

# Applications of the US EPA's CompTox Dashboard to support structure identification and chemical forensics using mass spectrometry

*Antony Williams<sup>1</sup>, Andrew D. McEachran<sup>2</sup>,  
Jon R. Sobus<sup>3</sup> and Emma Schymanski<sup>4</sup>*

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

*2) Oak Ridge Institute of Science and Education (ORISE) Research Participant, RTP, NC*

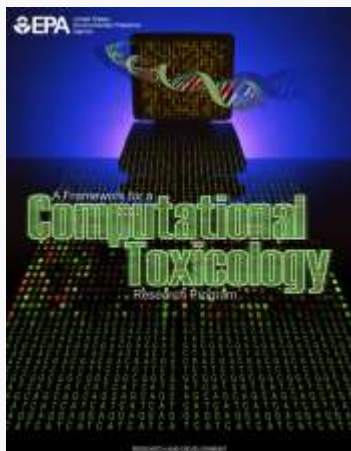
*3) National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*

*4) Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg*

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

*August 2018  
ACS Fall Meeting, Boston*

# National Center for Computational Toxicology




- National Center for Computational Toxicology established in 2005 to integrate:
  - High-throughput and high-content technologies
  - Modern molecular biology
  - Data mining and statistical modeling
  - Computational biology and chemistry
- Researching computational approaches to quickly evaluate the safety of chemicals for potential risk.
- Outputs: a lot of data, models, algorithms and software applications

- A publicly accessible website delivering access:
  - New entry portal for all NCCT dashboards
  - ~**762,000** chemicals with related property data
  - **Searchable by chemical, product use, gene and assay (ToxCast)**
  - Experimental and predicted physicochemical property data
  - **“Bioactivity data” for the ToxCast/Tox21 project**
  - **Generalized Read-Across (GenRA) module**
  - 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 Dashboard

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

 United States  
Environmental Protection  
Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share

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

762 Thousand Chemicals

Search for chemical by substance name, synonym, CAS number, TTXQID or INChIKey

Identifier substring search

See what people are saying, read the dashboard comments!  
[Cite the Dashboard Publication click here](#)

Latest News


[Read more news](#)

**YouTube video regarding using the Dashboard for Non-Targeted Analysis**

ch 7th, 2018 at 9:43:36 AM

Ar  
Mar


YouTube video discussing the application of the CompTox Chemistry Dashboard to support non-targeted analysis by mass spectrometry is available. This short video summarizes the advantages The dashboard in terms of data quality and focused data set for environmental non-targeted analysis. [View it here on Youtube](#).

 **Discover.**  
[About/Disclaimer](#)  
[Accessibility](#)  
[Privacy](#)

**Connect.**  
[ACToR](#)  
[DSSTox](#)  
[Downloads](#)


**Ask.**  
[Contact](#)  
[Help](#)

# CompTox Dashboard Chemicals










 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

762 Thousand Chemicals

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

Q Bisphenol A

-  Bisphenol A  
DTXSID0020182
-  Bisphenol A bis(2-hydroxyethyl ether) diacrylate  
DTXSID0066991
-  Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate  
DTXSID0066992
-  Bisphenol A bis(2-hydroxypropyl) ether  
DTXSID0051562
-  Bisphenol A carbonate polymer  
DTXSID0027640
-  Bisphenol A diglycidyl ether  
DTXSID0024824
-  Bisphenol A glycidyl methacrylate  
DTXSID0044847
-  Bisphenol A propoxylate diglycidyl ether  
DTXSID0038098
-  Bisphenol A propoxylate glycerolate diacrylate  
DTXSID00400728

comptox-prod.epa.gov/dashboard

# Detailed Chemical Pages

## DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

► ADME

► EXPOSURE

► BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

► LITERATURE

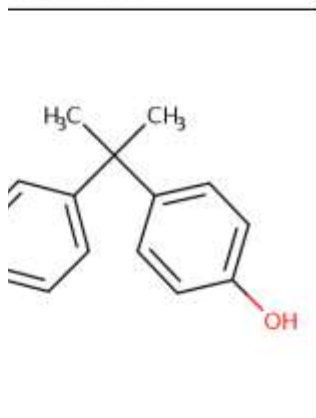
LINKS

COMMENTS

Bisphenol A

DTXSID7020182

SSTox Substance Id.



### Wikipedia

**Bisphenol A (BPA)** is an organic synthetic compound with the chemical formula  $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4)_2$  belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water. It has been in commercial use since 1967.

BPA is a starting material for the synthesis of plastics, primarily

[Read more](#)

### Intrinsic Properties

### Structural Identifiers

### Linked Substances

### Presence in Lists

### Record Information

### Quality Control Notes

# Access to Chemical Hazard Data

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search at EPA

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

Data Type  
Point of Departure  
Download

Human Eco

Columns 10

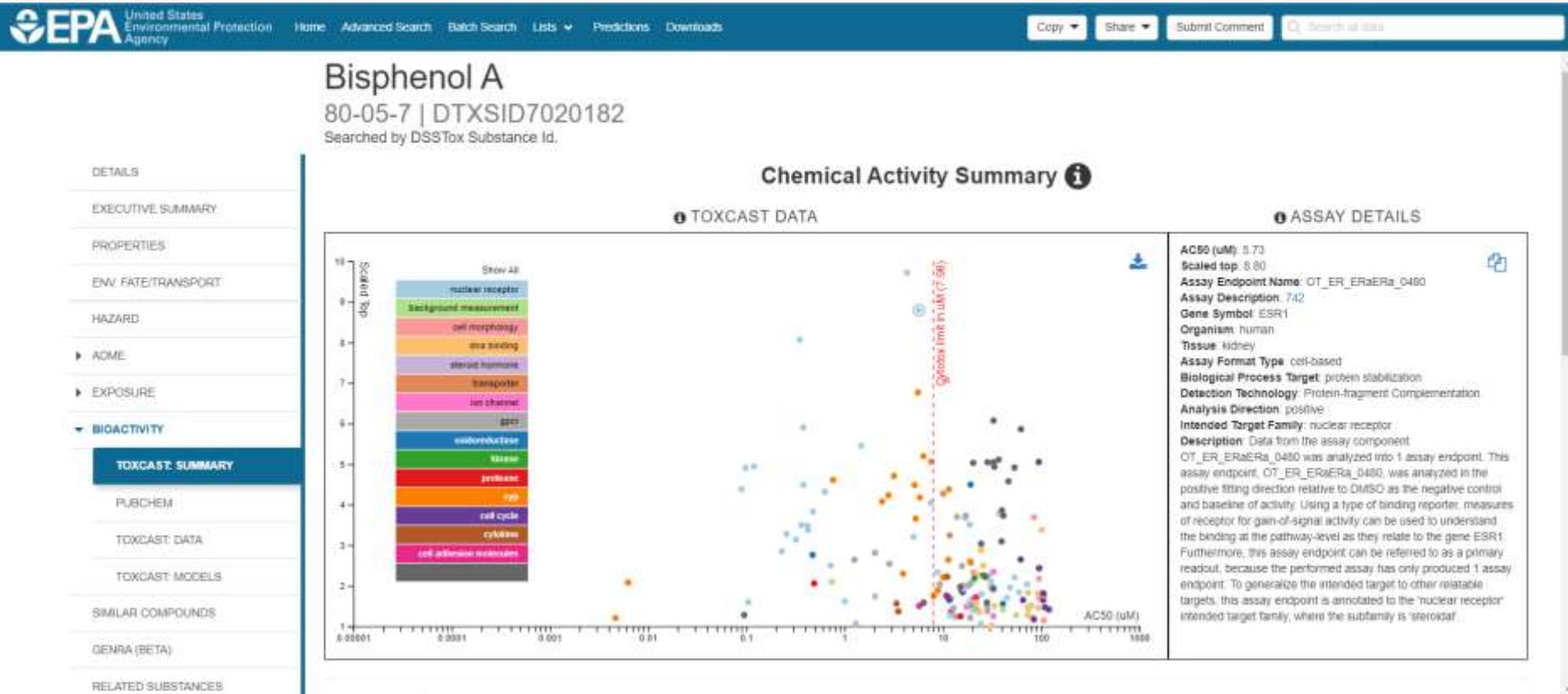
Search query

More	Priority	Toxval type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	5	BMDL-10	-	chronic	0.009	mg/kg-day	human	-	mouse	EFSA CEF	EFSA
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECOA
	6	NOAEL	-	reproductive	3.75	mg/kg-day	reproductive	oral	rat	-	HPVIS
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECOA
	5	NOEL	Systemic	repeat dose	4.0	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECOA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECOA
	7	LEL	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB
	7	nel	-	chronic	5	mg/kg-day	reproductive multigeneration	oral	rat	open_III	ToxRefDB
	5	NOAEL	-	chronic	5	mg/kg-day	human	-	mouse	EFSA AFC	EFSA
	7	nel	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB




# In Vitro Bioassay Screening

## ToxCast and Tox21





# Sources of Exposure to Chemicals

 **United States  
Environmental Protection  
Agency**

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Copy Share Submit Comment Search at EPA

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ACME

Download

Columns: 10

Search query

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

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

EXPOSURE

**PRODUCT & USE CATEGORIES**

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

TOXICS RELEASE INVENTORY

MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

# External Links to ~80 websites

## 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
- Chempidier
- 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
- ECHA Infocard
- ChemAgora



### Health Perspectives

### Program

### on Profiles

### de

### d

### Search Engine

### search

### 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

### Prediction

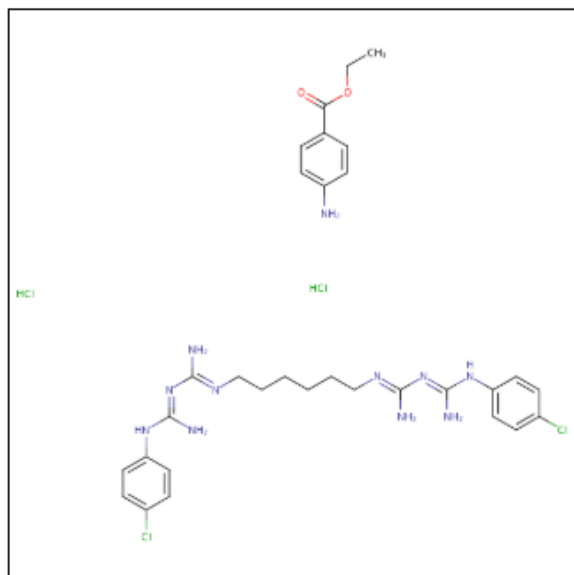
- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRTP Predictor
- LSERD

# MS-Ready Mappings

## Progaron

108532-15-6 | DTXSID20148579

Searched by DSSTox Substance Id.



Intrinsic Properties

Structural Identifiers

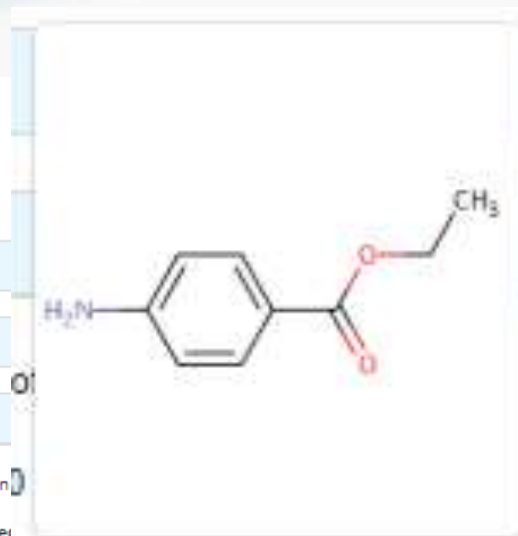
Linked Substances

Same Connectivity: 1 record (based on D

Mixtures, Components and Neutralizer

MS-Ready Mappings: DTXCID0013314

Similar Compounds: 0 records



DTXCID301804:11 records; D

Presence in Lists

Record Information

Quality Control Notes

# MS-Ready Mappings Set

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

9 of 11 chemicals visible

Download / Send

Sort by: DTXSID

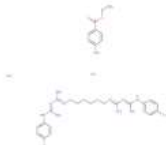


Show info: DTXSID CASRN Select all

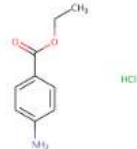
Filter by: Name or CASRN Isotopes



Anesthesine oxalate  
DTXSID: DTXSID20148337  
CASRN: 107948-47-0



Progaron  
DTXSID: DTXSID20148579  
CASRN: 108532-15-8



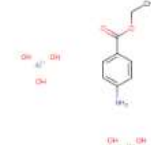
Benzocaine hydrochloride  
DTXSID: DTXSID0177812  
CASRN: 23239-88-5



Anesthesine succinate  
DTXSID: DTXSID60148336  
CASRN: 107948-46-9



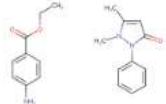
Almagel A-neo  
DTXSID: DTXSID60227559  
CASRN: 76741-92-9



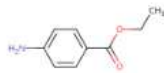
Almagel  
DTXSID: DTXSID70227560  
CASRN: 76741-95-2



Ethyl 4-aminobenzoate-2,4,6-trinitrophenyl  
DTXSID: DTXSID70767033  
CASRN: 5982-70-7



Antipyrine mixture with benzocaine  
DTXSID: DTXSID80212888  
CASRN: 63448-01-1



Benzocaine  
DTXSID: DTXSID8021804  
CASRN: 94-09-7

# Mass and Formula Searches

## Supporting Mass Spectrometry

### Advanced Search

#### Mass Search

Select Adduct:


Neutral ▼

 Da


±

Error Da


Da ppm

Search 

#### Molecular Formula Search

☒ MS Ready Formula 

☐ Exact Formula 

Search 


#### Generate Molecular Formula(e)

  Da

±

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]

# Advanced Searches

## Mass Based Search

### Mass Search

±

Min/Max

M

191.131


Da

±

5

Da


ppm

Search 

### Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

298 of 298 chemicals visible


Download / Send 

Show info:


DTXSID 

CASRN 

TOXCAST 


Mass Diff 

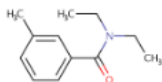
Select all 

Sort by: Mass Difference 



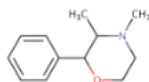
Filter by: Name or CASRN

Multicomponent Chemicals 



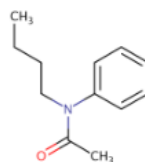
DEET

DTXSID: DTXSID2021995  
CASRN: 134-62-3  
TOXCAST: 14/663  
Mass Diff: 0.000014



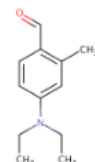
Phendimetrazine

DTXSID: DTXSID1023447  
CASRN: 634-03-7  
TOXCAST: 0  
Mass Diff: 0.000014



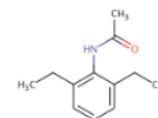
N-Butylacetanilide

DTXSID: DTXSID2042197  
CASRN: 91-49-6  
TOXCAST: 0  
Mass Diff: 0.000014



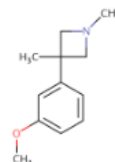
Benzaldehyde, 4-(diethylamino)-2-methyl-

DTXSID: DTXSID4059041  
CASRN: 92-14-8  
TOXCAST: 0  
Mass Diff: 0.000014



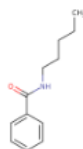
Acetanilide, 2',6'-diethyl-

DTXSID: DTXSID90168148  
CASRN: 16665-89-7  
TOXCAST: 0  
Mass Diff: 0.000014



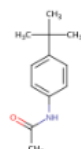
Azetidine, 1,3-dimethyl-3-(m-methoxyphenyl)-

DTXSID: DTXSID40173560  
CASRN: 19832-26-8  
TOXCAST: 0  
Mass Diff: 0.000014



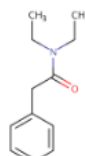
Benzamide, N-pentyl-

DTXSID: DTXSID20174198  
CASRN: 20308-43-4



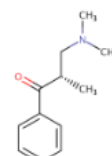
p-t-Butylacetanilide

DTXSID: DTXSID80174238  
CASRN: 20330-45-4



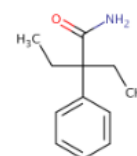
N,N-Diethylphenylacetamide

DTXSID: DTXSID00179048  
CASRN: 2431-98-1



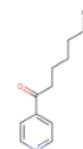
3-(Dimethylamino)-2-methylpropiphen-

DTXSID: DTXSID60180798  
CASRN: 26171-60-6



Butyramide, 2-ethyl-2-phenyl-

DTXSID: DTXSID60184653  
CASRN: 30568-39-9



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

DTXSID: DTXSID40188594  
CASRN: 32841-30-3



# Advanced Searches

## Mass Based Search

### Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

296 of 296 chemicals visible

Download / Send

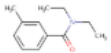
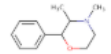
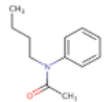
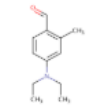
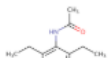
Select all

Sort by: Mass Difference



Filter by: Name or CASRN

Multicomponent Chemicals

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass	Mass Difference
	<a href="#">DTXSID2021995</a> ToxCast™	DEET	134-92-3	Level 1	111	111	155	753	191.131014	0.000014
	<a href="#">DTXSID1023447</a>	Phendimetrazine	634-03-7	Level 2	12	28	35	50	191.131014	0.000014
	<a href="#">DTXSID2042197</a>	N-Butylacetanilide	91-49-6	Level 2	1	26	50	1	191.131014	0.000014
	<a href="#">DTXSID4059041</a>	Benzaldehyde, 4-(diethylamino)-2-methyl-	92-14-8	Level 3	0	7	51	0	191.131014	0.000014
	<a href="#">DTXSID90168148</a>	Acetanilide, 2',6'-diethyl-	16665-89-7	Level 4	0	4	33	0	191.131014	0.000014

- Singleton searches are useful but we work with thousands of chemicals!
- 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?








## Batch Search?



### Step Three: Select Download Data or Display Chemicals

Please enter one identifier per line

#### Select Input Type(s)

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



Chemical Data


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

Fuel oil, no. 1  
Ethylene oxide  
Chloromethane  
1-Chloropropan-2-one  
n-Hexane  
Ammonia  
Nickel carbonyl  
Phosgene  
Potassium cyanide  
Chlorodimethylsilane

# Batch Searching

Select Output Format:






 Excel 

 Download






## Customize Results

- ☐ Select All
- ☐ Select All in Lists






## Chemical Identifiers

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

## Structures

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

## Intrinsic And Predicted Properties

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

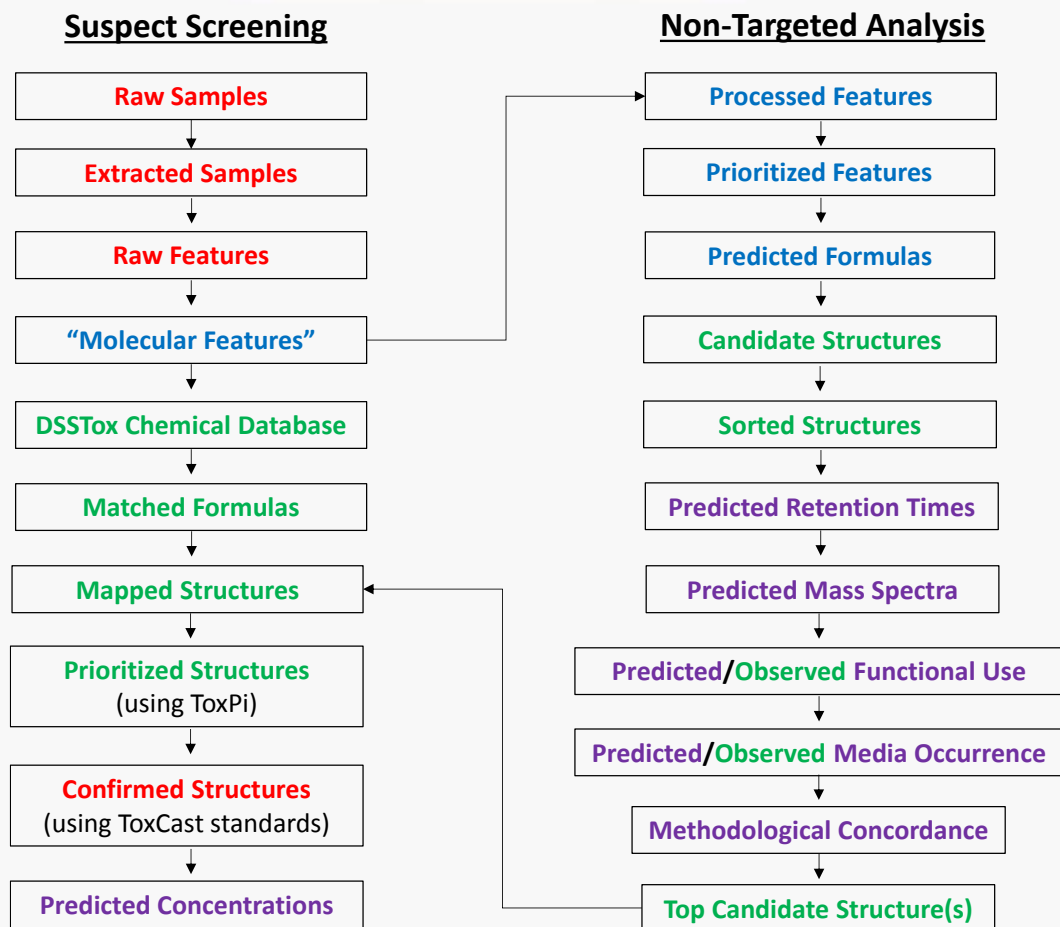
## Presence in Lists:

- ☐ ICCVAM test method evaluation report: in vitro ocular toxicity test methods
- ☐ 40CFR355
- ☐ A list of all PBDEs (Polybrominated diphenyl ethers)
- ☐ A list of all PCBs (Polychlorinated biphenyls)
- ☐ A list of polycyclic aromatic hydrocarbons
- ☐ Acute exposure guideline levels
- ☐ Algal Toxins
- ☐ Androgen Receptor Chemicals
- ☐ APCRA Chemicals for Prospective Analysis
- ☐ APCRA Chemicals for Retrospective Analysis
- ☐ APCRA Chemicals for Retrospective Analysis\_App\_List\_448\_Chemicals
- ☐ ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances
- ☐ ATSDR Toxic Substances Portal Chemical List
- ☐ Bisphenol Compounds
- ☐ California Office of Environmental Health Hazard Assessment
- ☐ Chemicals with interesting names
- ☐ CMAP
- ☐ DNT Screening Library
- ☐ Drinking Water Suspects, KWR Water, Netherlands
- ☐ EDSP Universe
- ☐ EPA Chemicals associated with hydraulic fracturing
- ☐ EPA Chemicals associated with hydraulic fracturing

# 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	-

# Suspect Screening and Non-Targeted Analysis Workflow



## Color Key

**Red** = Analytical Chemistry  
**Blue** = Data Processing & Analysis  
**Purple** = Mathematical & QSPR Modeling  
**Green** = Informatics & Web Services





# 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

Default Options: C[1-50] H[0-100] O[0-20] N[0-20]  
Include Halogens: ☐ F[0-20] ☐ Cl[0-20] ☐ Br[0-20]

Options ▾

Step One

Step Two

Step Three

Step Four

Step Five

Step Six

Step Five: Choose Data Fields to Download

Please enter one identifier per line

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)

C14H22N2O3

C10H12N2O

C14H18N4O3

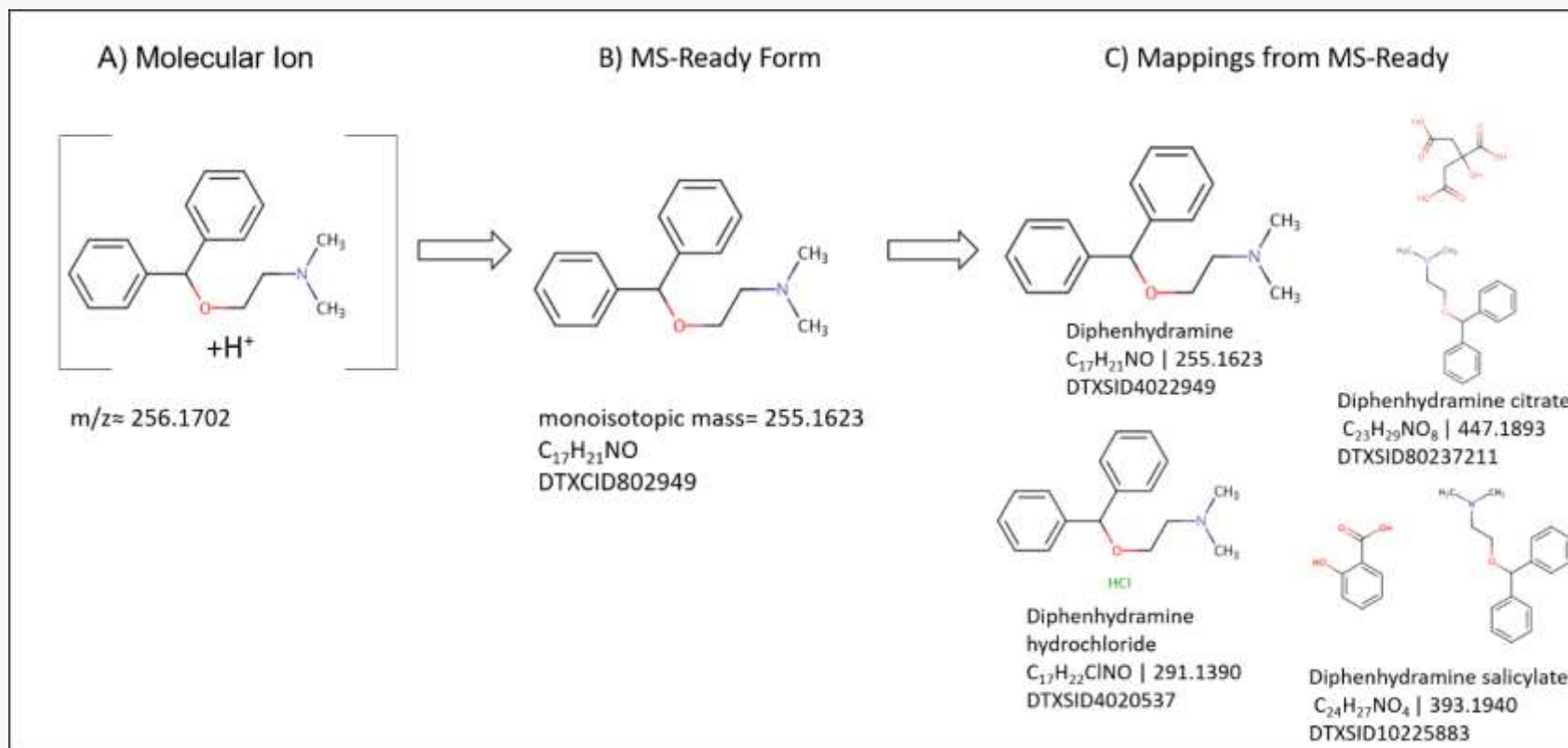
C12H11N7

C8H9NO2

Display All Chemicals

Download Chemical Data

# Specific Data-Mappings “MS-Ready Structures”




# MS-Ready Mappings


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

## Molecular Formula Search

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

☐ MS Ready Formula 

☒ Exact Formula 

Search 

3 of 3 chemicals visible

Download / Send 


Show info:

DTXSID 

CASRN 


Select all

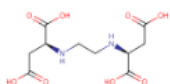


Sort by: DTXSID 

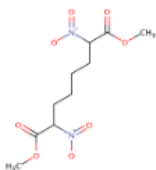


Filter by: Name or CASRN

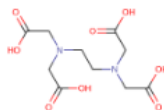
Multicomponent Chemicals 



N,N'-Ethylenedi-L-aspartic acid  
DTXSID: DTXSID1051852  
CASRN: 20846-91-7



Dimethyl 2,7-dinitrooctanedioate  
DTXSID: DTXSID20498804  
CASRN: 67404-09-5



Ethylenediaminetetraacetic acid  
DTXSID: DTXSID6022977  
CASRN: 60-00-4

# 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: 93 Chemicals**

93 chemicals

Download / Send

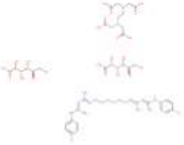


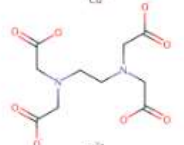
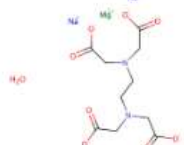
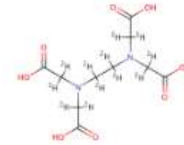
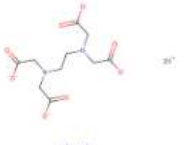
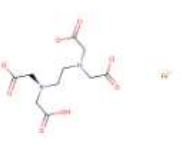
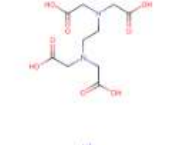

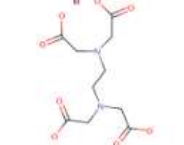
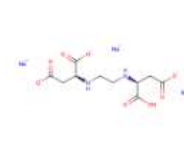
Sort by: DTXSID

Show info: DTXSID CASRN

Filter by: Name or CASRN

Select all

Hide

 <p>Trisdine DTXSID: DTXSID00153984 CASRN: 123354-94-9</p>	 <p>Acetic acid, (ethylenedinitrilo)tetra-, yttrium salt DTXSID: DTXSID00154799 CASRN: 12558-71-3</p>	 <p>Acetic acid, (ethylenedinitrilo)tetra-, aluminum salt DTXSID: DTXSID00183708 CASRN: 29507-62-8</p>	 <p>EDTA, copper salt DTXSID: DTXSID0034564 CASRN: 12278-01-8</p>	 <p>Magnesium sodium 2,2',2'',2'''-(ethane-1,2-diylbis(nitrilo))tetraacetate DTXSID: DTXSID00583348 CASRN: 29932-54-5</p>	 <p>2,2',2'',2'''-[(1,2-ethylenediylbis(nitrilo))tetraacetate DTXSID: DTXSID00583949 CASRN: 203806-08-0</p>
 <p>Zincate(2-), [[N,N'-1,2-ethanediy]bis(N-(carboxymethyl)carbamoyl)] DTXSID: DTXSID0065696 CASRN: 14025-21-9</p>	 <p>Ferrate(1-), [[N,N'-1,2-ethanediy]bis(N-(carboxymethyl)carbamoyl)] DTXSID: DTXSID0066163 CASRN: 17099-81-9</p>	 <p>PUBCHEM_64611985 DTXSID: DTXSID00715445 CASRN: 22239-30-1</p>	 <p>Glycine, N,N'-1,2-ethanediybis(N-(carboxymethyl)carbamoyl) DTXSID: DTXSID10236595 CASRN: 87731-78-0</p>	 <p>Bismuth Sodium Ethylenediaminetetraacetate DTXSID: DTXSID10437000 CASRN: 12558-49-5</p>	 <p>N,N'-Ethylenedi-(L-aspartic acid) trisodium salt DTXSID: DTXSID1051806 CASRN: 178949-82-1</p>

- 93 chemicals returned in total
  - Only 7 of the 93 are single component chemicals
  - Only 4 of the 93 are non-isotope-labeled
  - 3 are neutral compounds and 1 is charged

# Complexity to Simplicity

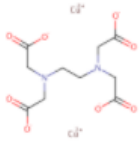

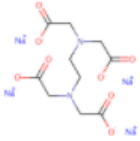

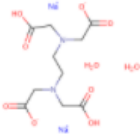

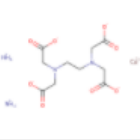

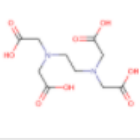

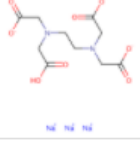

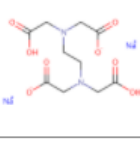

## 93 Chemicals – 7 in EPAHFR

	INPUT	DTXCID	INDIVID	FORMULA	SMILES	DTXSID	CASRN	EXPOCAST	MEXPOCAST	DATA_SOURCE	TOXVAL	DTXCAST	TOXCAST	# OF PUBMED	PUBCHEM	EPAHFR	
2	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID6022977	60-00-4	7.96e-05	Y			71	Y	2.65	3/113	25251	158	Y
3	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID9027073	139-33-3	-	-			41	Y	-	-	25251	56	Y
4	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID3026350	64-02-8	-	-			37	Y	-	-	-	57	Y
5	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID7020556	150-38-9	-	-			30	Y	-	-	-	33	Y
6	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID5049609	67989-88-2	-	-			20	Y	-	-	-	8	Y
7	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID5049576	6381-92-6	-	-			19	Y	-	-	25251	31	Y
8	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID0034564	12276-01-6	-	-			11	-	-	-	-	8	Y
9	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID5027774	15708-41-5	-	-			48	Y	1.98	6/303	241	53	-
10	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID2036409	62-33-9	4.64e-06	Y			37	Y	0.0	0/64	25251	42	-
11	C10H16N2O8	DTXCID00197424		C10H16N2(OC(=O)Cl	DTXSID1051852	20846-91-7	-	-			36	Y	-	-	89	25	-
12	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID6042107	15375-84-5	-	-			25	Y	-	-	97	25	-
13	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID3036442	5964-35-2	-	-			23	Y	-	-	25251	25	-
14	C10H16N2O8	DTXCID00197424		C10H16N2(OC(=O)Cl	DTXSID1051806	178949-82-1	-	-			22	Y	-	-	-	5	-
15	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID0065696	14025-21-9	-	-			22	Y	-	-	-	43	-
16	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID9027813	20824-56-0	-	-			21	Y	-	-	-	12	-
17	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID9027815	21265-50-9	-	-			20	Y	-	-	241	24	-
18	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID5058272	17421-79-3	-	-			19	Y	-	-	25251	25	-
19	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID3058612	2001-94-7	-	-			18	Y	-	-	25251	19	-
20	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID8027820	22473-78-5	-	-			16	Y	-	-	-	11	-
21	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID8058324	17572-97-3	-	-			15	-	-	-	-	36	-
22	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID8028343	67859-51-2	-	-			14	Y	-	-	-	5	-
23	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID4051328	13235-36-4	-	-			14	-	-	-	-	18	-
24	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID6070980	68015-77-0	-	-			14	Y	-	-	-	13	-
25	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID9058317	15934-01-7	-	-			11	-	-	-	-	5	-
26	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID0066163	17099-81-9	-	-			11	-	-	-	241	14	-
27	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID1068988	54959-35-2	-	-			11	-	-	-	241	14	-
28	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID5074266	60816-63-9	-	-			11	-	-	-	1	10	-
29	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID4048197	39208-15-6	-	-			10	-	-	-	-	28	-
30	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID2065830	14931-83-0	-	-			10	-	-	-	47	9	-
31	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID70189997	36499-65-7	-	-			10	-	-	-	25298	26	-
32	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID7051420	61916-40-3	-	-			9	-	-	-	-	4	-
33	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID2051425	73513-47-0	-	-			8	Y	-	-	-	3	-
34	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID7051426	73637-19-1	-	-			8	Y	-	-	-	5	-
35	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID2051427	73637-20-4	-	-			8	Y	-	-	-	-	-
36	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID3058741	10378-23-1	-	-			8	Y	-	-	-	31	-
37	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID6065925	15708-48-2	-	-			8	-	-	-	-	19	-
38	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID20217976	6766-87-6	-	-			8	-	-	-	-	13	-
39	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID5065807	14689-29-3	-	-			7	-	-	-	-	12	-
40	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID6069408	60544-70-9	-	-			7	-	-	-	-	12	-
41	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID00153984	123354-94-9	-	-			7	-	-	-	2	6	-
42	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID70190705	37209-61-3	-	-			7	-	-	-	6	9	-
43	C10H16N2O8	DTXCID902977		C10H16N2(OC(=O)Cl	DTXSID7051424	67401-50-7	-	-			6	-	-	-	-	4	-



# Complexity to Simplicity

## 93 Chemicals – 7 in the list

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source	Monoisotopic Mass
	<a href="#">DTXSID0034564</a>	EDTA, copper salt	12276-01-6	Level 1	10	11	8	0	413.918561 
	<a href="#">DTXSID3026350</a>	Ethylenediaminetetraacetic acid tetrasodium salt	64-02-8	Level 1	1227	37	57	0	380.018442 
	<a href="#">DTXSID5049576</a>	Disodium ethylenediaminetetraacetate dihydrate	6381-92-6	Level 1	93	19	31	25251	372.075683 
	<a href="#">DTXSID5049609</a>	Ethylenediaminetetraacetic acid, diammonium copper salt	67989-88-2	Level 2	9	20	8	0	387.057712 
	<a href="#">DTXSID6022977</a>	Ethylenediaminetetraacetic acid	60-00-4	Level 1	346	71	158	25251	292.090665 
	<a href="#">DTXSID7020556</a>	Trisodium ethylenediaminetetraacetate	150-38-9	Level 1	85	30	33	0	358.036498 
	<a href="#">DTXSID9027073</a>	Ethylenediaminetetraacetic acid, disodium salt	139-33-3	Level 1	1358	41	56	25251	336.054554 

# Searching batches

## Formula (or mass) searching

	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3	338.116398	19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23ClN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S	406.213757997	35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35ClN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35ClN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O	176.094963014	40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13ClN2O	212.0716407	11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxy pyrimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026	13140-86-8	N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3	290.137890456	68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3	290.137890456	51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19ClN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2-	C14H18N4O3	290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quinaz	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxyphenyl	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid-6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6025567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	50

# Downloadable Data

## [DSSTox MS Ready Mapping File](#)

Posted: 11/14/2016

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

## [DSSTox SDF File](#)

Posted: 12/14/2016

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

## [PHYSPROP Analysis File](#)

Posted: 12/14/2016

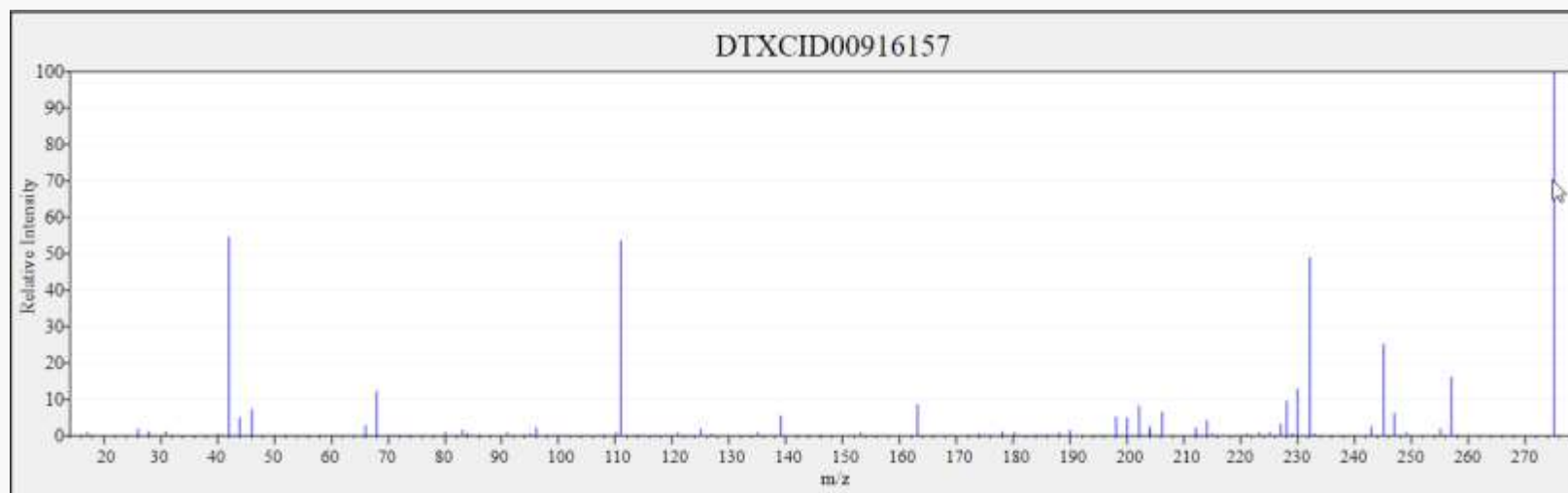
- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search
- Generation of MS-ready structures:
  - Upload file, download results
  - Service based generation

# Predicted Mass Spectra

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



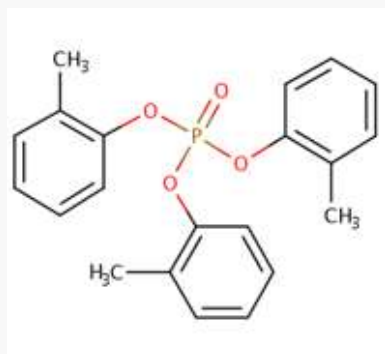
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >700,000 structures, to be accessible via Dashboard



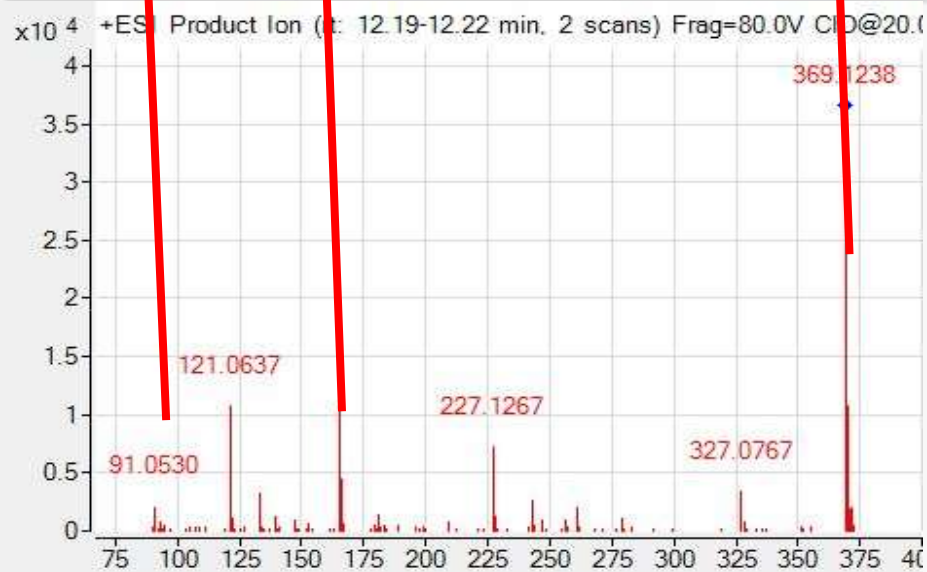
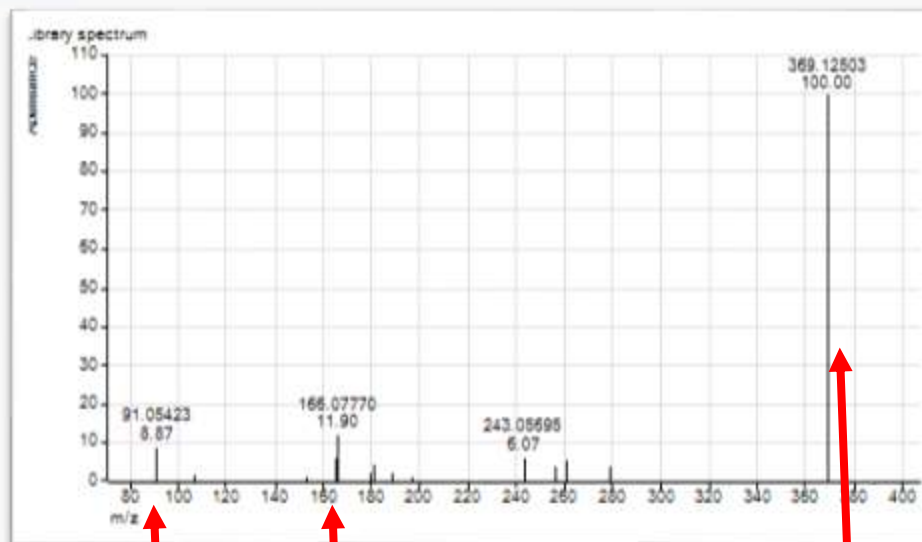


# Predicted Mass Spectra

Library Fragmentation  
Spectra (20eV)



Observed Fragmentation  
Spectra (20eV)



Match  
Score



# Search Expt. vs. Predicted Spectra

## Mass Search

±

Min/Max

Mass

Da

±

Error

Da

ppm

## Molecular Formula Search

Molecular Formula

*Mass or Formula must be entered before searching spectrum*

## Ionization Type

ESI+ ▼

## Spectra Input

Single Energy

Multiple

Peak Match Window:

0.02

Da

ppm

Search

# Prototype Development

## AADashboard

atrazine

100%

Select properties to predict

H

C ☐ Exact

N ☒ Substructure

O

Search result **2540** Show ☐ Isotopically Labeled ☐ Charged ☐ Salts or Mixtures Sort Similarity

Search result **2540** Show ☐ Isotopically Labeled

# Prototype Development

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):

Search

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

Elements per page 50

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

- The CompTox Chemistry Dashboard provides access to data for ~760,000 chemicals
- Multiple prediction models available for data gap filling
  - OPERA models and TEST models – PhysChem and Tox endpoints
  - Models based on *in vitro* data – classification models
  - Generalized Read-Across development in progress
- Real time prediction models rollout has started
- Web services available for some physchem and toxicity endpoints
- 2 years development as a CompTox Integration Hub

## **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>