

# Using the US EPA's CompTox Chemistry Dashboard for structure identification and non-targeted analyses

*Antony Williams<sup>1</sup>, Andrew D. McEachran<sup>3</sup>, Seth Newton<sup>2</sup>,  
Kristin Isaacs<sup>2</sup>, Katherine Phillips<sup>2</sup>, Nancy Baker<sup>1</sup>,  
Chris Grulke<sup>1</sup> and Jon R. Sobus<sup>2</sup>*

*1) National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC*

*2) National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*

*3) Oak Ridge Institute of Science and Education (ORISE) Research Participant, Research Triangle Park, 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*

*March 2018  
ACS Spring Meeting, New Orleans*

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

# CompTox Chemistry Dashboard

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



Home

Advanced Search

Batch Search

Lists

Predictions

Downloads

## Chemistry Dashboard

Aa Aa Aa



761 Thousand Chemicals

Search a chemical by systematic name, synonym, CAS number, or InChIKey



☐ Identifier substring search



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

## Latest News

[Read more news](#)

### A Movie Regarding how to Identify "Known Unknowns" Using the CompTox Dashboard

March 28th, 2017 at 7:35:41 PM

Recently we published a paper regarding [Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard](#), Analytical and Bioanalytical Chemistry, March 2017, Volume 409, Issue 7, pp 1729–1735. A movie explaining the paper in full animated detail has been put on YouTube. Enjoy the movie interlude [here](#).



### Discover.

[About/Disclaimer](#)  
[Accessibility](#)

### Connect.

[ACToR](#)  
[DSTox](#)

### Ask.

[Contact](#)  
[Help](#)



United States  
Environmental Protection  
Agency

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

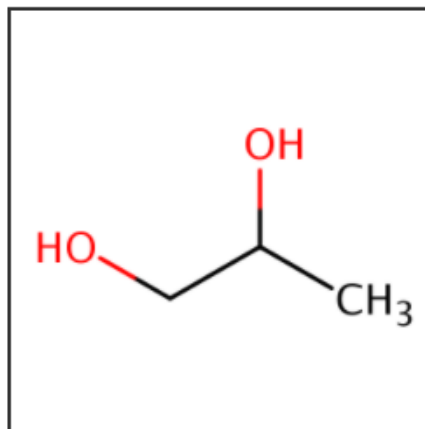
## Chemistry Dashboard | EPAHFR

[Submit Comment](#)[Copy](#)[Aa](#)[Aa](#)[Aa](#)

### 1,2-Propylene glycol

57-55-6 | DTXSID0021206

© Searched by DSSTox\_Substance\_Id: Found 1 result for 'DTXSID0021206'.



#### Wikipedia

Propylene glycol (IUPAC name: propane-1,2-diol) is a synthetic organic compound with the chemical formula C<sub>3</sub>H<sub>8</sub>O<sub>2</sub>. It is a viscous colorless liquid which is nearly odorless but possesses a faintly sweet taste. Chemically it is classed as a diol and is miscible with a broad range of solvents, including water, acetone, and chloroform. It is produced on a large scale and is primarily used in the production of polymers, but also sees use in food...[Read more](#)

#### Intrinsic Properties

#### Structural Identifiers

#### Linked Substances

#### Presence in Lists

#### Record Information

#### Quality Control Notes

[Chemical Properties](#)[Env. Fate/Transport](#)[Hazard](#)[ADME \(Beta\)](#)[Exposure](#)[Bioassays](#)[Similar Compounds](#)[Related Substances](#)[Synonyms](#)[Literature](#)[Links](#)[Comments](#)

# Access to Chemical Hazard Data



## Chemistry Dashboard | EPAHFR

[Submit Comment](#)[Copy](#)[Aa](#)[Aa](#)[Aa](#)[Chemical Properties](#)[Env. Fate/Transport](#)[Hazard](#)[ADME \(Beta\)](#)[Exposure](#)[Bioassays](#)[Similar Compounds](#)[Related Substances](#)[Synonyms](#)[Literature](#)[Links](#)[Comments](#)[Exposure Limit](#)[Lethality Effect Level](#)[Point of Departure](#)[Toxicity Value](#)

Download table as:

[TSV](#)[Excel](#)[Human](#)[Eco](#)

	Priority	Type	Subtype	Risk Assessment Class	Values	Units	Study Type	Exposure Route	Species	Subsource	Source
+	8	NOEL	Cardiova...	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	NOEL	Endocrine	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	LOEL	Hematol...	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	LOEL	Hepatic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	NOEL	Immune	immunot...	5000.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	NOEL	Renal	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	LOEL	Systemic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaill...	PPRTV (...)
+	8	NOEL	Hematol...	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaill...	PPRTV (...)
+	8	NOEL	Systemic	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaill...	PPRTV (...)

# Sources of Exposure to Chemicals

## Product & Use Categories

Chemical Weight Fraction

Chemical Functional Use

Monitoring Data

Exposure Predictions

Production Volume

Exposure

Bioassays

Similar Compounds

Related Substances




Synonyms

Literature

Links

Comments

### Product & Use Categories (PUCs)

 <a href="#">Categorization type</a>	 <a href="#">Number of Unique Products</a> 
PUC	288
PUC	208
PUC	117
PUC	107
PUC	107
PUC	101
PUC	101
PUC	90
PUC	89

- Structure Identification using the dashboard
  - Formula/mass-based searching – 1 chemical at a time


# Advanced Searches

## Advanced Search

### Mass Search

$\pm$

Min/Max

M 

Mass


Da

$\pm$

Error


Da


ppm


Search 

### Molecular Formula Search

Molecular Formula

☒ MS Ready Formula 

☐ Exact Formula 

Search 

### Generate Molecular Formula(e)

$\pm$

Min/Max

Mass


Da

$\pm$

Error


Da

ppm

Search 

Default Options: C[1-50] H[0-100] O[0-20] N[0-20] P[0-20] S[0-10]

Include Halogens: ☐ F[0-20] ☐ Cl[0-20] ☐ Br[0-20] ☐ I[0-20]

Options 



# Advanced Searches

## Mass Based Search

### Mass Search

▼

# Advanced Searches

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

## Chemistry Dashboard

Search Results  
Searched by Mass: '191.081 +/- 5 ppm'.  
298 of 298 chemicals visible

Download / Send Sort by: Mass Difference


Hide: **Multicomponent Chemicals**

 <b>DEET</b> 134-62-3	 <b>Phendimetrazine</b> 634-03-7	 <b>N-Butylacetanilide</b> 91-49-6	 <b>Benzaldehyde, 4-(diethylamino)-...</b> 92-14-8	 <b>Acetanilide, 2',6'-diethyl-...</b> 16665-89-7	 <b>Azetidine, 1,3-dimethyl-3-(m-met...</b> 19832-26-9
 <b>Benzamide, N-pentyl-...</b> 20308-43-4	 <b>p-t-Butylacetanilide</b> 20330-45-4	 <b>N,N-Diethylphenylacetamide</b> 2431-96-1	 <b>3-(Dimethylamino)-2-methylpropi...</b> 26171-50-6	 <b>Butyramide, 2-ethyl-2-phenyl-...</b> 30568-39-9	 <b>1-Heptanone, 1-(4-pyridyl)-...</b> 32941-30-3


# Formula Searches

## Molecular Formula Search

C<sub>12</sub>H<sub>17</sub>NO

☒ MS Ready Formula 

☐ Exact Formula 

Search 

# Exact Formula Search: C<sub>12</sub>H<sub>17</sub>NO

## 298 Chemicals

**EPA** United States Environmental Protection Agency

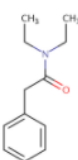
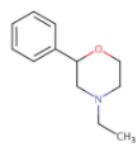
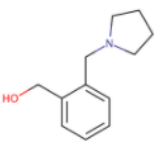
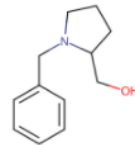
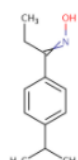
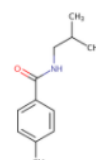
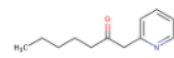
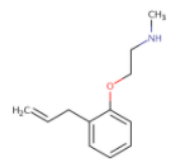
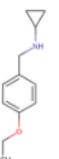
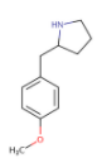
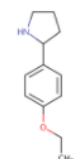
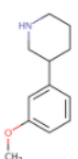
Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

### Chemistry Dashboard

Search Results  
Searched by Exact Molecular Formula: 'C<sub>12</sub>H<sub>17</sub>NO'.

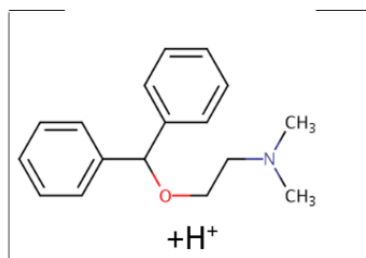
Download / Send Sort by: DTXSID 298 of 298 chemicals visible Hide: Multicomponent Chemicals Select all

 <p>N,N-Diethylphenylacetamide 2431-96-1</p>	 <p>MLS002639407 83081-04-3</p>	 <p>{2-[(Pyrrolidin-2-yl)methyl]phenyl}methanol 91271-58-8</p>	 <p>(1-benzylpyrrolidin-2-yl)methanol 67131-44-6</p>	 <p>1-Propanone, 1-[4-(1-methylethyl)phenyl] 78575-14-1</p>	 <p>4-methyl-N-(2-methylpropyl)benzamide 88358-24-1</p>
 <p>2-Heptanone, 1-(2-pyridinyl) 21211-98-3</p>	 <p>2-(2-ALLYLPHENOXY)-N-METHYLPYRROLIDINE 57162-94-4</p>	 <p>N-(4-ethoxybenzyl)cyclopropanamine 892571-13-0</p>	 <p>2-[(4-methoxyphenyl)methyl]pyrrolidine 66162-38-7</p>	 <p>2-(4-ethoxyphenyl)pyrrolidine 383127-28-4</p>	 <p>3-(3-methoxyphenyl)piperidine 79601-21-1</p>

- Structure Identification using the dashboard
  - Formula/mass-based searching – 1 chemical at a time
  - Distilling structures into “MS-Ready form”

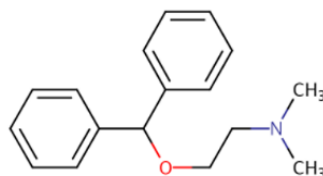
# Specific Data-Mappings “MS-Ready Structures”

A) Molecular Ion



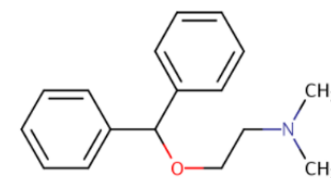
$m/z \approx 256.1702$

B) MS-Ready Form

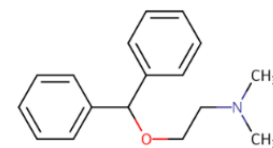


monoisotopic mass= 255.1623  
C17H21NO  
DTXCID802949

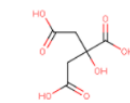
C) Mappings from MS-Ready



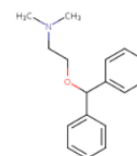
Diphenhydramine  
C17H21NO | 255.1623  
DTXSID4022949



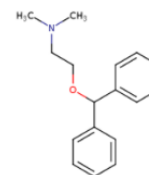
Diphenhydramine  
hydrochloride  
C17H22ClNO | 291.1390  
DTXSID4020537



Diphenhydramine citrate  
C23H29NO8 | 447.1893  
DTXSID80237211



Diphenhydramine salicylate  
C24H27NO4 | 393.1940  
DTXSID10225883



# Diphenhydramine

## 15 Total MS-Ready Mappings

**EPA** United States Environmental Protection Agency


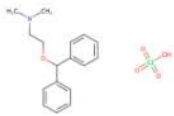
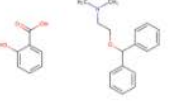

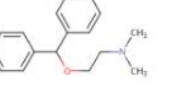
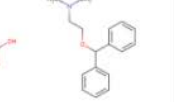
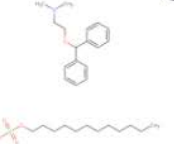
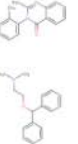
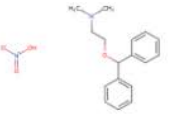
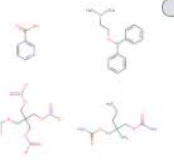
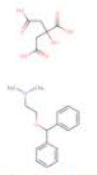
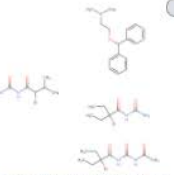
Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

### Chemistry Dashboard

MS-Ready Mappings of Diphenhydramine (Isotopes pre-filtered)

Download / Send Sort by: DTXSID 15 of 15 chemicals visible Hide: Isotopes Select all

 Dampa D 101052-67-9	 Perchloric acid-2-(diphenylmethoxy)-2-methylpropane 17626-30-1	 Diphenhydramine salicylate 7491-10-3	 Diphenhydramine hydrochloride 147-24-0	 Diphenhydramine 58-73-1	 Acetic acid-2-(diphenylmethoxy)-2-methylpropane 840517-01-3
 Diphenhydramine laurylsulfate 103659-13-8	 Diphenhydramine hydrochloride 8076-99-1	 Nitric acid-2-(diphenylmethoxy)-2-methylpropane 17626-28-7	 Visano cor 81246-67-5	 Diphenhydramine citrate [USP] 88637-37-0	 n-(acetylcarbamoyl)-2-bromo-2-methylpropane 52232-04-9

# “MS Ready” Formula Search C<sub>12</sub>H<sub>17</sub>NO

## 354 Chemicals

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

### Chemistry Dashboard

Search Results  
Searched by MS Ready Formula: 'C<sub>12</sub>H<sub>17</sub>NO'

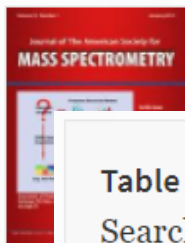
Download / Send Sort by: DTXSID 354 chemicals Hide: Select all

 N,N-Diethylphenylacetamide 2431-96-1	 MLS002639407 83081-04-3	 {2-[(Pyrrolidin-2-yl)methyl]phenyl}methanol 91271-58-8	 (1-benzylpyrrolidin-2-yl)methanol 67131-44-6	 1-Propanone, 1-[4-(1-methylethyl)phenyl] 78575-14-1	 4-methyl-N-(2-methylpropyl)benzamide 88358-24-1
 2-Heptanone, 1-(2-pyridinyl)- 21211-98-3	 2-(2-ALLYLPHENOXY)-N-METHYLPYRROLIDINE 57162-94-4	 N-(4-ethoxyphenyl)cyclopropanamine 892571-13-0	 2-[(4-methoxyphenyl)methyl]pyrrolidine 66162-38-7	 2-(4-ethoxyphenyl)pyrrolidine 383127-28-4	 3-(3-methoxyphenyl)piperidine 79601-21-1



- Structure Identification using the dashboard
  - Formula/mass-based searching – 1 chemical at a time
  - Distilling structures into “MS-Ready form”
  - Ranking based on metadata

# Identifying Known Unknowns by reference ranking



[Journal of The American Society for Mass Spectrometry](#)  
January 2012, Volume 23, [Issue 1](#), pp 179-185 | [Cite as](#)

**Table 1**

Searching ChemSpider by Elemental Composition then Sorting by Number of Associated References

Class of compounds	Number compounds in class	Position of compound sorted in descending order by number of references					
		#1	#2	#3	#4	#5	>#5
Drugs	45	43	1	1			
Pesticides	8	7	1				
Toxins	2	2					
Polymer antioxidants	15	15					
Polymer UV stabilizers	10	8	1	1			

# Data source ranking using the Dashboard

Anal Bioanal Chem  
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>

	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

# Additional Metadata Ranking

- US EPA CompTox Chemistry Dashboard Data Sources
- “CPDat” Consumer Product Database
- PubChem Data Source Count
- PubMed Reference Count

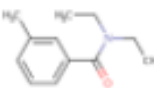
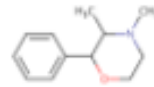
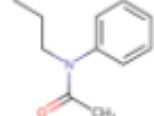
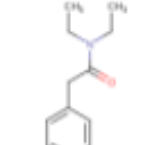
# Additional Metadata Ranking

## C12H17NO: 354 Chemicals

Chemistry Dashboard									
Search Results									
Searched by MS Ready Formula: 'C12H17NO'.									
354 chemicals									
Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source	Monoisotopic Mass
	<a href="#">DTXSID2021995</a>	DEET	134-62-3	Level 1	111	104	155	753	191.131014
	<a href="#">DTXSID1023447</a>	Phendimetrazine	634-03-7	Level 2	12	27	35	50	191.131014
	<a href="#">DTXSID2042197</a>	N-Butylacetanilide	91-49-6	Level 2	1	26	50	1	191.131014
	<a href="#">DTXSID00179048</a>	N,N-Diethylphenylacetamide	2431-96-1	Level 4	0	18	52	34	191.131014
	<a href="#">DTXSID60865298</a>	N,N-Diethyl-2-methylbenzamide	2728-04-3	Level 1	0	11	49	0	191.131014
	<a href="#">DTXSID4016123</a>	4-Aminobenzonitrile	28237-75-2	Level 4	0	11	26	1	101.121014

# Additional Metadata Ranking

## C12H17NO: 354 Chemicals

Structure	DTXSID	Preferred Name	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source
	<a href="#">DTXSID2021995</a>	DEET	111	104	155	753
	<a href="#">DTXSID1023447</a>	Phendimetrazine	12	27	35	50
	<a href="#">DTXSID2042197</a>	N-Butylacetanilide	1	26	50	1
	<a href="#">DTXSID00179048</a>	N,N-Diethylphenylacetamide	0	18	52	34

# Top Ranked Chemical

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search All Data

Chemistry Dashboard

Submit Comment Share Copy Aa Aa Aa

## DEET

134-62-3 | DTXSID2021995

Searched by DSSTox\_Substance\_Id: Found 1 result for 'DTXSID2021995'.

Chemical structure of DEET (N,N-Diethyl-m-tolamide) is shown. It consists of a benzene ring with a methyl group (H<sub>3</sub>C) at the meta position and a diethylcarbamoyl group (-C(=O)N(CH<sub>2</sub>CH<sub>3</sub>)<sub>2</sub>) at the 1 position.

### Wikipedia

N,N-Diethyl-m-tolamide, also called DEET () or diethyltoluamide, is the most common active ingredient in insect repellents. It is a slightly yellow oil intended to be applied to the skin or to clothing and provides protection against mosquitoes, ticks, fleas, chiggers, leeches and many biting insects. ...[Read more](#)

### Intrinsic Properties

### Structural Identifiers

### Linked Substances

### Presence in Lists

### Record Information

### Quality Control Notes

Executive Summary (Beta) Chemical Properties Env. Fate/Transport Hazard ADME (Beta) Exposure Bioassays Similar Compounds Related Substances Synonyms Literature Links Comments

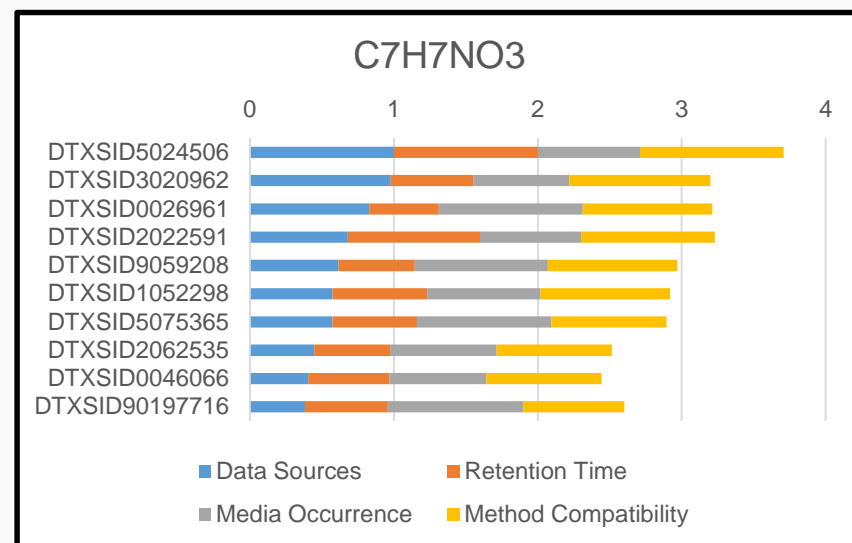
Executive Summary

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

22

# Additional data streams in development

- US EPA CompTox Chemistry Dashboard Data Sources
- “CPDat” Consumer Product Database
- PubChem Data Source Count
- PubMed Reference Count
- Retention Time Prediction  $SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$
- Predicted Environmental Media Occurrence
- Presence in Lists





# “Chemicals Detected in Water”

## Chemistry Dashboard

Aa ▼ Aa Aa ▲

### Select List


List Name	Number of Chemicals	List Description
<a href="#">Drinking Water Suspects, KWR Water, Netherlands</a>	136	KWRSJERPS is a list of prioritized suspects relevant for human health in drinking water from KWR Water in Nieuwegein, The Netherlands. The methods are detailed in Sjerps et al 2016, DOI: 10.1016/j.watres.2016.02.034
<a href="#">EPA Consumer Products Suspect Screening Results</a>	1705	This is a compiled list of the suspects reported in the supporting information of Phillips et al 2018, DOI: 10.1021/acs.est.7b04781 - Suspect Screening Analysis of Chemicals in Consumer Products with GCxGC-TOF/MS.
<a href="#">EPA Integrated Risk Information System (IRIS)</a>	510	EPA's IRIS Program identifies and characterizes the health hazards of chemicals found in the environment. Each IRIS assessment can cover a chemical, a group of related chemicals, or a complex mixture.
<a href="#">EPAHFR - EPA Chemicals associated with hydraulic fracturing</a>	1640	EPAHFR lists chemicals associated with hydraulic fracturing from 2005-20013, as reported in EPA's Hydraulic Fracturing Drinking Water Assessment Final Report (Dec 2016)
<a href="#">STOFF-IDENT Database of Water-Relevant Substances</a>	8885	STOFF-IDENT is a database of water relevant substances collated from various sources within the STOFF-IDENT and FOR-IDENT projects, hosted by LfU, HSWT and TUM. The database at <a href="https://www.lfu.bayern.de/stoffident/#!home">https://www.lfu.bayern.de/stoffident/#!home</a> has additional functionali...
<a href="#">Superfund Chemical Data Matrix</a>	220	The Superfund Chemical Data Matrix (SCDM) generates a list of the corresponding Hazard Ranking System (HRS) factor values, benchmarks, and data elements for a particular chemical.
<a href="#">Surfactant List Screened in Swiss Wastewater (2014)</a>	122	EAWAGSURF is a list of surfactants screened in Swiss wastewater effluents as part of a 2014 study. Structures/mixtures are being progressively curated and linked (Schymanski/Williams). Further details in Schymanski et al 2014, DOI: 10.1021/es4044374

- Structure Identification using the dashboard
  - Formula/mass-based searching – 1 chemical at a time
  - Distilling structures into “MS-Ready form”
  - Ranking based on metadata
  - Batch searching of formulae and masses

# Batch Search








## Batch Search ?

Step One      Step Two      Step Three      Step Four      Step Five      Step Six



**Step One: Select Input**

**Select Input Type(s)**

- ☐ Chemical Name 
- ☐ CASRN 
- ☐ InChIKey  ☐ Skeletal ID 
- ☐ DSSTox Substance 
- ☐ MS-Ready Formula(e) 
- ☐ Exact Formula(e) 
- ☐ Monoisotopic Mass








×

This search is based on what we refer to as 'Mass Spec Ready Formulae'. All chemicals within the database are treated in a manner that all are desalted and stereochemistry is removed as Mass Spectrometry detects the major components of a salt and is insensitive to stereochemistry. As an example, a search for the formula associated with phenol will return phenol, sodium phenolate and calcium phenoxide.

(Search results should be limited to <1000 identifiers)

# Batch Search

## Select Input Type(s)










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



Enter Identifiers to Search (searches should be limited to <1000 identifiers)

C6H12O3  
C7H7N3  
C8H11NO  
C7H5NOS  
C9H15NO  
C11H12O  
C9H8O3  
C6H12O5  
C9H15NO2

## Metadata

- ☐ Curation Level Details 
- ☒ Data Sources 
- ☒ Assay Hit Count 
- ☐ Include links to ACToR reports - SLOW! (BETA) 
- ☒ NHANES/Predicted Exposure 
- ☒ Include ToxVal Data Availability 
- ☒ Number of PubMed Articles 
- ☐ Abstract Sifter Input File (Beta) 
- ☐ MetFrag Input File(Beta)
- ☒ IRIS
- ☒ PPRTV
- ☒ PubChem Data Sources
- ☐ ToxPrint fingerprints 


- ☐ NIOSH IDLH Values
- ☐ NIOSH International Chemical Safety Cards
- ☐ NIOSH Pocket Guide to Chemical Hazards
- ☐ NIOSH Skin Notation Profiles
- ☐ NORMAN Collaborative Trial 2015 Targets and Suspects
- ☐ Norman Network PFAS (KEMI Report)
- ☐ NORMAN Network Priority List
- ☐ NormaNEWS: Norman Early Warning System
- ☐ PFAS list provided by X.Trier et al
- ☐ Pharmaceutical List with EU, Swiss and US Consumption Data
- ☐ Provisional Peer Reviewed Toxicity Values
- ☐ Stockholm Convention on Organic Pollutants
- ☒ STOFF-IDENT Database of **Water**-Relevant Substances
- ☐ Superfund Chemical Data Matrix
- ☒ Surfactant List Screened in Swiss Wastewater (2014)

# Excel Output

INPUT	FOUND_BY	DTXCID_IN	DATA_SOURCE	TOXVAL_D	TOXCAST	TOXCAST	NUMBER_C	PUBCHEM	STO
C6H12O3	MS Ready	<a href="#">DTXCID701</a>	51	Y	0.36	2/562	24	83	Y
C6H12O3	MS Ready	<a href="#">DTXCID003</a>	67	Y	0.36	1/276	376	80	Y
C6H12O3	MS Ready	<a href="#">DTXCID106</a>	65	Y	4.42	5/113	6	77	Y
C6H12O3	MS Ready	<a href="#">DTXCID105</a>	45	Y	0.0	0/163	3	94	-
C6H12O3	MS Ready	<a href="#">DTXCID901</a>	38	Y	-	-	14	110	Y
C6H12O3	MS Ready	<a href="#">DTXCID402</a>	34	Y	0.0	0/113	-	53	Y
C6H12O3	MS Ready	<a href="#">DTXCID202</a>	31	Y	-	-	-	36	Y
C6H12O3	MS Ready	<a href="#">DTXCID202</a>	30	-	2.54	7/276	-	54	-
C6H12O3	MS Ready	<a href="#">DTXCID109</a>	26	Y	-	-	-	46	-
C6H12O3	MS Ready	<a href="#">DTXCID202</a>	24	Y	0.0	0/113	-	47	-
C6H12O3	MS Ready	<a href="#">DTXCID303</a>	22	Y	-	-	-	89	-
C6H12O3	MS Ready	<a href="#">DTXCID302</a>	20	Y	-	-	2	25	Y
C6H12O3	MS Ready	<a href="#">DTXCID407</a>	19	Y	-	-	12	62	-
C6H12O3	MS Ready	<a href="#">DTXCID704</a>	17	Y	-	-	-	64	-
C6H12O3	MS Ready	<a href="#">DTXCID704</a>	16	Y	-	-	3	49	-

# Batch Search Integration to MetFrag

<http://c-ruttkies.github.io/MetFrag/projects/metfragweb/>



## MetFrag

In silico fragmentation for computer assisted identification of metabolite mass spectra

1. Candidate database

2a. Neutral mass of the precursor molecule

3. Relative mass deviation for candidate selection

4. Neutral molecular formula of the precursor (if available)

5. Database specific candidate identifiers

2b. Calculate neutral precursor mass from charged ion mass

Database:

Neutral Mass:  Search ppm:

Formula:

Identifiers:

Parent Ion:

Candidate Filter & Score Settings

Fragmentation Settings & Processing

Mzppm:

Mzabs:

Mode:

MS/MS Peak list

90.97445 681

106.94476 274

110.02750 110

115.98965 95

117.98540 384

124.93547 613


124.99015 146

125.99793 207

133.95592 777

143.98846 478

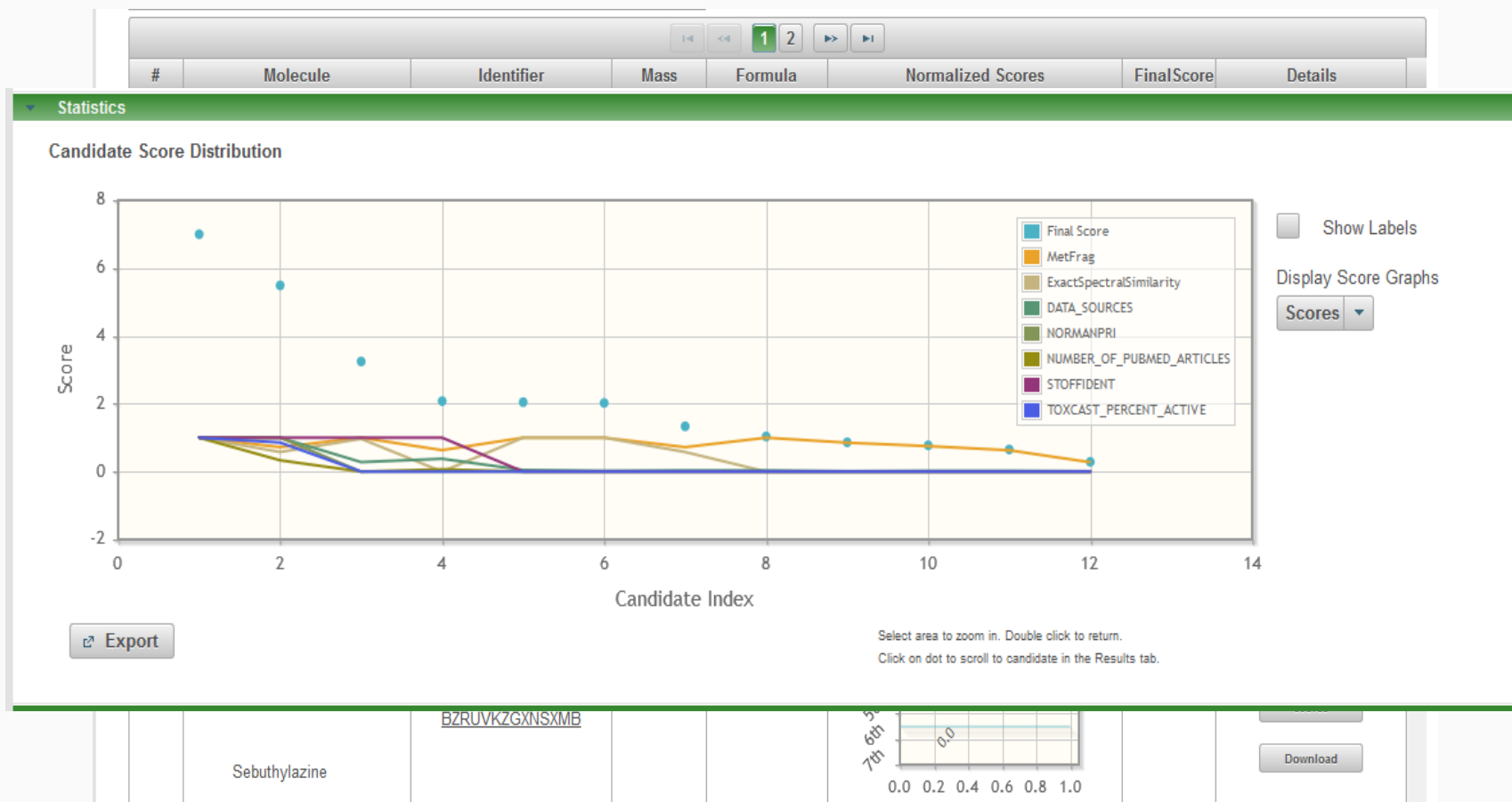
144.99625 352

 Metadata

- ☐ Curation Level Details
- ☒ Data Sources
- ☒ Assay Hit Count
- ☐ Include links to ACToR reports - SLOW! (BETA)
- ☒ NHANES/Predicted Exposure
- ☒ Include ToxVal Data Availability
- ☒ Number of PubMed Articles
- ☐ Abstract Sifter Input File (Beta)
- ☒ MetFrag Input File(Beta)
- ☒ IRIS
- ☒ PPRTV
- ☒ PubChem Data Sources

# Batch Search Integration to MetFrag

<http://c-ruttikies.github.io/MetFrag/projects/metfragweb/>





# The Dashboard to Support MS-Analysis

## MS-Ready Structures Underpin Analysis

### Mass Search

±

Min/Max

M

Mass

Da

±


Error

Da

ppm

### Molecular Formula Search

Molecular Formula

☒ MS Ready Formula 

☐ Exact Formula 

### Generate Molecular Formula(e)

±

Min/Max

Mass

Da

Step One

Step Two

Step Three

Step Four

Step Five

Step Six

Step Five: Choose Data Fields to Download


Default Options: C[1-50] H[0-100] O[0-20] N[0-

Include Halogens: ☐ F[0-20] Cl[0-20] Br[0-20]


Options ▼

#### Select Input Type(s)

☐ Chemical Name 

☐ CASRN 

☐ InChIKey  ☐ Skeleton 

☐ DSSTox Substance ID 

☒ MS-Ready Formula(e) 

☐ Exact Formula(e) 

☐ Monoisotopic Mass

Display All Chemicals

Download Chemical Data

Enter Identifiers to Search (searches should be limited to <1000 identifiers)

C14H22N2O3  
C10H12N2O  
C14H18N4O3  
C12H11N7  
C8H9NO2

# Downloadable Data



## Chemistry Dashboard

[Aa ▼](#)[Aa](#)[Aa ▲](#)

## Downloads

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

Posted: 11/14/2016

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

SID	CID	DTXSID
316388891	20404	DTXSID30873143
316388890	10142816	DTXSID70873142
316388889	50742127	DTXSID40873139
316388888	19073841	DTXSID20873137
316388887	11505215	DTXSID00873135
316388886	25021861	DTXSID80873133
316388885	2784427	DTXSID60873131
316388884	6731	DTXSID00873130

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

Posted: 11/14/2016

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

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

# Future Work: Combined Substructure/Formula Searching

atrazine Search

100%

Select properties to predict

**T.E.S.T.** 18 OPERA Search

☐ Exact  
☐ Substructure  
☐ Similarity  
☒ Molecular Formula  
☐ Molecular Weight


Input formula (e.g. C6 H6):  
C15H16O2 Search

Search result **5** Show ☐ Isotopically Labeled ☐ Chiral


Elements per page 50 1

<https://comptox.epa.gov/dashboard/DTXSID7020182>

# Future Work: Searching Against Predicted Spectra

 CFM-ID

☆ Utilities ▾ Help Data Publications Contact Us



## CFM-ID

Competitive Fragmentation Modeling for Metabolite Identification

### Spectra Prediction

---

Predicts the spectra for a given input molecule. Spectra are computed for low (10V), medium (20V) and high (40V) collision energy levels and are represented by a list of 'mass intensity' pairs, each corresponding to a peak in the spectra.

**Parent Compound Structure**  
InChI or SMILES format

InChI strings need to start with "InChI=" and are not expected to have any charge - an additional H+ will be added. Maximum compound size is 200 atoms. Load an [InChI example](#), [SMILES example](#), or [another SMILES example](#).

**Spectra Type**

**Ion Mode**

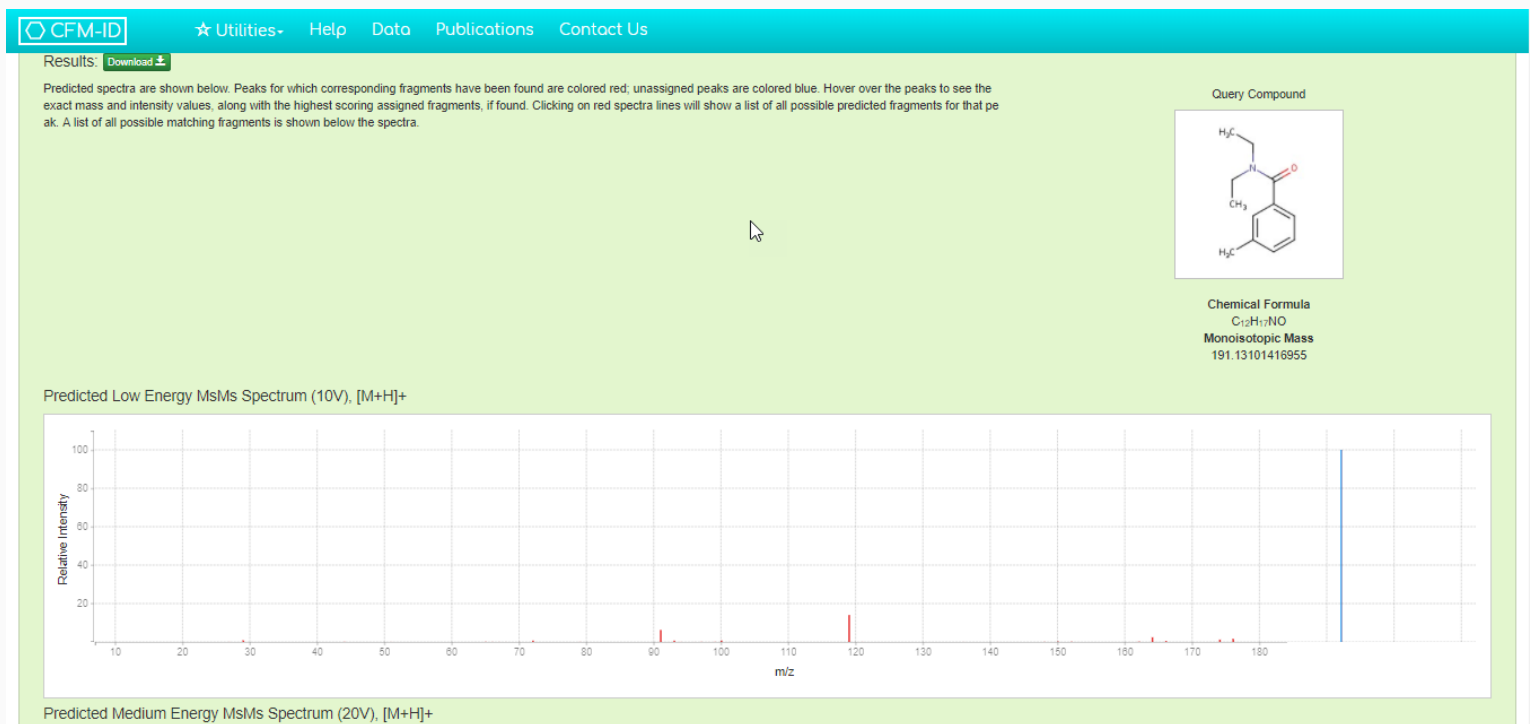
**Adduct Type**

If you wish to run multiple jobs, input larger query molecules, or customize the computation parameters, you can freely download the source code here: <http://sourceforge.net/projects/cfm-id>.

cfmid.wishartlab.com

# Future Work: Searching Against Predicted Spectra

- CFM-ID predicted spectra generated for 700,000 chemicals
  - Positive ion, Negative ion, Electron Impact
  - Three energies



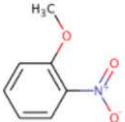

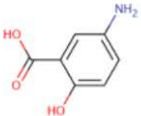

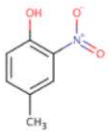

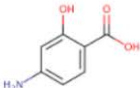

# Future Work

## Scoring scheme into results

### Search Results

🔍 Searched by molecular formula: Found 188 results.

Download as: TSV Excel SDF

ID	Structure	Preferred Name	CAS-RN	QC Level	CPCat Count	Number of S...	PubChem D...	Monoisotopic Mass	Identification Score
DTXSID3020962 ToxCast™		1-Methoxy-2-nitrobenzene	91-23-6	Level 2: Expert curate...	0	58	111	153.042593	<div> <span>Data Sources</span> <span>PubMed Ct</span> <span>Media Occurr</span> <span>CPDat C</span> </div>  3.6
DTXSID5024506 ToxCast™		5-Aminosalicylic acid	89-57-6	Level 2: Expert curate...	2	57	193	153.042593	<div> <span>Data Sources</span> <span>PubMed Ct</span> <span>Media Occurr</span> <span>CPDat C</span> </div>  3.3
DTXSID0026961 ToxCast™		4-Methyl-2-nitrophenol	119-33-5	Level 2: Expert curate...	0	46	93	153.042593	<div> <span>Data Sources</span> <span>PubMed Ct</span> <span>Media Occurr</span> <span>CPDat C</span> </div>  3.3
DTXSID2022591		Aminosalicic acid	65-49-6	Level 1: Expert curate...	0	40	177	153.042593	<div> <span>Data Sources</span> <span>PubMed Ct</span> <span>Media Occurr</span> <span>CPDat C</span> </div>  2.5

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$$

- The CompTox Chemistry Dashboard provides access to data for ~760,000 chemicals
- High quality curated data and rich metadata facilitates mass spec analysis
- “MS-Ready” processed data enables structure identification

- The CompTox Chemistry Dashboard team
- NERL colleagues:
  - Jon Sobus, Elin Ulrich, Mark Strynar, Seth Newton (NTA Analysis)
  - Katherine Phillips, Kathie Dionisio, Kristin Isaacs (Consumer Products Database)
- Emma Schymanski – Luxembourg Center for Systems Biomedicine (MS-ready/NTA)



## **Antony Williams**

US EPA Office of Research and Development

National Center for Computational Toxicology (NCCT)

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

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