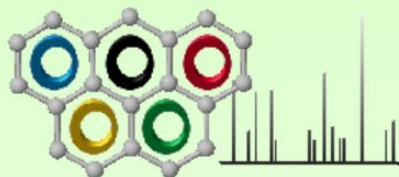
≡



Search result 2540

Show

☐ Isotopically Labeled



CASMI

Critical Assessment of Small Molecule Identification

The experimental and computational mass spectrometry communities are invited to participate in the fifth round of an open contest on the identification of small molecules from mass spectrometry data.

This year the contest will test the applicability of MS and MS/MS on natural products chemistry identifications. With 45 (Category 1) and up to 243 (Categories 2&3) natural products challenges - including a few tricky ones - there's something for everyone!

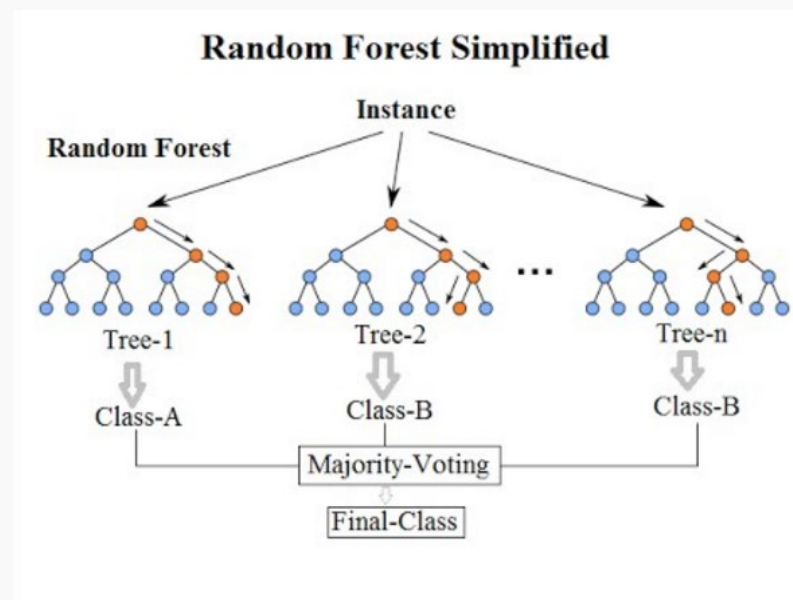
- Application of metadata candidate ranking and CFM-ID to all five years of CASMI data

Method Amenability Prediction

Charlie Lowe

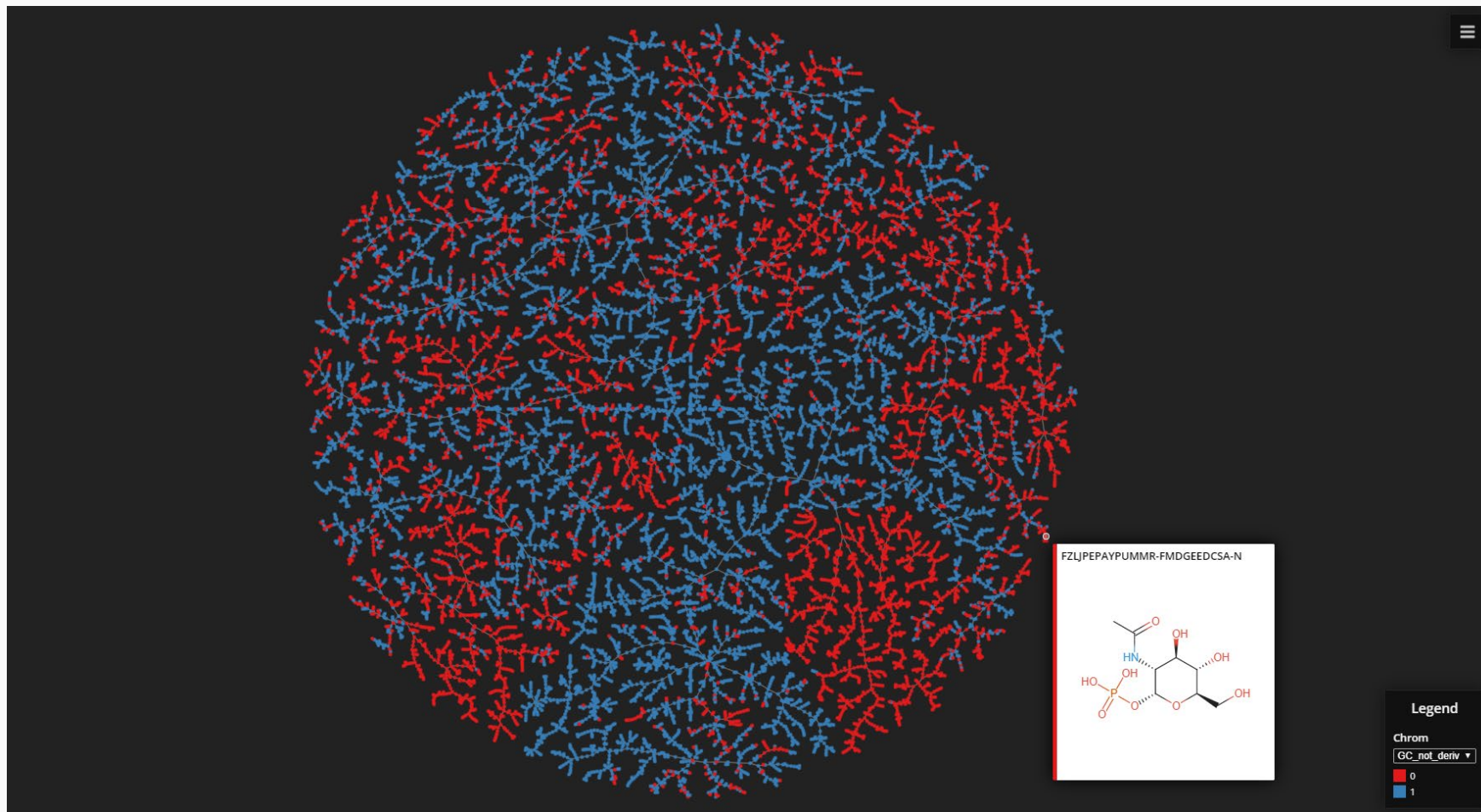
Why?

- Chromatography-mass spectrometry can be LC or GC
- Which phase is more appropriate for which chemicals?



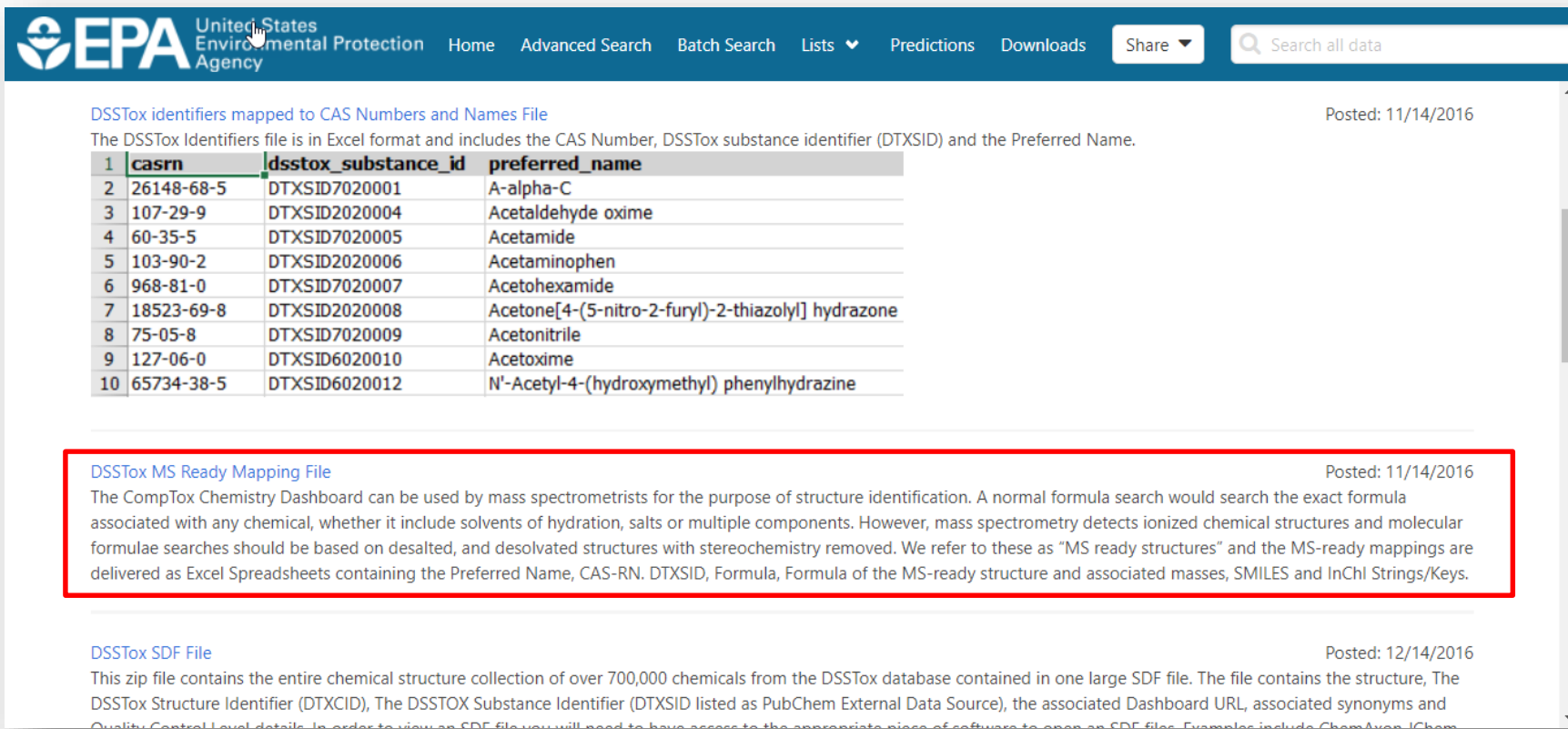
- Data sources to date
 - **Massbank of North America**
 - 9,275 chemicals for non-derivatized GC
 - 846 chemicals for derivatized GC
 - 816 chemicals for APCI+
 - 454 chemicals for APCI-
 - 4,907 chemicals for ESI+
 - 3,430 chemicals for ESI-
 - **EPA Non-targeted Analysis Collaborative Trial (ENTACT)**
 - 886 chemicals for non-derivatized GC
 - 44 chemicals for derivatized GC
 - 774 chemicals for APCI+
 - 431 chemicals for APCI-
 - 1,113 chemicals for ESI+
 - 648 chemicals for ESI-

TMAP Visualization of MoNA GC Data



API services and Open Data

- Web Services <https://actorws.epa.gov/actorws/>
- Data sets also available for download..



The screenshot shows the EPA ACTORWS website interface. At the top is the EPA logo and navigation links: Home, Advanced Search, Batch Search, Lists, Predictions, Downloads, and a Share button. A search bar is on the right. Below the navigation bar, there are two main sections. The first section is titled "DSSTox identifiers mapped to CAS Numbers and Names File" and is dated "Posted: 11/14/2016". It includes a description: "The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name." Below this is a table with 10 rows of data. The second section is titled "DSSTox MS Ready Mapping File" and is also dated "Posted: 11/14/2016". It includes a detailed description of the file's purpose and content. The third section is titled "DSSTox SDF File" and is dated "Posted: 12/14/2016". It includes a description of the file's contents.

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.

	casrn	dsstox_substance_id	preferred_name
2	26148-68-5	DTXSID7020001	A-alpha-C
3	107-29-9	DTXSID2020004	Acetaldehyde oxime
4	60-35-5	DTXSID7020005	Acetamide
5	103-90-2	DTXSID2020006	Acetaminophen
6	968-81-0	DTXSID7020007	Acetohexamide
7	18523-69-8	DTXSID2020008	Acetone[4-(5-nitro-2-furyl)-2-thiazolyl] hydrazone
8	75-05-8	DTXSID7020009	Acetonitrile
9	127-06-0	DTXSID6020010	Acetoxime
10	65734-38-5	DTXSID6020012	N'-Acetyl-4-(hydroxymethyl) phenylhydrazine

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.

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 file. Examples include ChemAxon JChem

<https://actorws.epa.gov/actorws/>

- Data in UI, JSON and XML format
- Our services are free of course..

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=80-05-7>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=80-05-7>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=80-05-7>

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=DTXCID60513>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=DTXCID60513>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=DTXCID60513>

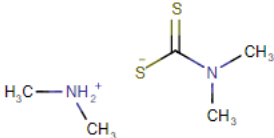
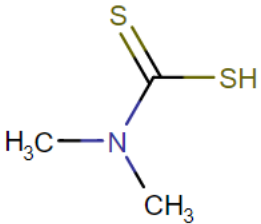
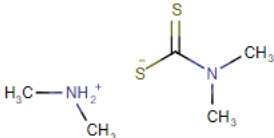
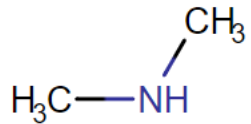
<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

InChIKey to DTXCIDs

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

Image	DTXCID	Smiles	Image	MsReady DTXCID	MsReady SMILES
	DTXCID60513	<chem>C[NH2+]C.CN(C)C([S-])=S</chem>		DTXCID0023797	<chem>CN(C)C(S)=S</chem>
	DTXCID60513	<chem>C[NH2+]C.CN(C)C([S-])=S</chem>		DTXCID704057	<chem>CNC</chem>

Data and Services used by the Community

NORMAN Suspect List Exchange

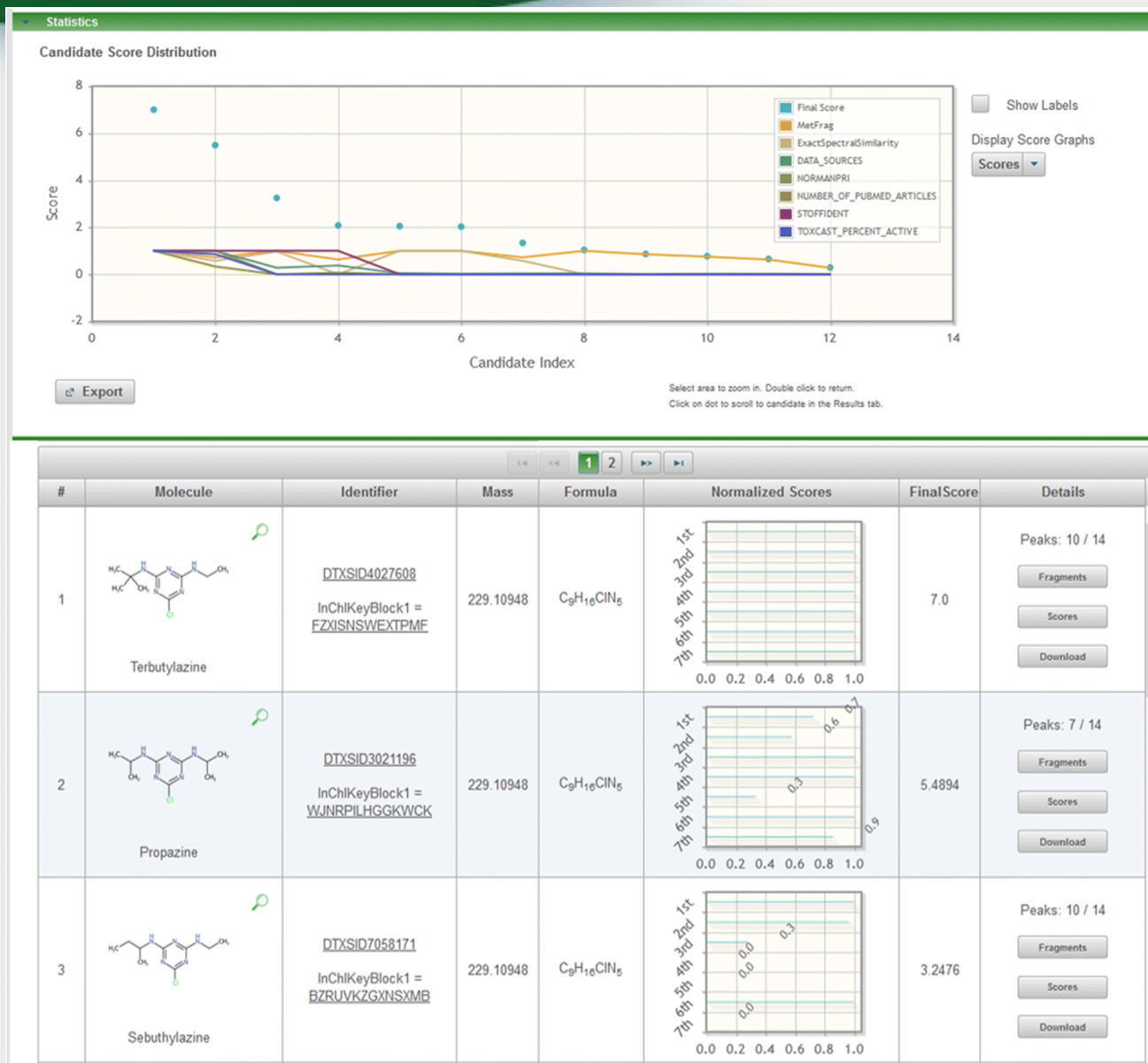
<https://www.norman-network.com/?q=node/236>



Wastewater Suspect List based on Swedish Product Data	Wastewater Suspect List Original File with Mapped DTXSIDs (12/02/2019)	KEMIWSUS InChIKeys (12/02/2019)	A prioritized list of 1,123 substances relevant for wastewater based on Swedish product registry data, including scores. Provided by Stellan Fischer, KEMI.
Algal toxins list from CompTox	ALGALTOX XLSX, CSV (14/02/2019) CompTox ALGALTOX List	ALGALTOX InChIKeys (14/02/2019)	List of algal toxins (generated during blooms) from the CompTox Chemicals Dashboard.
CCL 4 Chemical Candidate List	CCL4 XLSX, CSV (14/02/2019) CompTox CCL4 List	CCL4 InChIKeys (14/02/2019)	Contaminants that are not (yet) regulated in the USA but are known or anticipated to occur in public water systems; from CompTox.
Hydrogen Deuterium Exchange (HDX) Standard Set	HDXNOEX XLSX, CSV (14/02/2019) CompTox HDXNOEX List CompTox HDXEXCH List	HDXNOEX InChIKeys (14/02/2019)	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule HRMS (Ruttkies et al. submitted). HDXEXCH list also contains observed deuterated species.
Neurotoxicants Collection from Public Resources	NEUROTOXINS XLSX, CSV (14/02/2019) CompTox NEUROTOXINS List	NEUROTOXINS InChIKeys (14/02/2019)	A list of neurotoxicants compiled from public resources, details on CompTox and Schymanski <i>et al.</i> (submitted).
Statins Collection from Public Resources	STATINS XLSX, CSV (14/02/2019) CompTox STATINS List	STATINS InChIKeys (14/02/2019)	A list of statins (lipid-lowering medications) compiled from public resources, details on CompTox.
Synthetic Cannabinoids and Psychoactive Compounds	SYNTHCANNAB XLSX, CSV (14/02/2019) CompTox SYNTHCANNAB List	SYNTHCANNAB InChIKeys (14/02/2019)	A list of synthetic cannabinoids and psychoactive compounds assembled from public resources, from CompTox.

Integration to MetFrag in place

<https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0299-2>



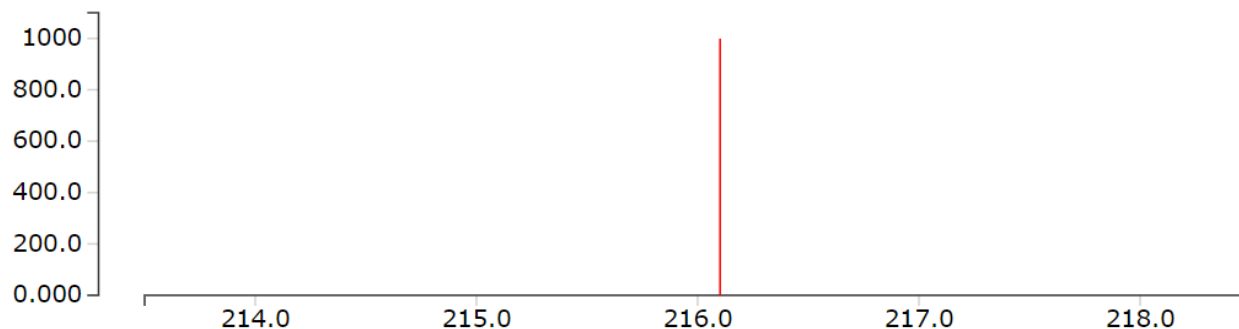
MassBank mapping to Dashboard Based on Web Service lookup

MassBank Record: EA028808

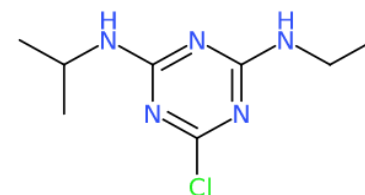
[Home](#) | [Search](#) | [Record Index](#) | [Data Privacy](#) | [Imprint](#) | MassBank ID:

Atrazine; LC-ESI-ITFT; MS2; CE: 15%; R=15000; [M+H]⁺

Mass Spectrum



Chemical Structure

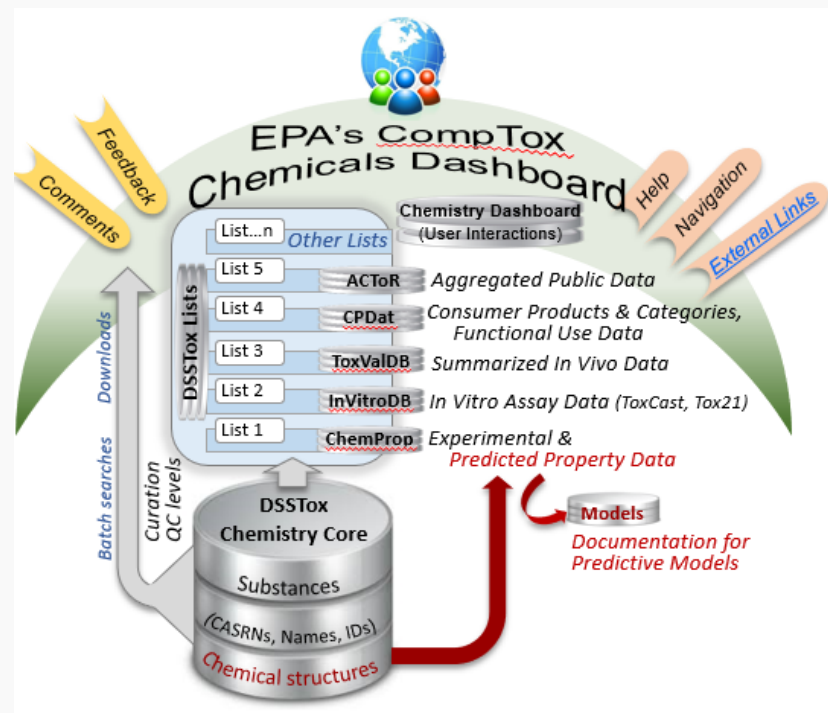


Options
● Labels

CH\$NAME: Atrazine
CH\$NAME: 6-chloro-N-ethyl-N'-isopropyl-1,3,5-triazine-2,4-diamine
CH\$NAME: 6-chloranyl-N4-ethyl-N2-propan-2-yl-1,3,5-triazine-2,4-diamine
CH\$COMPOUND_CLASS: N/A; Environmental Standard
CH\$FORMULA: [C8H14ClN5](#)
CH\$EXACT_MASS: 215.0932
CH\$SMILES: c1(nc(nc(n1)Cl)NCC)NC(C)C
CH\$IUPAC: InChI=1S/C8H14ClN5/c1-4-10-7-12-6(9)13-8(14-7)11-5(2)3/h5H,4H2,1-3H3,(H2,10,11,12,13,14)
CH\$LINK: CAS [1912-24-9](#)
CH\$LINK: CHEBI [15930](#)
CH\$LINK: KEGG [C06551](#)
CH\$LINK: PUBCHEM [CID:2256](#)
CH\$LINK: INCHIKEY [MXWJVTOOROXGIU-UHFFFAOYSA-N](#)
CH\$LINK: CHEMSPIDER [2162](#)
CH\$LINK: COMPTOX [DTXSID9020112](#)

Conclusion

- Dashboard access to data for ~875,000 chemicals (~895k in the Spring Release)
- MS-Ready data facilitates structure identification
- Related metadata facilitates candidate ranking
- Relationship mappings and chemical lists of great utility
- Curation and mutual sharing of chemical lists is important (e.g. NORMAN)



Acknowledgements



EPA ORD

Ann Richard
Chris Grulke
John Wambaugh
Jeremy Dunne
Jeff Edwards
Grace Patlewicz
Alex Chao
Kristin Isaacs
Charles Lowe
James McCord
Seth Newton
Katherine Phillips
Tom Purucker
Jon Sobus
Mark Strynar
Elin Ulrich
Joach Pleil

TEAMS

IT Development Team
Curation Team

ILS

Kamel Mansouri

GDIT

Ilya Balabin
Tom Transue
Tommy Cathey

Collaborators

Emma Schymanski &
the NORMAN Network

Antony Williams

CCTE, US EPA Office of Research and Development,

Williams.Antony@epa.gov

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

Williams et al. *J Cheminform* (2017) 9:61
DOI 10.1186/s13321-017-0247-6


 Journal of Cheminformatics

DATABASE

Open Access

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry



Antony J. Williams^{1*} , Christopher M. Grulke¹, Jeff Edwards¹, Andrew D. McEachran², Kamel Mansouri^{1,2,4}, Nancy C. Baker³, Grace Patlewicz¹, Imran Shah¹, John F. Wambaugh¹, Richard S. Judson¹ and Ann M. Richard¹

<https://doi.org/10.1186/s13321-017-0247-6>