

# Structure identification approaches using the EPA CompTox Chemicals Dashboard to support mass spectrometry analyses

***Antony Williams, Charles Lowe, Alex Chao, Elin Ulrich and Jon Sobus***

*Center for Computational Toxicology and Exposure, U.S. Environmental Protection Agency, RTP, NC*

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

*August 2021  
ACS Fall Meeting, Atlanta*

A publicly accessible website delivering:

- ~883,000 chemicals with related property data
- Experimental and predicted physicochemical property data
- Integration to “biological assay data” for 1000’s of chemicals
- Information regarding consumer products containing chemicals
- Links to other agency websites and public data resources
- “Literature” searches for chemicals using public resources
- “Batch searching” for thousands of chemicals
- Downloadable Open Data for reuse and repurposing

# A single app integrating...

## SEARCH

## BATCH SEARCH

Step Four: Select Data Output Format and Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

☒ Identifiers

☒ Chemical Name(s)

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

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

107-26-2

86-30-2

14024-55-6

34484-76-5

7784-40-9

7804-94-9

60332-96-5

2122-79-5

124833-17-8

7786-81-4

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

Download

## BATCH SEARCH

Step Four: Select Data Output Format and Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

☒ Identifiers

☒ Chemical Name(s)

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

☒ InChIKey Substructure

☒ DSDS Compound ID

☒ InChIKey Substructure

☒ MS-Ready Formula(s)

☒ Exact Formula(s)

☒ Monoisotopic Mass

☒ Display All Chemicals

☒ Download Chemical Data

Select Output Format:

☒ Excel

Customize Results

☒ Select All

☒ Select All in Lists

Chemical Identifiers

☒ DTSID

☒ Chemical Name

☒ CASRN

☒ InChIKey

# Detailed Chemical Pages

## DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES


SYNONYMS

▶ LITERATURE

LINKS

COMMENTS

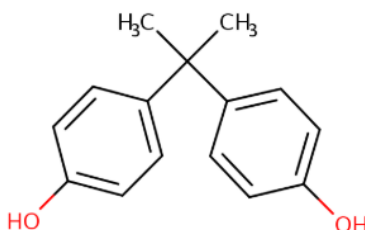
[Home](#) [Advanced Search](#) [Batch Search](#) [Lists](#) [Predictions](#) [Downloads](#)



### Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.






Chemical structure of Bisphenol A (BPA), showing two phenol rings connected by a central carbon atom bonded to two methyl groups (H<sub>3</sub>C and CH<sub>3</sub>). The hydroxyl groups (OH) are highlighted in red.



#### Wikipedia


**Bisphenol A (BPA)** is an organic synthetic compound with the chemical formula (CH<sub>3</sub>)<sub>2</sub>C(C<sub>6</sub>H<sub>4</sub>OH)<sub>2</sub> 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:** C<sub>15</sub>H<sub>16</sub>O<sub>2</sub>  [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

# Sources of Exposure to Chemicals

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

Download ▼

Columns ▼ 10 ▼

Search query

### Product and Use Categories (PUCs) ⓘ

Product or Use Categorization	Categorization type	Number of Unique Products
manufacturing, metals	CPCat Cassette	17
machines	CPCat Cassette	17
plastics	CPCat Cassette	16
flooring	CPCat Cassette	12
nt, metals	CPCat Cassette	11
chemical	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	7
	CPCat Cassette	6

### EXPOSURE

#### PRODUCT & USE CATEGORIES

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

TOXICS RELEASE INVENTORY

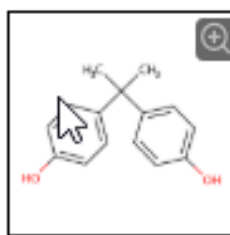
MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

First << < 1 2 3 4 5 6 7 8 9 10 > >> Last

# Physicochemical properties and environmental fate and transport



**Bisphenol A**

**80-05-7 | DTXSID7020182**

Searched by DSSTox Substance Id.

Property



Summary



Download

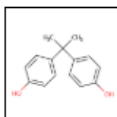


Columns



Property	Experimental average	Predicted average	
<a href="#">LogP: Octanol-Water</a>	3.32 (1)	3.29	
<a href="#">Melting Point</a>	155 (7)	139	
<a href="#">Boiling Point</a>	200 (1)	363	
<a href="#">Water Solubility</a>	5.26e-4 (1)	9.62e-4	
<a href="#">Vapor Pressure</a>	-	8.37e-7	
<a href="#">Flash Point</a>	-	190	

# Link farm to public resources



## 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](#)





# NIST WebBook

<https://webbook.nist.gov/chemistry/>


## Analytical


 FOR-IDENT

 NEMI: National Environmental  
Methods Index

 RSC Analytical Abstracts

 Tox21 Analytical Data

 MONA: MassBank North  
America

 mzCloud

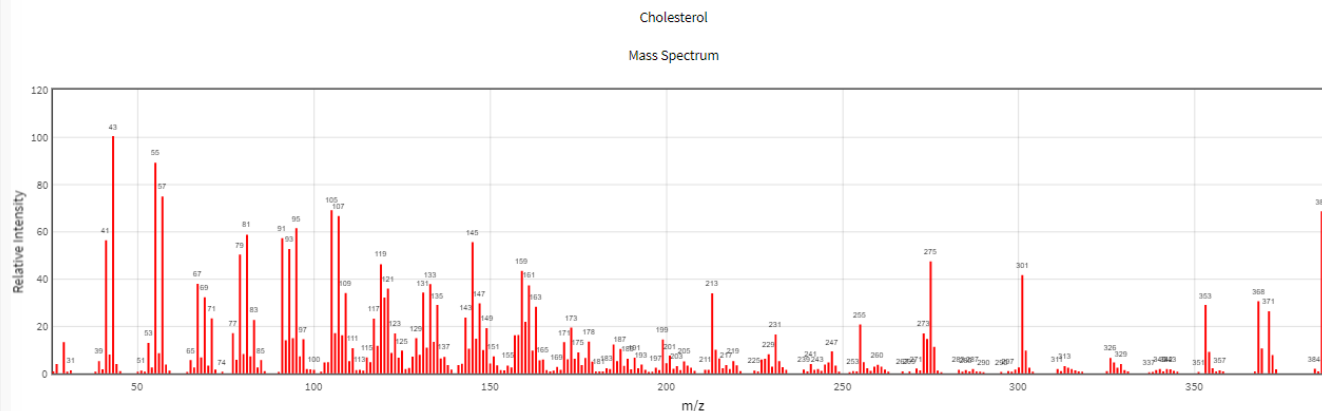
 NIST IR Spectrum

 NIST MS Spectrum

## Spectrum

Plot

Help / Software credits



# 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: ★★★★★



Oc1ccc(cc1)C(C)(C)c2ccc(O)cc2

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] <sup>+</sup>
ionization mode	positive
accession	EA016309

Originally submitted to the MassBank High Quality Mass Spectral Database

MassBank | LC-MS | Display Full Record

# ***Chemical lists***

# >300 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 ▾
<a href="#">HDXEXCH</a>	<a href="#">MASSPECDB: Hydrogen Deuterium Exchange Standard Set - Under HDX Conditions</a>	2018-11-07	592	Observed species (deuterated and undeuterated) from the HDXNOEX list under hydrogen deuterium exchange conditions (Ruttkies, Schymanski et al. in prep.)
<a href="#">HDXNOEX</a>	<a href="#">MASSPECDB: Hydrogen Deuterium Exchange Standard Set - No Exchange</a>	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.)
<a href="#">MASSBANKEUSP</a>	<a href="#">MASSPECDB: MassBank.EU Collection: Special Cases</a>	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.
<a href="#">MASSBANKREF</a>	<a href="#">MASSPECDB: MassBank Reference Spectra Collection</a>	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.
<a href="#">MYCOTOXINS</a>	<a href="#">MASSPECDB: Mycotoxins from MassBank.EU</a>	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

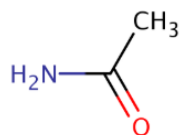


DTXSID

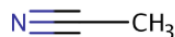


Hide chemicals that are:

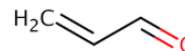
Filter by Name or CASRN



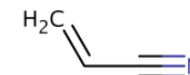
Acetamide  
CASRN:60-35-5  
DTXSID:DTXSID7020005



Acetonitrile  
CASRN:75-05-8  
DTXSID:DTXSID7020009



Acrolein  
CASRN:107-02-8  
DTXSID:DTXSID5020023



Acrylonitrile  
CASRN:107-13-1  
DTXSID:DTXSID5020029

# “Volatilome” Saliva

## LIST: VOLATILOME: Saliva

☐ Identifier substring search

### List Details

**Description:** This list is a subset of compounds detected in saliva and reported in the peer-reviewed literature and identified in experimental work at US-EPA. 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)).

**Number of Chemicals:** 307

307 chemicals

Select all

Download

Send to Batch Search

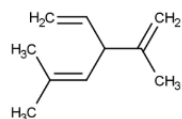
Name

CASRN

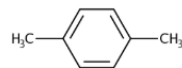
DTXSID

Hide chemicals that are:

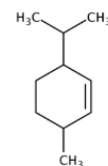
Filter by Name or CASRN



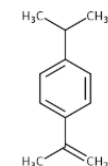
santolina triene  
CASRN:2153-66-4  
DTXSID:DTXSID90880656



p-Xylene  
CASRN:106-42-3  
DTXSID:DTXSID2021868



p-Menth-2-ene  
CASRN:5256-65-5  
DTXSID:DTXSID30333757



p-Isopropyl-alpha-methylstyrene  
CASRN:2388-14-9  
DTXSID:DTXSID90178580

# 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



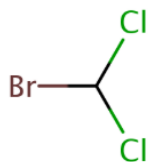
CASRN

DTXSID



Hide chemicals that are:

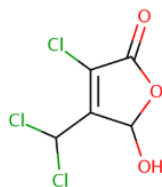
Filter by Name or CASRN



Bromodichloromethane

CASRN:75-27-4

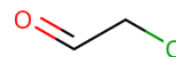
DTXSID:DTXSID1020198



3-Chloro-4-(dichloromethyl)-5-hydroxy-2(1H)-furan-2-one

CASRN:77439-76-0

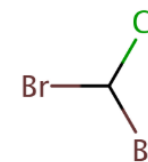
DTXSID:DTXSID6020276



Chloroacetaldehyde

CASRN:107-20-0

DTXSID:DTXSID4020292



Chlorodibromomethane

CASRN:124-48-1

DTXSID:DTXSID1020300

# 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.
- Part 1 of this report presents the findings of the literature search.
- The scope of this study was limited to the chemicals identified in the literature search.

### 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

298 chemicals

Select all

Download

Send to Batch Search

Default



CASRN

DTXSID

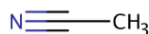
x

x

▼

Hide chemicals that are:

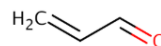
Filter by Name or CASRN



Acetonitrile

CASRN:75-05-8

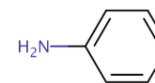
DTXSID:DTXSID7020009



Acrolein

CASRN:107-02-8

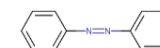
DTXSID:DTXSID5020023



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

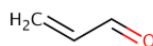
CASRN x

DTXSID x

1640 chemicals

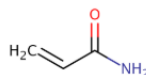
Hide chemicals that are:

Filter by Name or CASRN



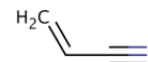
Acrolein

CASRN:107-02-8  
DTXSID:DTXSID5020023



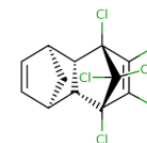
Acrylamide

CASRN:79-06-1  
DTXSID:DTXSID5020027



Acrylonitrile

CASRN:107-13-1  
DTXSID:DTXSID5020029



Aldrin

CASRN:309-00-2  
DTXSID:DTXSID8020040



United States  
Environmental Protection  
Agency

[Home](#)
[Advanced Search](#)
[Batch Search](#)
[Lists](#)
[Predictions](#)
[Downloads](#)

[Share](#)

## PFAS|EPA: PFAS structures in DSSTox (update August 2020)

☐ Identifier substring search

### List Details

**Description:** List consists of all DTXSID records with a structure assigned, and using a set of substructural filters based on community input. The substructural filters ([visible here](#)) are designed to be simple, reproducible and transparent, yet general enough to encompass the largest set of structures having sufficient levels of fluorination to potentially impart PFAS-type properties.


**Number of Chemicals:** 8163

3500 of 8163 chemicals loaded

[Select all](#)
[Download](#)
[Send to Batch Search](#)
[Default](#)

[DTXSID](#)
[CASRN](#)
[TOXCAST](#)

[Hide chemicals that are:](#)
[Filter by Name or CASRN](#)



Enflurane  
DTXSID:DTXSID1020562  
CASRN:13838-16-9  
TOXCAST:0/235



1,1,1,2-Tetrafluoroethane  
DTXSID:DTXSID1020562  
CASRN:811-9  
TOXCAST:-





Analytical and Bioanalytical Chemistry  
<https://doi.org/10.1007/s00216-021-03392-7>

### RESEARCH PAPER

## FluoroMatch 2.0—making automated and comprehensive non-targeted PFAS annotation a reality

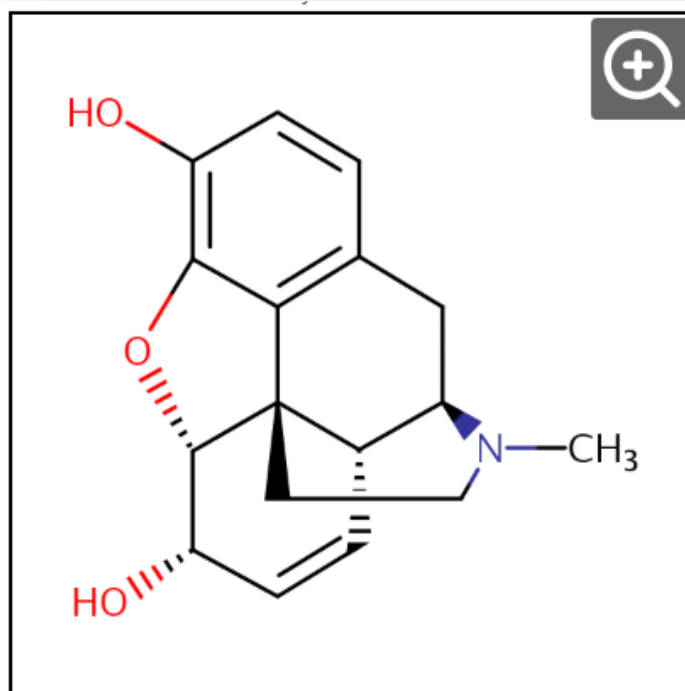
Jeremy P. Koelmel<sup>1</sup> • Paul Stelben<sup>1</sup> • Carrie A. McDonough<sup>2</sup> • David A. Dukes<sup>2</sup> • Juan J. Aristizabal-Henao<sup>3</sup> • Sara L. Nason<sup>4</sup> • Yang Li<sup>5</sup> • Sandi Sternberg<sup>6</sup> • Elizabeth Lin<sup>1</sup> • Manfred Beckmann<sup>7</sup> • Antony J. Williams<sup>8</sup> • John Draper<sup>7</sup> • Jasen P. Finch<sup>7</sup> • Jens K. Munk<sup>9</sup> • Chris Deigl<sup>10</sup> • Emma E Rennie<sup>11</sup> • John A. Bowden<sup>3,12</sup> • Krystal J. Godri Pollitt<sup>1</sup> 

Received: 5 February 2021 / Revised: 16 April 2021 / Accepted: 4 May 2021  
 © Springer-Verlag GmbH Germany, part of Springer Nature 2021



# ***Related Searches to Support Mass Spectrometry***

# Find me “related structures” Formula-Based Search



## Wikipedia

**Morphine** is a pain medication of the opiate family which is found naturally in a number of plants and animals. It acts directly on the central nervous system (CNS) to decrease the feeling of pain. It can be taken for both acute pain and chronic pain. It is frequently used for pain from myocardial infarction and during labor. It can be given by mouth, by injection into a muscle, by injection under the skin, intravenously, injection into the space around the

...

[Read more](#)

## Intrinsic Properties



**Molecular Formula:** C<sub>17</sub>H<sub>19</sub>NO<sub>3</sub>

Mol File

Find All Chemicals



**Average Mass:** 285.343 g/mol

Isotope Mass Distribution



**Monoisotopic Mass:** 285.136493 g/mol

## Structural Identifiers

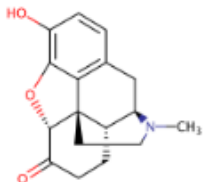
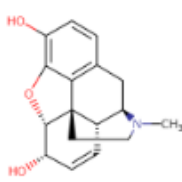
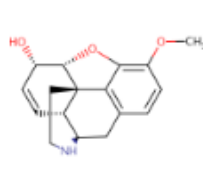
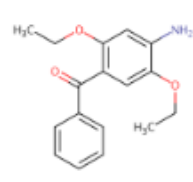
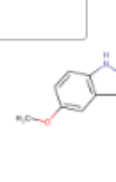
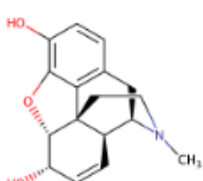
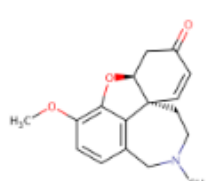
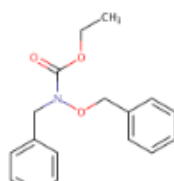
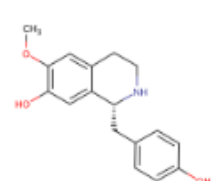
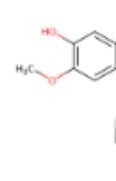
# Select Chemicals of Interest

Searched by Exact Molecular Formula: C17H19NO3

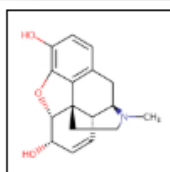
6 of 120 chemicals selected

Search  Default

Hide chemicals that are:

 <b>Hydromorphone</b> DTXSID: DTXSID8023133 PubChem: 39 CPDAT: 21	 <b>Morphine</b> DTXSID: DTXSID9023336 PubChem: 52 CPDAT: 37	 <b>Norcodeine</b> DTXSID: DTXSID8046327 PubChem: 20 CPDAT: 2	 <b>Methanone, (4-amino-2,5-diethoxyphenyl)phenyl</b> DTXSID: DTXSID3071696 PubChem: 16 CPDAT: 0	 <b>Carbazone</b> DTXSID: DTXSID... PubChem: 1 CPDAT: 0
 <b>Morphinan-3,6-diol, 7,8-didehydro-4,5-epoxy</b> DTXSID: DTXSID40167308 PubChem: 8 CPDAT: 0	 <b>Narwedine, (+/-)-</b> DTXSID: DTXSID60168190 PubChem: 43 CPDAT: 0	 <b>Carbamic acid, (phenylmethoxy)(phenyl)methyl</b> DTXSID: DTXSID50171931 PubChem: 11 CPDAT: 0	 <b>7-Isoquinolinol, 1,2,3,4-tetrahydro-1-((4-hydroxyphenyl)methyl)</b> DTXSID: DTXSID70176367 PubChem: 21 CPDAT: 0	 <b>Cherylin</b> DTXSID: DTXSID... PubChem: 1 CPDAT: 0

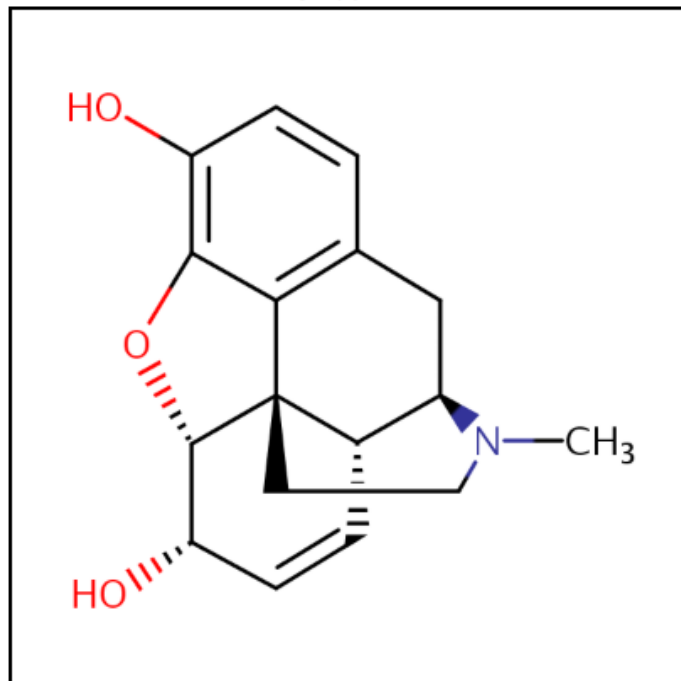
# Find me “related structures” Based on Structure Similarity



## Morphine

57-27-2 | DTXSID9023336

Searched by Approved Name.



Wikipedia

Intrinsic Properties

Structural Identifiers

Linked Substances

Same Connectivity: [3 records](#) (based on first layer of InChI)

Mixtures, Components and Neutralized Forms: [13 records](#) (based on QSAR ready mappings and with the compound as a component of a mixture)

MS-Ready Mappings: [DTXCID60196731: 21 records](#);

**Similar Compounds: [178 records](#) (based on Tanimoto coefficient >0.8)**

Presence in Lists

# Find me “related structures” Based on Structure Similarity

Searched with a similarity threshold of 0.8

178 chemicals

Select all

Download

Send to Batch Search

Similarity



DTXSID

CASRN

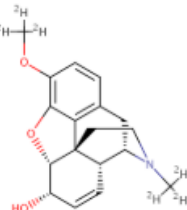
TOXCAST

Similarity



Hide chemicals that are:

Filter by Name or CASRN



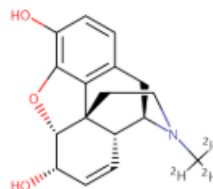
(5alpha,6alpha,9alpha)-17-(<sup>2</sup>H<sub>3</sub>)Methyl-...

DTXSID: DTXSID30747375

CASRN: 1007844-34-9

TOXCAST: -

Similarity: 1.00



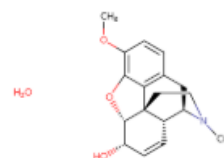
Morphinan-3,6-alpha-diol, 7,8-didehydr...

DTXSID: DTXSID00217656

CASRN: 67293-88-3

TOXCAST: -

Similarity: 1.00



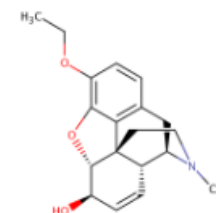
Codeine hydrate

DTXSID: DTXSID20209391

CASRN: 6059-47-8

TOXCAST: -

Similarity: 1.00



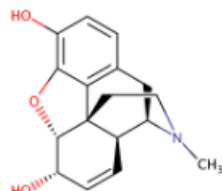
Ethyl-alpha-isomorphine

DTXSID: DTXSID30197073

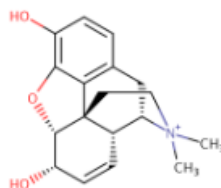
CASRN: 47252-01-7

TOXCAST: -

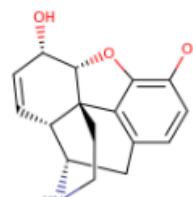
Similarity: 1.00



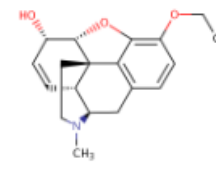
Morphinan-3,6-diol, 7,8-didehydro-4,5-e...



N-Methylmorphine



Normorphine



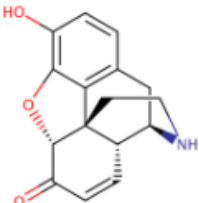
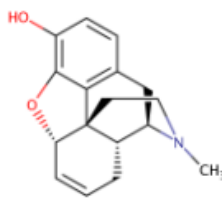
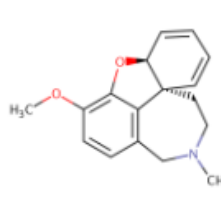
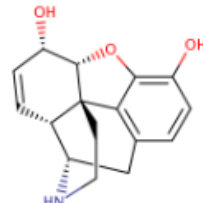
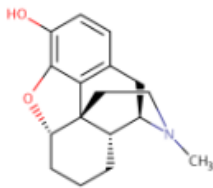
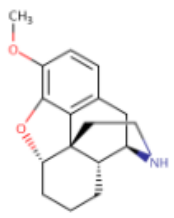
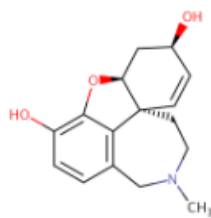
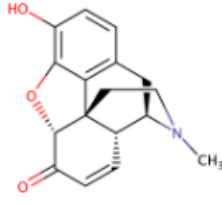
Ethylmorphine

# Find me “related structures”

## Structure Similarity – sort on mass

178 chemicals

Select all Download Send to Batch Search **Mass** CASRN DTXSID Mass Hide chemicals that are: Filter by Name or CASRN

 <p>(5alpha)-7,8-Didehydro-4,5-epoxy-3-hy... CASRN: 6872-48-6 DTXSID: DTXSID00218845 Mass: 269.105193</p>	 <p>Desoxymorphine C CASRN: 63732-65-0 DTXSID: DTXSID20213170 Mass: 269.141579</p>	 <p>R-116937 CASRN: 664995-65-7 DTXSID: DTXSID60216701 Mass: 269.141579</p>	 <p>Normorphine CASRN: 466-97-7 DTXSID: DTXSID1049019 Mass: 271.120843</p>
 <p>Desomorphine CASRN: 427-00-9 DTXSID: DTXSID10195390</p>	 <p>Morphinan-14-ol, 4,5-alpha-epoxy-3-me... CASRN: 55256-27-4 DTXSID: DTXSID90203765</p>	 <p>O-Desmethylgalantamine CASRN: 60755-80-8 DTXSID: DTXSID40209575</p>	 <p>Morphinone CASRN: 467-02-7 DTXSID: DTXSID50196907</p>



# ***Mass & 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

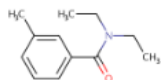
CASRN

TOXCAST

Mass Diff

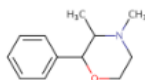
Multicomponent Chemicals

Filter by Name or CASRN



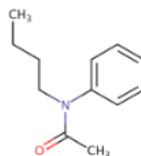
DEET

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



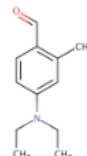
Phendimetrazine

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



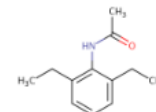
N-Butylacetanilide

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



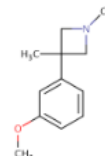
Benzaldehyde, 4-(diethylamino)-2-methyl-

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



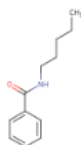
Acetanilide, 2',6'-diethyl-

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



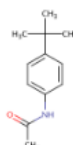
Azetidine, 1,3-dimethyl-3-(m-methoxyphenyl)-

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



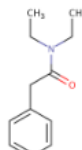
Benzamide, N-pentyl-

DTXSID: DTXSID20174196  
CASRN: 20330-43-4  
TOXCAST: -  
Mass Diff: 0.000014



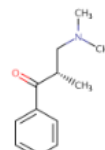
p-t-Butylacetanilide

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



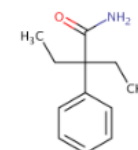
N,N-Diethylphenylacetamide

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



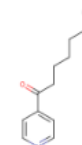
3-(Dimethylamino)-2-methylpropiofenone

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: C<sub>6</sub>H<sub>8</sub>O<sub>2</sub> or C<sub>6</sub>H(8-10)O(0-2)

Search 

# “MS-Ready Structures”

<https://doi.org/10.1186/s13321-018-0299-2>

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

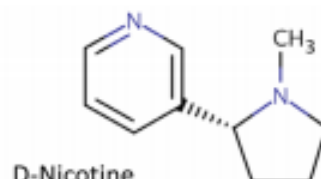
Andrew D. McEachran<sup>1,2\*</sup>, Kamel Mansouri<sup>1,2,3</sup>, Chris Grulke<sup>2</sup>, Emma L. Schymanski<sup>4</sup>, Christoph Ruttkies<sup>5</sup> and Antony J. Williams<sup>2\*</sup>

## Open Science for Identifying “Known Unknown” Chemicals

Emma L. Schymanski<sup>\*,†</sup> and Antony J. Williams<sup>\*,‡</sup>

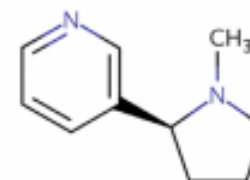


CN1CCC[C@H]1C1=CN=CC=C1  
DTXSID1020930 | SNICXCGAKADSCV  
54-11-5 | **162.1157** | 0.929 | **72**  
Tox: **yes** | Expo: **yes** | Bioassay: **yes**

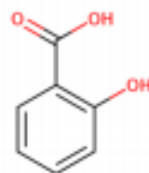
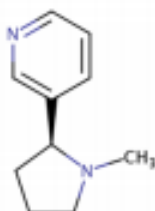


CN1CCC[C@@H]1C1=CN=CC=C1  
DTXSID004635 | SNICXCGAKADSCV  
25162-00-9 | **162.1157** | 0.929 | **20**  
Tox: **no** | Expo: **yes** | Bioassay: **yes**

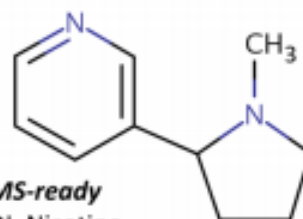
**LEGEND:** Name, SMILES  
DTXSID | InChIKey 1<sup>st</sup> Block  
CAS | Monoiso. Mass | logP | Sources  
Data on: Toxicity | Exposure | Bioassays



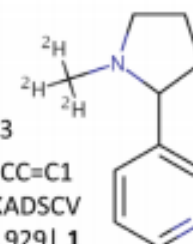
Nicotine hydrochloride  
Cl.CN1CCC[C@H]1C1=CN=CC=C1  
DTXSID602093 | HDJBTCAJIMNXEW  
2820-51-1 | **198.0924** | 0.929 | **9**  
Tox: **no** | Expo: **yes** | Bioassay: **yes**



Benzoic acid, 2-hydroxy-, compd. with  
3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)  
OC(=O)C1=CC(=O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1  
DTXSID5075319 | AIBWPBUAKCMKNS  
29790-52-1 | **300.1474** | 0.929 | **6**  
Tox: **no** | Expo: **yes** | Bioassay: **no**



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



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

- EXACT Formula:** C<sub>10</sub>H<sub>16</sub>N<sub>2</sub>O<sub>8</sub>: 3 Hits

☐ MS Ready Formula ☒ Exact Formula

Formula

C<sub>10</sub>H<sub>16</sub>N<sub>2</sub>O<sub>8</sub>

Select all Download Send to Batch Search Default DTXSID PubChem CPDAT 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

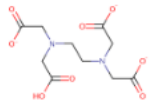
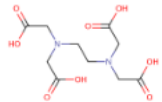
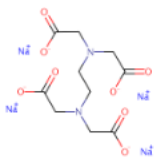
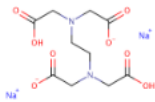
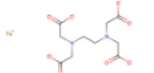
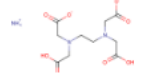
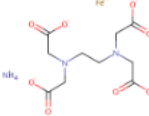
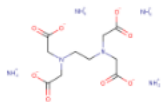
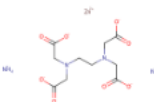
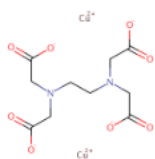
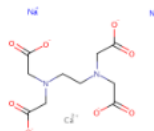
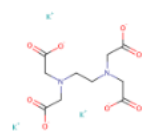
# MS-Ready Mappings

- **Same Input Formula: C<sub>10</sub>H<sub>16</sub>N<sub>2</sub>O<sub>8</sub>**
- **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**

# MS-Ready Mappings

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

## Progaron

108532-15-6 | DTXSID20148579  
Searched by DSSTox Substance Id.

**DETAILS**

- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS

**Intrinsic Properties**

Molecular Formula:  $C_{31}H_{43}Cl_4N_{11}O_2$  Mol File Find All Chemicals

Average Mass: 743.56 g/mol Isotope Mass Distribution

**Structural Identification**

**Linked Substances**

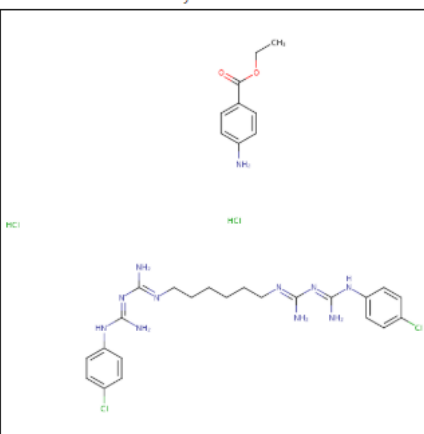
Same Connectivity: 1 record (nChl)

Mixtures, Component: records

**MS-Ready Mappings:** DTXCID301804: 12 records; DTXCID0013314: 11 records;

**Similar Compounds:** 0 records

**Quality Control Notes**



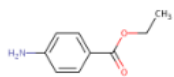
# MS-Ready Mappings Set

## MS-Ready Mappings of Benzocaine (Isotopes pre-filtered)

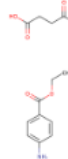
Select all Download Send to Batch Search Default DTXSID PubChem CPDAT

10 of 12 chemicals visible

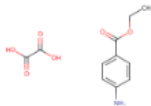
Isotopes Filter by Name or CASRN



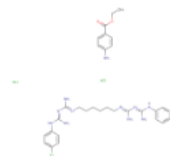
**Benzocaine**  
DTXSID: DTXSID8021804  
PubChem: 184  
CPDAT: 42



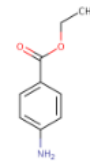
**Anesthesine succinate**  
DTXSID: DTXSID60148336  
PubChem: 10  
CPDAT: 0



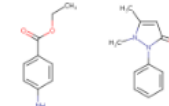
**Anesthesine oxalate**  
DTXSID: DTXSID20148337  
PubChem: 6  
CPDAT: 0



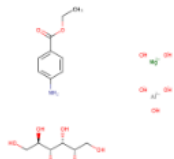
**Progaron**  
DTXSID: DTXSID20148579  
PubChem: 5  
CPDAT: 0



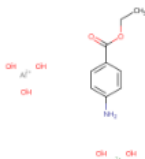
**Benzocaine hydrochloride**  
DTXSID: DTXSID50177812  
PubChem: 33  
CPDAT: 0



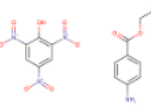
**Antipyrine mixture with benzocaine**  
DTXSID: DTXSID80212866  
PubChem: 9  
CPDAT: 0



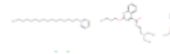
**Almagel A-neo**  
DTXSID: DTXSID60227559  
PubChem: 9  
CPDAT: 0



**Almagel**  
DTXSID: DTXSID70227560  
PubChem: 9  
CPDAT: 0



**Ethyl 4-aminobenzoate-2,4,6-trinitroph...**  
DTXSID: DTXSID70787033  
PubChem: 5  
CPDAT: 0

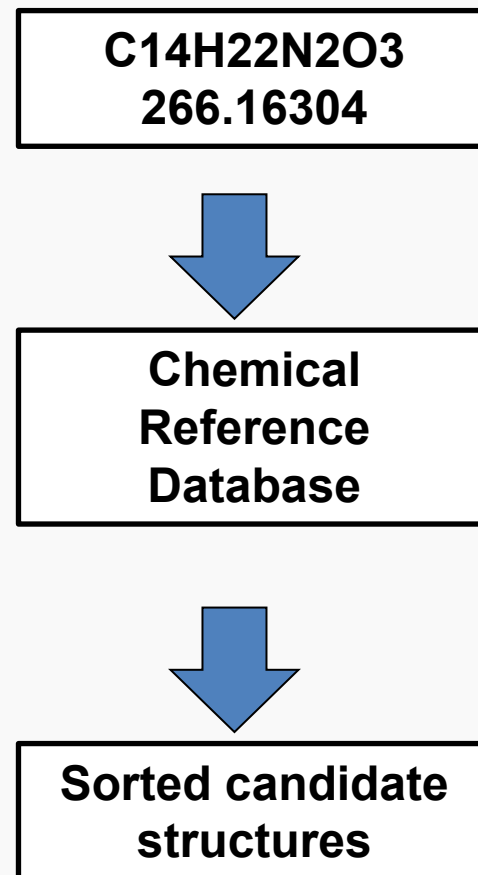


**1-Hexadecylpyridin-1-ium 2-butoxy-N-...**  
DTXSID: DTXSID50997335  
PubChem: 0  
CPDAT: 0

# ***Candidate ranking***

# Data Source Ranking of “known unknowns”

- Mass and/or formula is for an unknown chemical but contained within a reference database
- Most likely candidate chemicals have the most associated data sources, most associated lit. articles or both



# Is a bigger database better?



© 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

- ChemSpider was 26 million chemicals then
- Much BIGGER today
- Is bigger better??

**110** Million  
chemical structures

- Use available metadata to rank candidates
  - Associated data sources
    - Associated lists in the underlying database
    - Associated data sources in PubChem
    - Specific types (e.g. water, surfactants, pesticides etc.)
  - Number of associated literature articles (Pubmed)
  - **Chemicals in the environment** – the number of products/categories containing the chemical is a very important source of data

<sup>w</sup>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. McEachran<sup>1</sup> • Jon R. Sobus<sup>2</sup> • Antony J. Williams<sup>3</sup>

- Dashboard content *was* 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 <sup>a</sup>	1.3	2	1			

# How did performance compare?

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

<sup>a</sup> 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**

# How did performance compare?

Summary statistics and rank-ordered position in the CompTox Chemistry Dashboard and ChemSpider of the 89 compound subset from the Little et al. [7] study

		Average rank	Number in each position rank-ordered				
		( $\pm$ SD)	#1	#2	#3	#4	#5+
Mass-based	Dashboard	1.2 $\pm$ 0.7	77 <sup>a</sup>	5	3	3	
	ChemSpider	2.2 $\pm$ 6.1 <sup>b</sup>	68	8	7	1	5
Formula-based	Dashboard	1.1 $\pm$ 0.4	78 <sup>a</sup>	8	2		
	ChemSpider	1.3 $\pm$ 1.0	77	8	2	1	2

<sup>a</sup>One chemical (tephrosin) not present in the Dashboard

# Will the correct Microcystin LR Stand Up? ChemSpider Skeleton Search

Matches any text strings used to describe a molecule.

ZYZCGGRZINLQBL



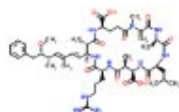
Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID ?

FILTER ▾

Search Hits Limit: 100 ▾

Found 9 results

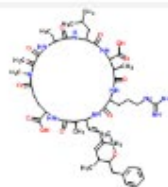
Search term: ZYZCGGRZINLQBL (Found by InChIKey (skeleton match))



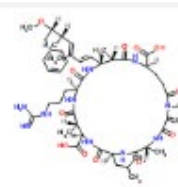
[Cyanoginosin](#)



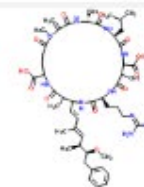
[MCYST-LR](#)



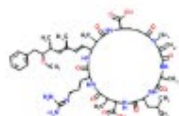
[15-\(3-Carbamimidamidopropyl\)-8-isobutyl-18-\(\(1E,3E\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)](#)



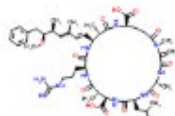
[\(5R,8S,11R,12S,15S,18R,19S,22R\)-15-\(3-Carbamimidamidopropyl\)-8-isobutyl-18-\(\(1E,3E,5S,6S\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)](#)



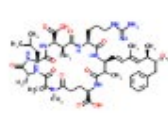
[\(5R,8S,15S\)-15-\(3-\(\(Diaminomethylene\)amino\)propyl\)-8-isobutyl-18-\(\(1E,3E,5S,6S\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)](#)



[15-\(3-\(\(Diaminomethylene\)amino\)propyl\)-8-isobutyl-18-\(6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)](#)



[\(5R,8S,11R,12S,15S,18S,19S,22R\)-15-\(3-Carbamimidamidopropyl\)-8-isobutyl-18-\(\(1E,5S,6S\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)](#)

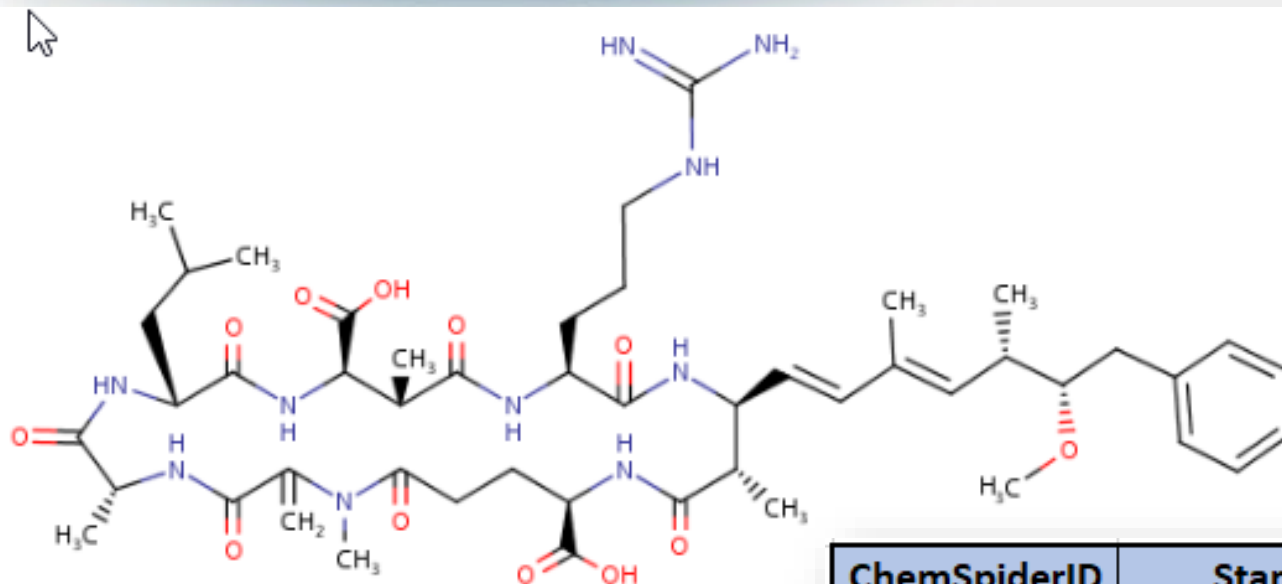


[\(5R,8R,11R,12S,15S,18S,22R\)-15-\(3-Carbamimidamidopropyl\)-8-isobutyl-18-\(\(1E,3E,5R,6R\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)](#)



[Diamino-N-\(3-\(\(5R,8S,11R,12S,15S,18S,19S,22R\)-11,2-dicarboxy-8-isobutyl-18-\(\(1E,3E,5S,6S\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)propyl\)-8-isobutyl-18-\(\(1E,3E,5S,6S\)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-yl\)\)](#)

# Comparing ChemSpider Structures



ChemSpiderID	Standard InChIKey Stereolayer
<b>WIKIPEDIA</b>	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
<b>CompTox</b>	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
<a href="#">4941647</a>	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
<a href="#">393078</a>	t28-,29-,30-,31+,34-,35-,36+, <b>37-</b> ,38-,40+
<a href="#">57618348</a>	t28-,29-,30-,31+,34-,35-,36+, <b>37-</b> ,38-,40+
<a href="#">29342071</a>	t28-,29-,30-,31+, <b>34+</b> ,35-,36+, <b>37-</b> ,38-,40+
<a href="#">7987594</a>	t28-, <b>29?</b> , <b>30?</b> ,31+, <b>34?</b> ,35-, <b>36?</b> , <b>37-</b> ,38-, <b>40?</b>
<a href="#">22900854</a>	t28-, <b>29?</b> , <b>30+</b> , <b>31-</b> , <b>34+</b> , <b>35+</b> , <b>36-</b> , <b>37-</b> ,38-, <b>40-</b>
<a href="#">19692240</a>	NONE
<a href="#">2831283</a>	NONE

# Comparing ChemSpider Structures

ChemSpiderID	InChIKey	# Stereocenters	# Different
WIKIPEDIA	ZYZCGGRZINLQBL-JCGNTXOTSA-N	10/10	0
CompTox	ZYZCGGRZINLQBL-JCGNTXOTSA-N	10/10	0
<a href="#">4941647</a>	ZYZCGGRZINLQBL-JCGNTXOTSA-N	10/10	0
<a href="#">393078</a>	ZYZCGGRZINLQBL-GWRQVWKTSA-N	10/10	1
<a href="#">57618348</a>	ZYZCGGRZINLQBL-UPPCHHEJSA-N	10/10	1
<a href="#">29342071</a>	ZYZCGGRZINLQBL-IIJTUTQBSA-N	10/10	2
<a href="#">7987594</a>	ZYZCGGRZINLQBL-BESLYTPASA-N	5/10	6
<a href="#">22900854</a>	ZYZCGGRZINLQBL-QAXSDTKVSA-N	9/10	8
<a href="#">19692240</a>	ZYZCGGRZINLQBL-ORZJCNCZSA-N	0/10	10
<a href="#">2831283</a>	ZYZCGGRZINLQBL-UHFFFAOYSA-N	0/10	10

# Other Searches



**PubChem** [About](#)

SEARCH FOR

**ZYZCGGRZINLQBL**

Treating this query as a text search.

**Compounds**  
**(17)**

Show **All** entries

CMR. Query InChI...	src_id	Source	src_compound_id
...matches...	1	ChEMBL	<a href="#">CHEMBL444092</a>
...matches...	4	Guide to Pharmacology	<a href="#">4735</a>
...matches...	6	KEGG Ligand	<a href="#">C05371</a>
...matches...	7	ChEBI	<a href="#">6925</a>
...matches...	9	ZINC	<a href="#">ZINC000169715525</a>
...matches...	9	ZINC	<a href="#">ZINC000255288110</a>
...matches...	9	ZINC	<a href="#">ZINC000255288111</a>
...matches...	9	ZINC	<a href="#">ZINC000255288112</a>
...matches...	9	ZINC	<a href="#">ZINC000255288113</a>
...matches...	9	ZINC	<a href="#">ZINC000255288114</a>
...matches...	9	ZINC	<a href="#">ZINC000255288115</a>
...matches...	9	ZINC	<a href="#">ZINC000583653042</a>
...matches...	9	ZINC	<a href="#">ZINC000669680403</a>
...matches...	10	eMolecules	<a href="#">26754757</a>
...matches...	10	eMolecules	<a href="#">31239828</a>
...matches...	11	IBM Patent System	<a href="#">DA3C2F25F29692734272194ED0E2C009</a>
...matches...	14	FDA SRS	<a href="#">EQ8332842Y</a>

# ***Batch Searches***



# List of Opioids – Presence in Lists?

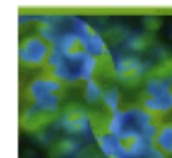


ELSEVIER


## Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059

 **TrEAC**  
TRENDS IN ENVIRONMENTAL ANALYTICAL CHEMISTRY



### Opioid occurrence in environmental water samples—A review

Marina Celia Campos-Mañas <sup>a</sup>, Imma Ferrer <sup>b</sup>  , E.Michael Thurman <sup>b</sup>, Ana Agüera <sup>a</sup>

 **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

Display All Chemicals Download Chemical Data

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

# 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"/>	Naltiben	-	-	-	-	-	-
<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

# Opioids and Metabolites (160)

## DRUGS: Opioids and related metabolites

Search OPIOIDS Chemicals

☐ 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

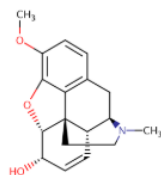
DTXSID

180 chemicals

Hide chemicals that are:

Filter by Name or CASRN

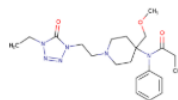
⌵



Codeine

CASRN:76-57-3

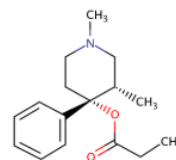
DTXSID:DTXSID2020341



Alfentanil

CASRN:71195-58-9

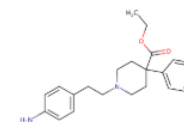
DTXSID:DTXSID9022570



Alphaprodine

CASRN:77-20-3

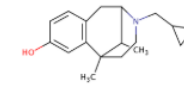
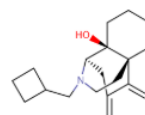
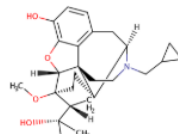
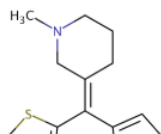
DTXSID:DTXSID4022575



Anileridine

CASRN:144-14-9

DTXSID:DTXSID8022610



- We work with **thousands** of masses/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 Five: Choose Data Fields to Download

Please enter one identifier per line



#### Select Input Type(s)

+/-  ppm

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

Display All Chem

#### 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-phenoxypyrimidine	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-quin	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-trimethoxypheny	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

# ***Benefits of Open Data***



# API services and Open Data

- Available API and web services
- Open Data available for download



United States  
Environmental Protection  
Agency

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

---

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

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

Posted: 11/14/2016

	casm	dsstox_substance_id	preferred_name
1	26148-68-5	DTXSID7020001	A-alpha-C
2	107-29-9	DTXSID2020004	Acetaldehyde oxime
3	60-35-5	DTXSID7020005	Acetamide
4	103-90-2	DTXSID2020006	Acetaminophen
5	968-81-0	DTXSID7020007	Acetohexamide
6	18523-69-8	DTXSID2020008	Acetone[4-(5-nitro-2-furyl)-2-thiazolyl] hydrazone
7	75-05-8	DTXSID7020009	Acetonitrile
8	127-06-0	DTXSID6020010	Acetoxime
9	65734-38-5	DTXSID6020012	N'-Acetyl-4-(hydroxymethyl) phenylhydrazine

---

[DSSTox MS Ready Mapping File](#)

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.

Posted: 11/14/2016

---

[DSSTox SDF File](#)

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 /Chem

Posted: 12/14/2016

- Dozens of web services to provide access to data
- Data in UI, JSON and XML format

<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>

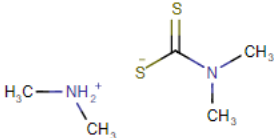
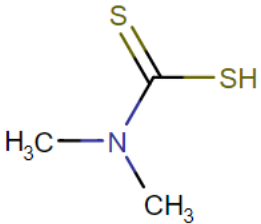
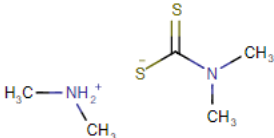
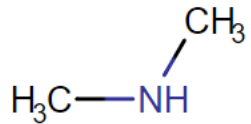
<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

<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>

# Example: 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>

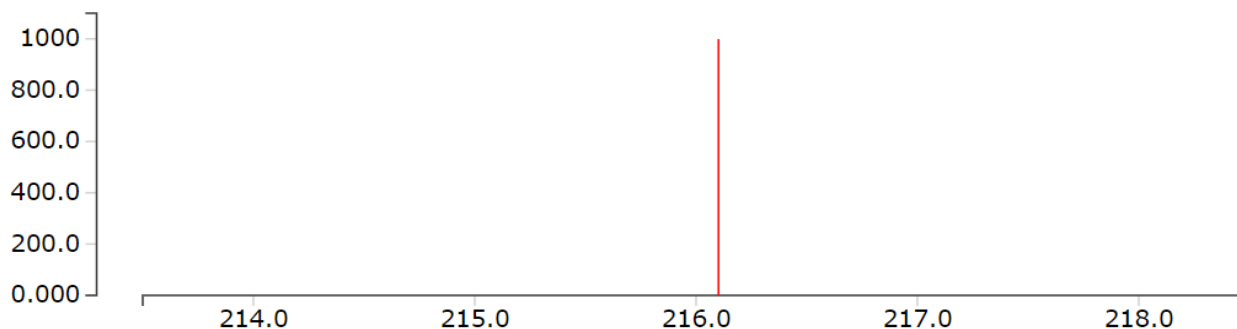
# MassBank mapping to Dashboard

## MassBank Record: EA028808

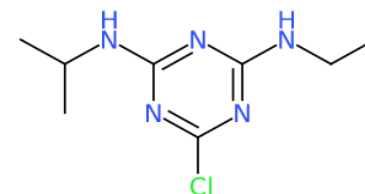
[Home](#) | [Search](#) | [Record Index](#) | [Data Privacy](#) | [Imprint](#) | MassBank ID:

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

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](#)

# NORMAN Suspect List Exchange

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



<b>Wastewater Suspect List based on Swedish Product Data</b>	Wastewater Suspect List <b>Original File with Mapped DTXSIDs</b> (12/02/2019)	KEMIWSUS <b>InChIKeys</b> (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.
<b>Algal toxins list from CompTox</b>	ALGALTOX <b>XLSX, CSV</b> (14/02/2019) CompTox <b>ALGALTOX List</b>	ALGALTOX <b>InChIKeys</b> (14/02/2019)	List of algal toxins (generated during blooms) from the CompTox Chemicals Dashboard.
<b>CCL 4 Chemical Candidate List</b>	CCL4 <b>XLSX, CSV</b> (14/02/2019) CompTox <b>CCL4 List</b>	CCL4 <b>InChIKeys</b> (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.
<b>Hydrogen Deuterium Exchange (HDX) Standard Set</b>	HDXNOEX <b>XLSX, CSV</b> (14/02/2019) CompTox <b>HDXNOEX List</b> CompTox <b>HDXEXCH List</b>	HDXNOEX <b>InChIKeys</b> (14/02/2019)	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule HRMS (Ruttkies et al. submitted). <b>HDXEXCH</b> list also contains observed deuterated species.
<b>Neurotoxicants Collection from Public Resources</b>	NEUROTOXINS <b>XLSX, CSV</b> (14/02/2019) CompTox <b>NEUROTOXINS List</b>	NEUROTOXINS <b>InChIKeys</b> (14/02/2019)	A list of neurotoxicants compiled from public resources, details on CompTox and Schymanski <i>et al.</i> (submitted).
<b>Statins Collection from Public Resources</b>	STATINS <b>XLSX, CSV</b> (14/02/2019) CompTox <b>STATINS List</b>	STATINS <b>InChIKeys</b> (14/02/2019)	A list of statins (lipid-lowering medications) compiled from public resources, details on CompTox.
<b>Synthetic Cannabinoids and Psychoactive Compounds</b>	SYNTHCANNAB <b>XLSX, CSV</b> (14/02/2019) CompTox <b>SYNTHCANNAB List</b>	SYNTHCANNAB <b>InChIKeys</b> (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>



# ***Work in Progress***

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Structure/substructure/similarity search
- The EPA NTA WebApp
- Access to API and web services

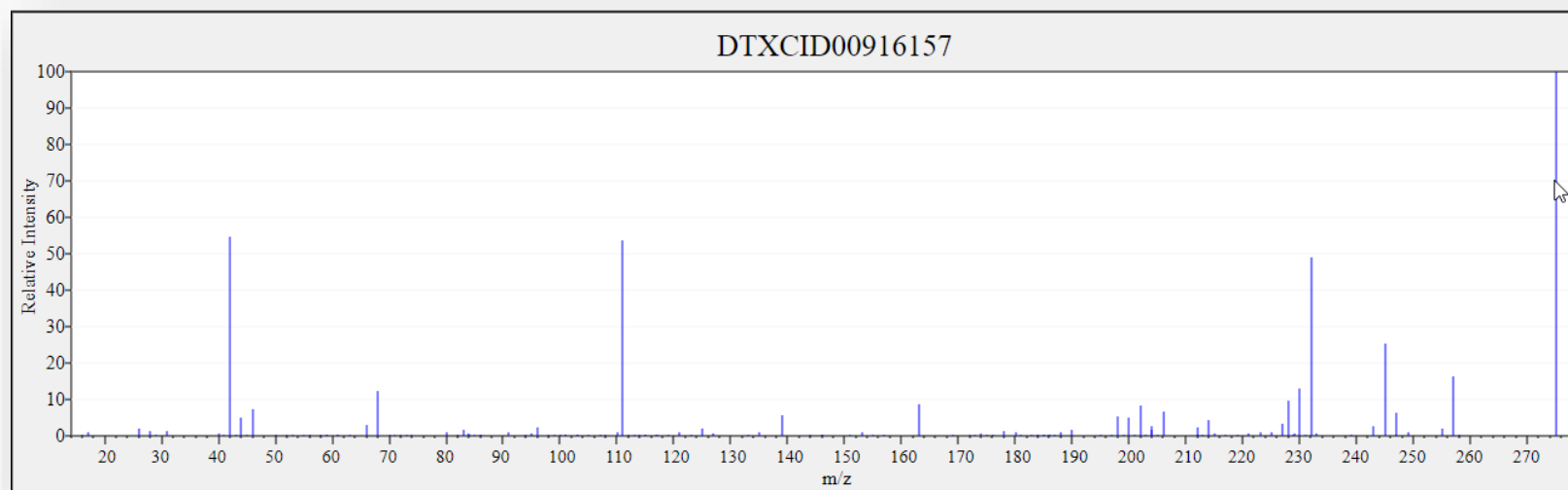


# Predicted Mass Spectra

<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




- Predictions generated and stored for >800,000 structures
- Python code to score experimental vs predicted spectra
- Cosine dot product match score calculation

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](#) 

# 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



## Spectra Input

Single Ener

304.1332052 11.61  
198.0913404 7.30  
123.0440559 6.53  
196.0756904 5.28  
216.1019051 4.70  
200.1080005 4.80

## Peak Match

Search

[TSV](#) [CSV](#) [Excel](#)

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

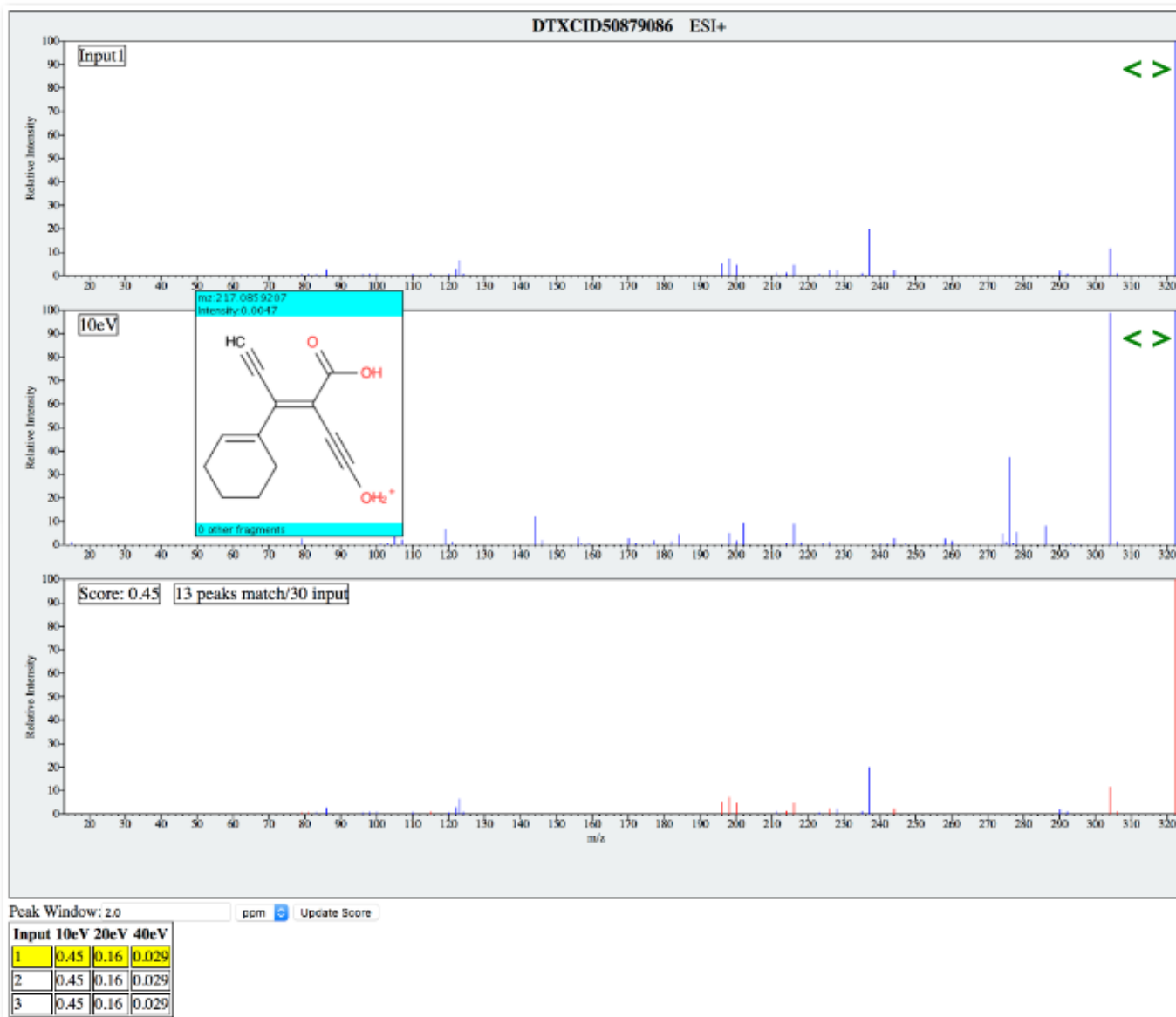
[DTXCID60349982](#)

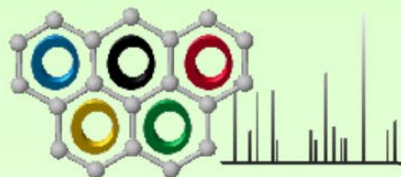
0.11

[DTXCID10316649](#)

0.09

# Spectral Viewer Comparison





# 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

Open Access

Article

# Revisiting Five Years of CASMI Contests with EPA Identification Tools

by  Andrew D. McEachran<sup>1,\*</sup>  ,  Alex Chao<sup>1</sup>  ,  Hussein Al-Ghoul<sup>1</sup>  ,  Charles Lowe<sup>2</sup>  ,  
 Christopher Grulke<sup>2</sup>  ,  Jon R. Sobus<sup>2</sup>   and  Antony J. Williams<sup>2,\*</sup>  

<sup>1</sup> Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA

<sup>2</sup> Center for Computational Toxicology and Exposure, Office of Research and Development, U.S. Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, USA

\* Authors to whom correspondence should be addressed.

*Metabolites* 2020, 10(6), 260; <https://doi.org/10.3390/metabo10060260>

Received: 3 May 2020 / Revised: 3 June 2020 / Accepted: 17 June 2020 / Published: 23 June 2020



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<sup>1,2</sup> • Hussein Al-Ghoul<sup>1,2</sup> • Andrew D. McEachran<sup>1,3</sup> • Ilya Balabin<sup>4</sup> • Tom Transue<sup>4</sup> • Tommy Cathey<sup>4</sup> • Jarod N. Grossman<sup>2,3</sup> • Randolph Singh<sup>1,5</sup> • Elin M. Ulrich<sup>2</sup> • Antony J. Williams<sup>6</sup> • Jon R. Sobus<sup>2</sup>

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

© The Author(s) 2019



# NTA WebApp Development

## NTA WebApp: Input Page

**EPA** United States Environmental Protection Agency

Environmental Topics   Laws & Regulations   About EPA   Search EPA.gov

[Contact Us](#)

### NTA: non-targeted analysis of MS data (beta)

**Tools**

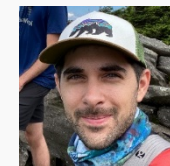
- MS1 Tool**
  - Run MS1 Tool
  - MS1 Tool Algorithms
  - MS1 Tool QA/QC
  - MS1 Tool References
  - MS2 CFMID Tool

**Documentation**

- Source Code

### Run NTA MS1 Tool

Input	Value
Project name:	<input type="text" value="Example nta"/>
Positive MPP file (csv):	<input type="button" value="Choose File"/> No file chosen
Negative MPP file (csv):	<input type="button" value="Choose File"/> No file chosen
Adduct mass accuracy units:	<input type="text" value="ppm"/>
Adduct mass accuracy:	<input type="text" value="10"/>
Adduct retention time accuracy (mins):	<input type="text" value="0.05"/>
Tracer file (csv; optional):	<input type="button" value="Choose File"/> No file chosen
Tracer mass accuracy units:	<input type="text" value="ppm"/>
Tracer mass accuracy:	<input type="text" value="5"/>
Tracer retention time accuracy (mins):	<input type="text" value="0.1"/>
Min sample:blank cutoff:	<input type="text" value="3"/>
Min replicate hits:	<input type="range" value="2"/>
Max replicate CV:	<input type="text" value="0.8"/>
Parent ion mass accuracy (ppm):	<input type="range" value="5"/>
Discard features below this retention time (mins):	<input type="text" value="0.0"/>
Search dashboard by:	<input type="text" value="mass"/>
Save top result only?	<input type="text" value="no"/>
DSSTox search batch size (debugging):	<input type="text" value="150"/>



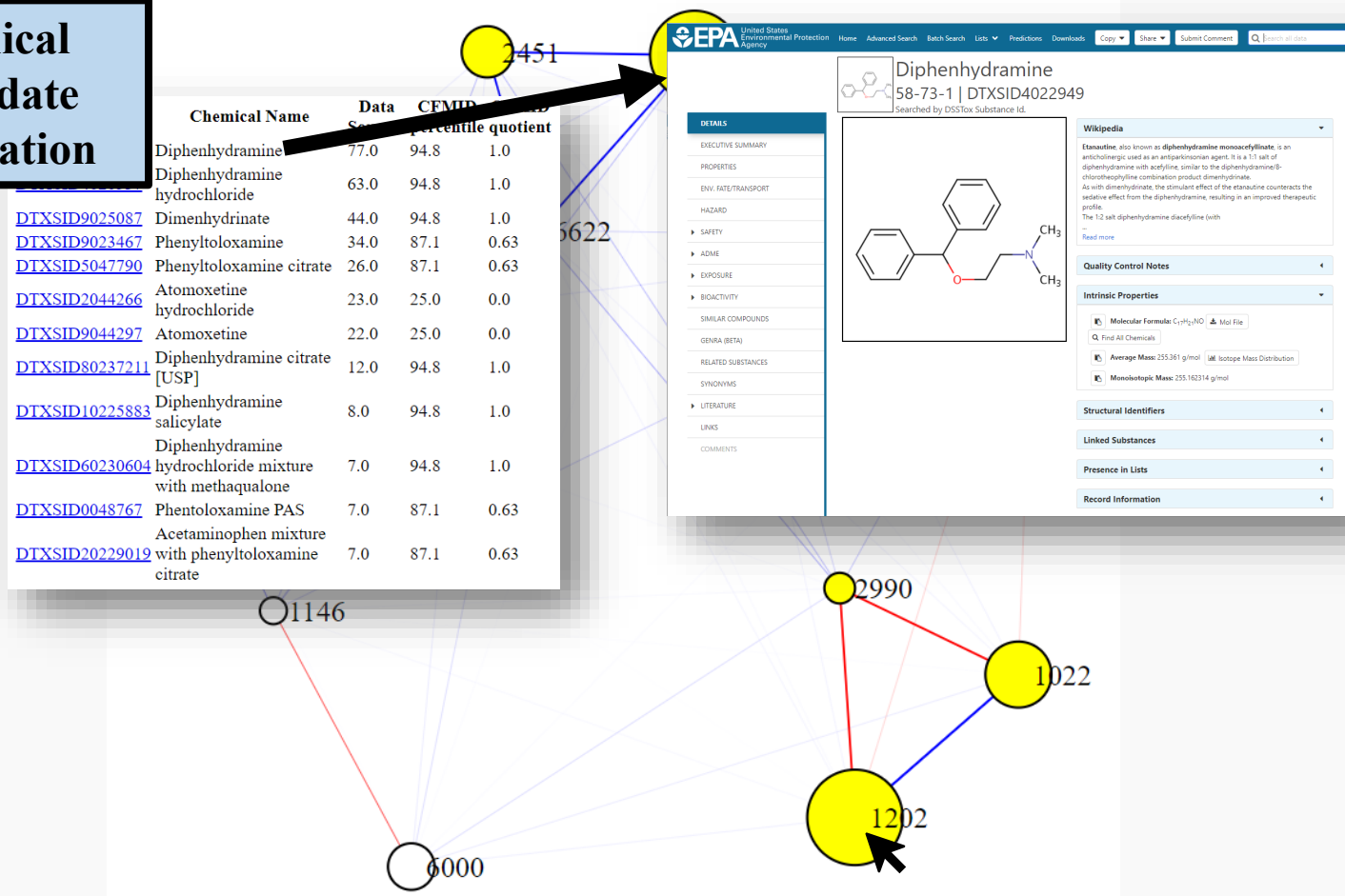
*Jeff  
Minucci*

NTA data input  
files

Parameters for  
NTA clean-up /  
database  
searching

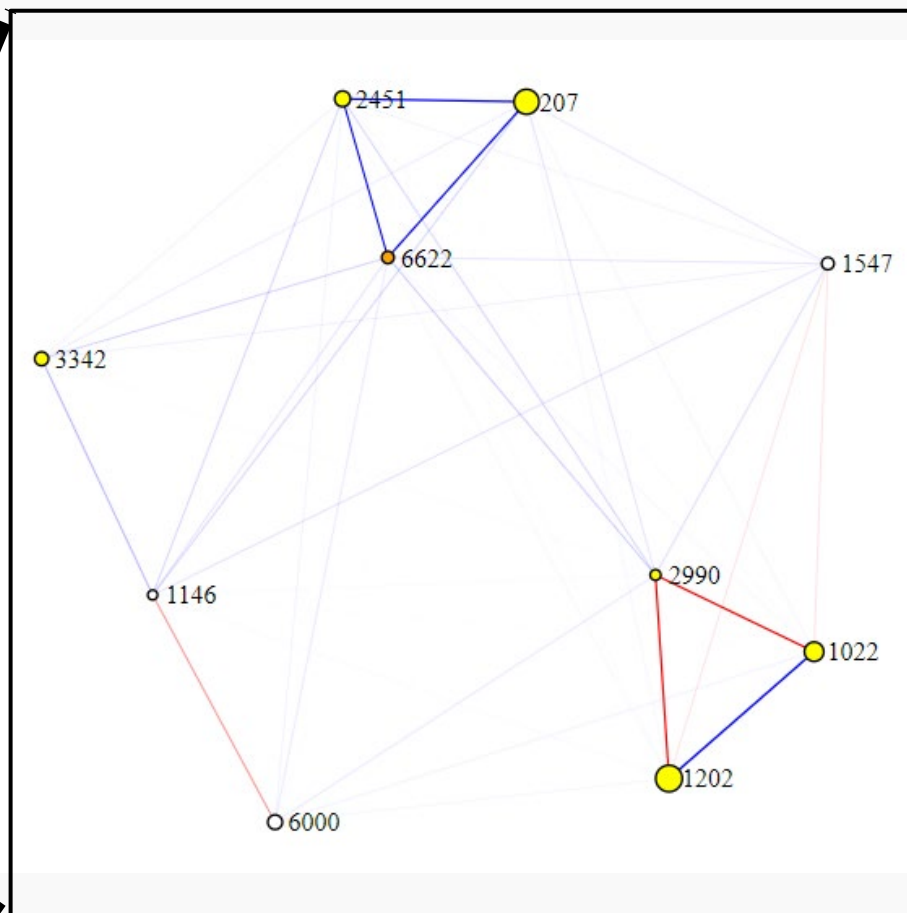
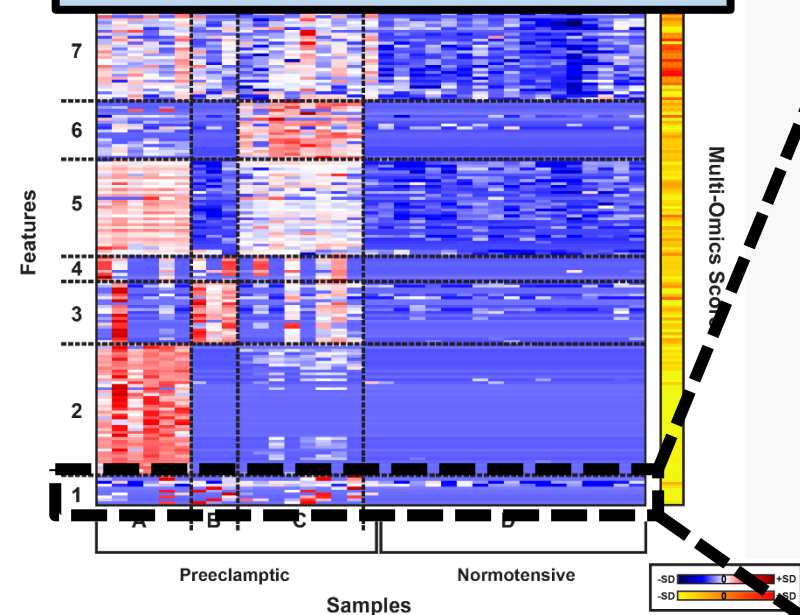
# Visualizations of Data – Cluster 1

## Chemical Candidate Information



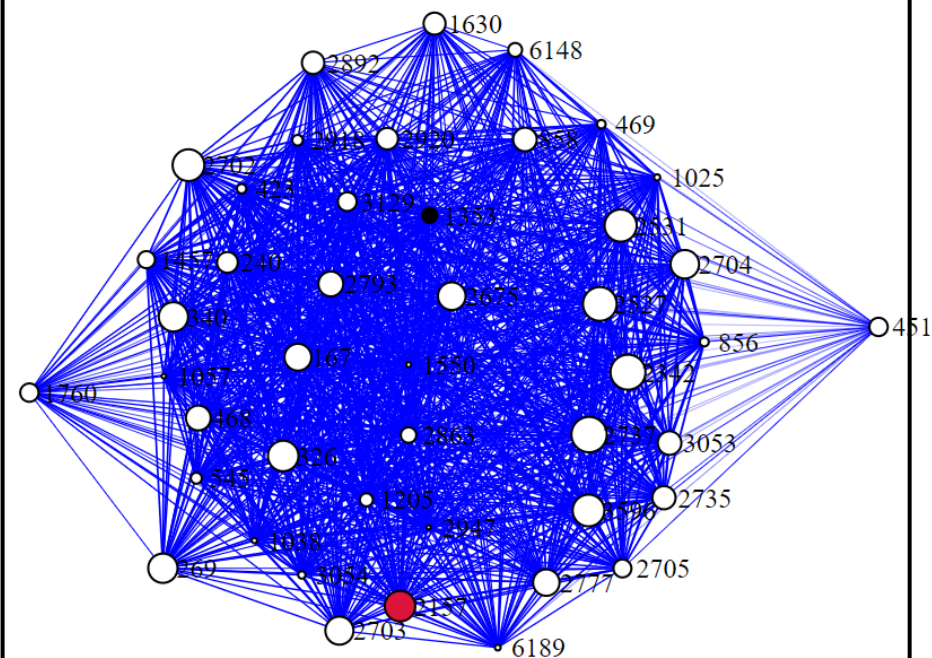
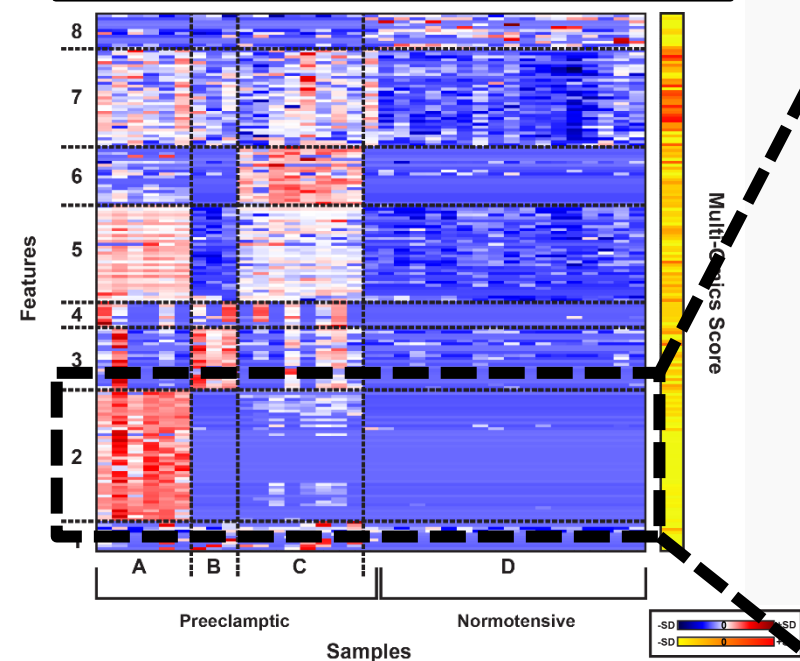
# Patterns of Cluster Results

**Cluster 1: Primarily drugs  
(e.g. diphenhydramine,  
labetalol, clindamycin)**

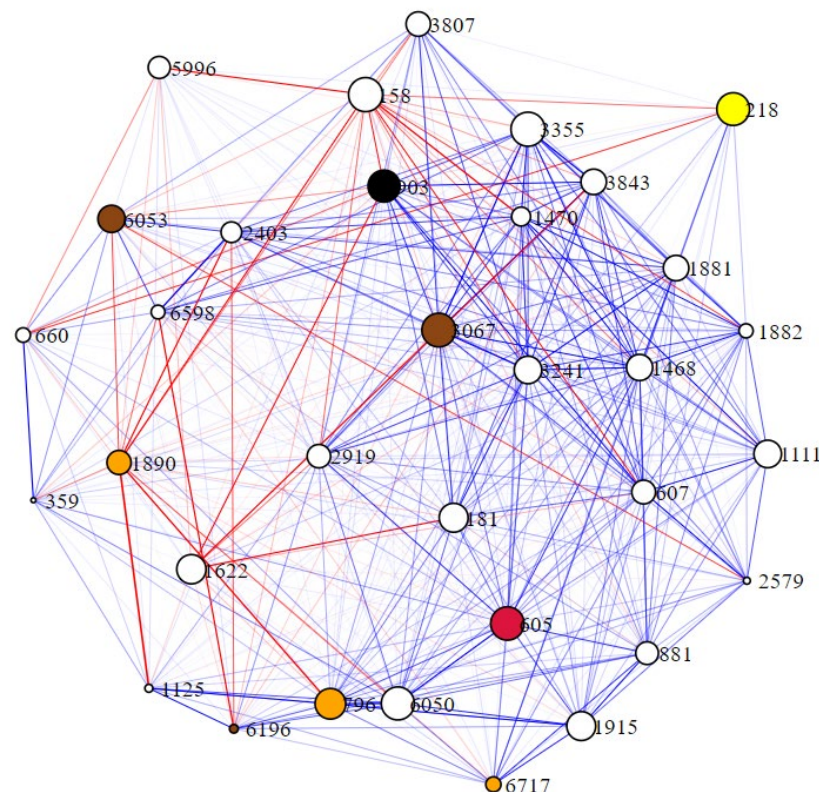


# Patterns of Cluster Results

## Cluster 2: Likely xenobiotics (ethanolamides)



Heatmap visualization showing the Multi-Omics Score across samples. The y-axis represents samples 1 through 7. The x-axis is divided into Preeclamptic (A, B, C) and Normotensive (D) samples. A color scale on the right indicates the Multi-Omics Score, ranging from -SD (blue) to +SD (red). A dashed line separates the Preeclamptic and Normotensive groups. A dashed line also separates the samples into two groups: 1-3 and 4-7.





# 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

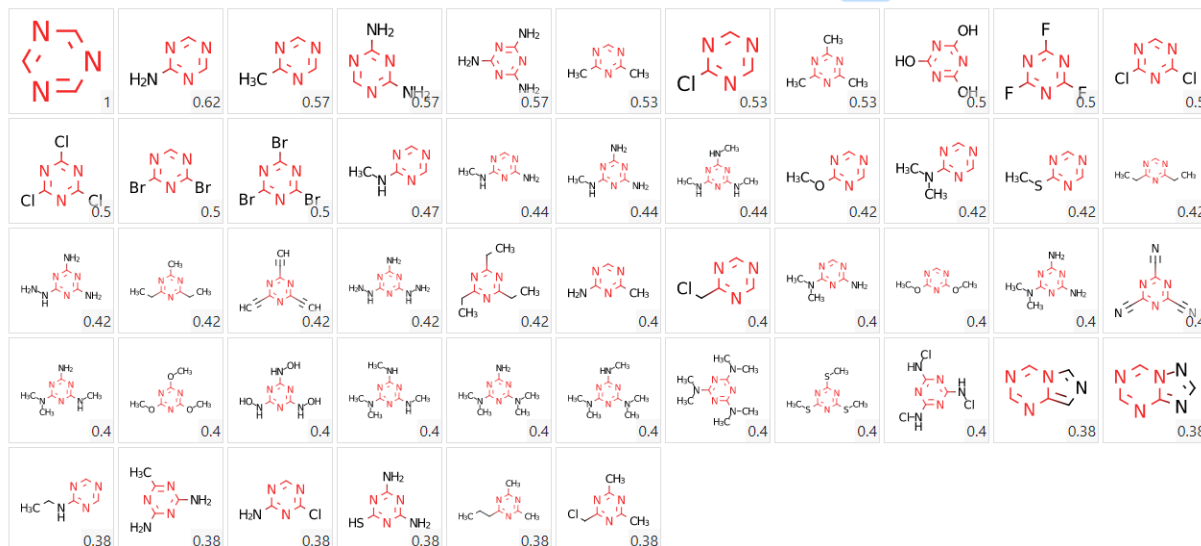
☐ Charged

☐ Salts or Mixtures

Sort

Similarity

≡



Search result 2540

Show

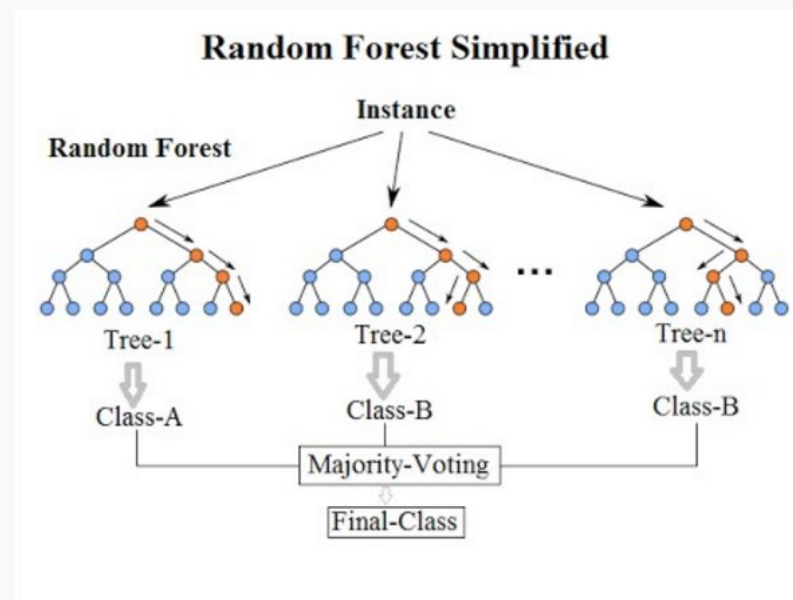
☐ Isotopically Labeled

# Method Amenability Prediction

## Charlie Lowe

### Why?

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

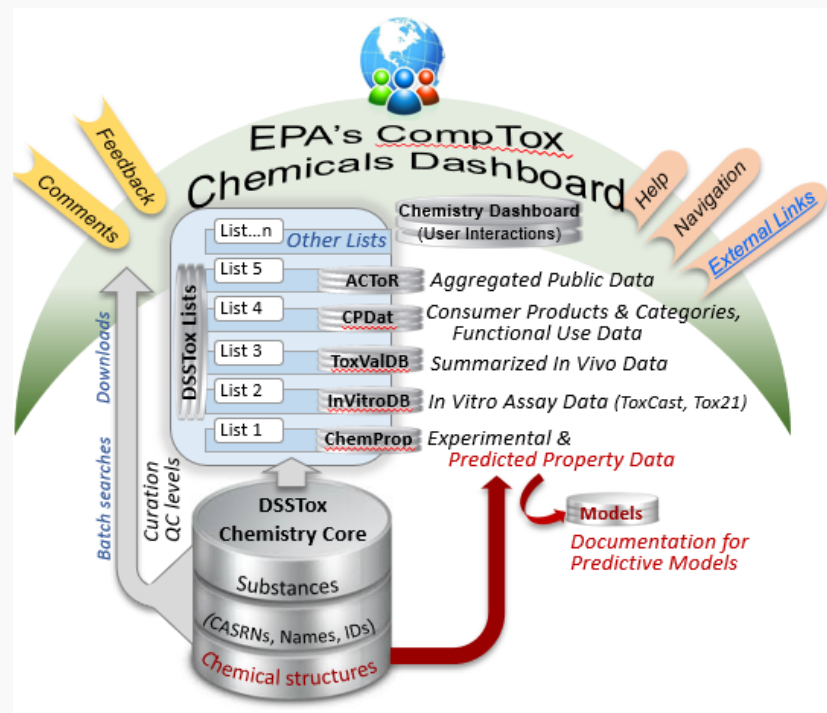


- Data sources to date
  - **Massbank of North America**
    - 9,275 chemicals for non-derivatized GC
    - 846 chemicals for derivatized GC
    - 816 chemicals for APCI+
    - 454 chemicals for APCI-
    - 4,907 chemicals for ESI+
    - 3,430 chemicals for ESI-
  - **EPA Non-targeted Analysis Collaborative 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-



# Conclusion

- Dashboard access to data for ~883,000 chemicals
- 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



- *NCCT IT development team*
- *Tommy Cathey, ACTOR Web Services*
- *Nancy Baker, Abstract Sifter*
- *Todd Martin & Valery Tkachenko, WebTEST*
- *Kathie Dionisio & Kristin Isaacs, CPDat*
- *Thanks to Emma Schymanski, University of Luxembourg, for coordinating all efforts with the NORMAN Network for curation of lists on the Suspect Exchange*

## Antony Williams

US EPA Office of Research and Development  
National Center for Computational Toxicology

**EMAIL:** [Williams.Antony@epa.gov](mailto:Williams.Antony@epa.gov)

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

[RETURN TO ISSUE](#) | [< PREV](#) **SUSTAINABLE SYSTEMS** [NEXT >](#)

### Chemical Characterization of Recycled Consumer Products Using Suspect Screening Analysis

Charles N. Lowe, Katherine A. Phillips, Kristin A. Favela, Alice Y. Yau, John F. Wambaugh, Jon R. Sobus, Antony J. Williams, Ashley J. Pfirman, and Kristin K. Isaacs\*

✓ **Cite this:** *Environ. Sci. Technol.* 2021, 55, 16, 11375–11387

Publication Date: August 4, 2021 ▾

<https://doi.org/10.1021/acs.est.1c01907>

Copyright © 2021 American Chemical Society

[RIGHTS & PERMISSIONS](#) ✓ Subscribed

Article Views | Altmetric | Citations

530

16

-

[LEARN ABOUT THESE METRICS](#)

Share Add to Export

