

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

*Antony Williams<sup>1</sup> and Andrew D. McEachran<sup>2,3</sup>*

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

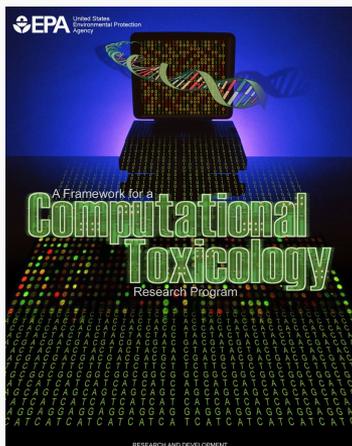
2) *Oak Ridge Institute of Science and Education (ORISE) Research Participant, RTP, NC*

3) *Present Address: Agilent Inc., Santa Clara, CA*

*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 2019  
Pittcon, Philadelphia*

# 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:
  - ~**875,000** 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 Chemicals Dashboard

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



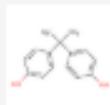
The screenshot shows the EPA CompTox Chemicals Dashboard. At the top is a blue navigation bar with the EPA logo and text: "United States Environmental Protection Agency", "Home", "Advanced Search", "Batch Search", "Lists", "Predictions", "Downloads", and a "Share" button. Below the navigation bar is a white header area with the EPA logo on the left and the text "875 Thousand Chemicals" on the right. Under the logo is a circular menu with three tabs: "Chemicals" (selected), "Product/Use Categories", and "Assay/Gene". Below the menu is a search bar with the placeholder text "Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey". To the left of the search bar is a checkbox labeled "Identifier substring search". To the right of the search bar are two links: "See what people are saying, read the dashboard [comments!](#)" and "Cite the Dashboard Publication [click here](#)". Below this is a section titled "Latest News" with a link "Read more news". The main content area features a news item titled "Journal of Cheminformatics article regarding 'MS-Ready structures'" dated "March 9th, 2019 at 1:09:45 PM". The text of the article reads: "A recent article describes 'MS-Ready structures', what they are, how they are generated and details regarding the benefits of these structures in navigating structure relationships across the dashboard. The article is published in the Journal of Cheminformatics [here](#)." At the bottom of the news item are four dots, with the first one filled, indicating the current position in a carousel.

Chemicals

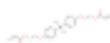
Product/Use Categories

Assay/Gene

 Bisphenol



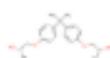
Bisphenol A  
DTXSID7020182



Bisphenol A bis(2-hydroxyethyl ether) diacrylate  
DTXSID6066991

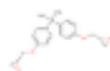


Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate  
DTXSID1066992

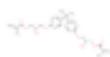


Bisphenol A bis(2-hydroxypropyl) ether  
DTXSID8051592

Bisphenol A carbonate polymer  
DTXSID6027840



Bisphenol A diglycidyl ether  
DTXSID6024624



Bisphenol A glycidyl methacrylate  
DTXSID7044841

## DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

COMMENTS

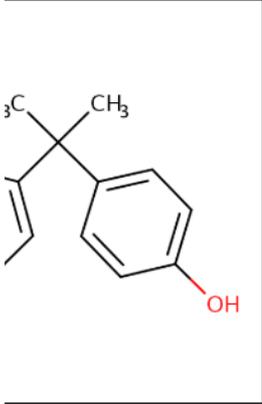
Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

## Bisphenol A

105-7 | DTXSID7020182

Identified by DSSTox Substance Id.



**Wikipedia**

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

[Read more](#)

**Intrinsic Properties**

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

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

**Monoisotopic Mass:** 228.11503 g/mol

**Structural Identifiers**

**Linked Substances**

**Presence in Lists**

**Record Information**

**Quality Control Notes**

# Access to Chemical Hazard Data

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

**HAZARD**

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

COMMENTS

## Hazard

Data Type: Toxicity Value

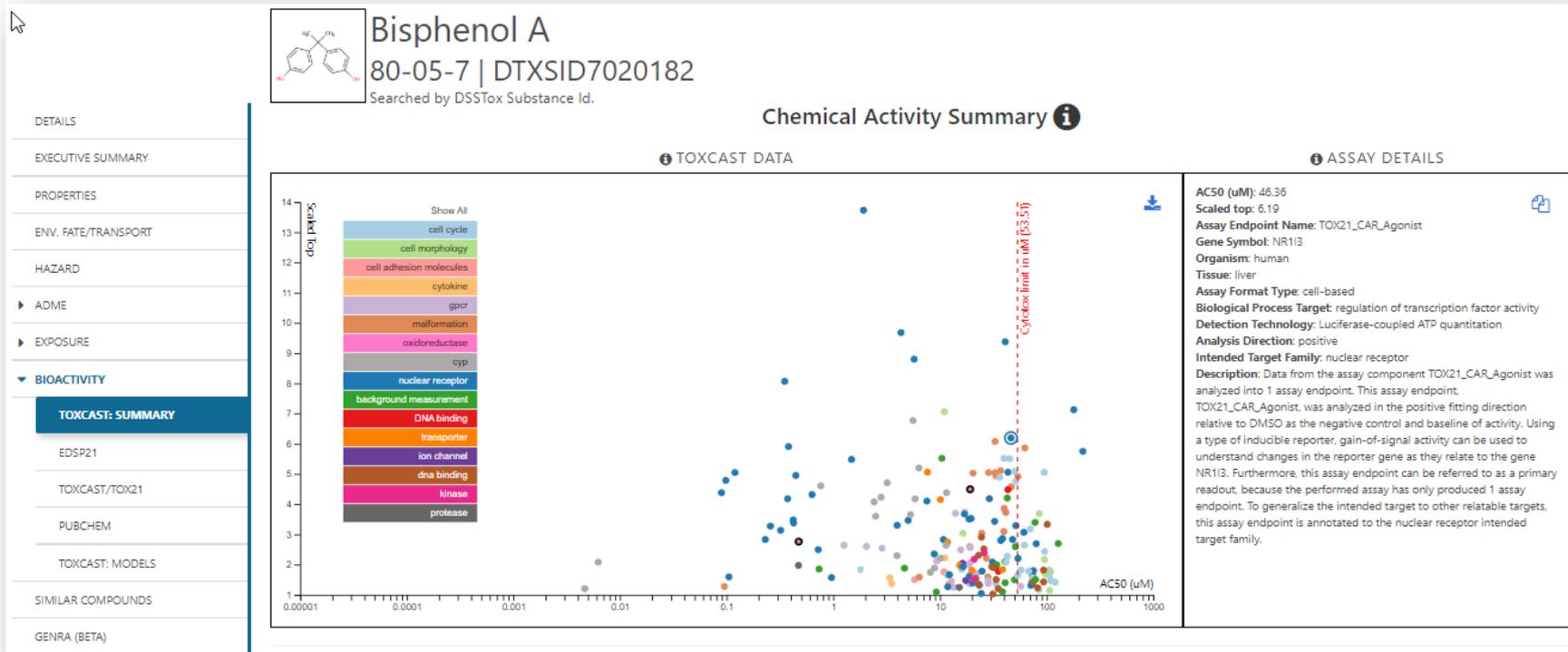
Human
Eco

Download
Columns
Search query

More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	7	MEG	Short-term Critical Air	short-term	500	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Short-term Marginal Air	short-term	100	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Short-term Negligible Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Soil Negligible Soil	chronic	106000	mg/kg	-	soil	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Long-Term, SL/d Negligible Water	chronic	7	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	6	RfD	-	chronic	0.05	mg/kg-day	-	oral	rat	Wignall	<a href="#">Wignall</a>
	5	RfD	-	chronic	0.05	mg/kg-day	-	-	-	MSC Table 5	<a href="#">Pennsylvania DEP ToxValues</a>
	4	RfD	-	chronic	0.05	mg/kg-day	chronic	oral	rat	IRIS	<a href="#">Chiu</a>
	3	RfD	-	chronic	0.6	mg/kg-day	-	oral	rat	EPA/ORNL/OLEM	<a href="#">HEAST</a>
	1	RfD	-	chronic	0.05	mg/kg-day	-	oral	-	EPA NCEA	<a href="#">IRIS</a>

# In Vitro Bioassay Screening

## ToxCast and Tox21



# Sources of Exposure to Chemicals

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

### EXPOSURE

#### PRODUCT & USE CATEGORIES

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

TOXICS RELEASE INVENTORY

MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

Download

Columns 10

Search query

### Product and Use Categories (PUCs) i

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

# MS-Ready Mappings

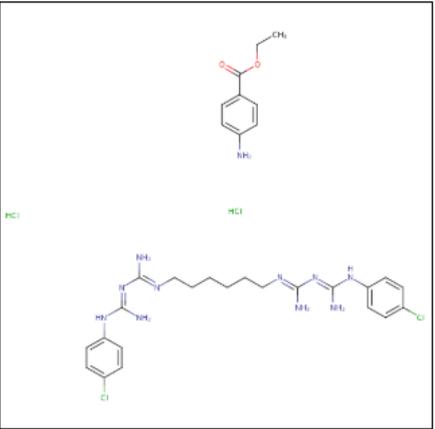
 United States Environmental Protection Agency

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)

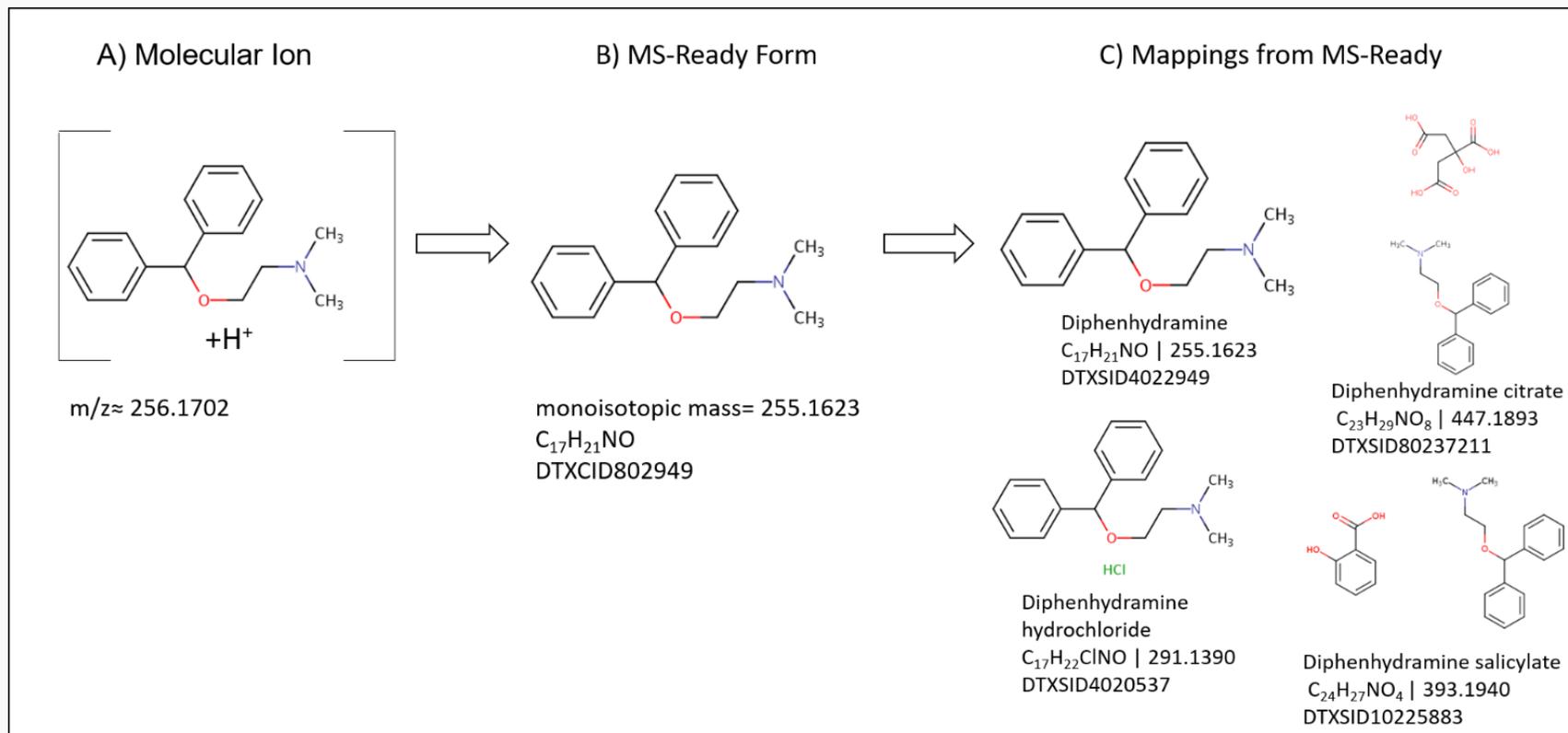
Mixtures, Component: 0 records

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

Similar Compounds: 0 records

Quality Control Notes

# Specific Data-Mappings “MS-Ready Structures”



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

Journal of Cheminformatics

METHODOLOGY

Open Access

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



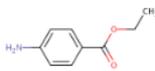
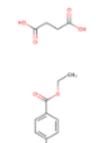
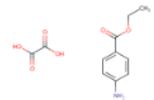
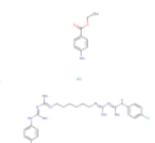
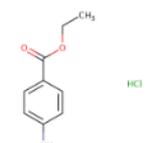
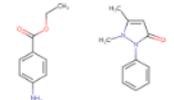
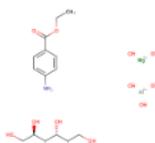
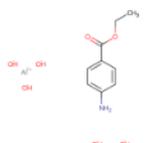
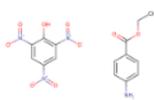
Andrew D. McEachran<sup>1,2\*</sup>, Kamel Mansouri<sup>1,2,3</sup>, Chris Grulke<sup>2</sup>, Emma L. Schymanski<sup>4</sup>, Christoph Ruttkies<sup>5</sup> and Antony J. Williams<sup>2\*</sup>

# MS-Ready Mappings Set

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

10 of 12 chemicals visible

Select all | Download | Send to Batch Search | Default | DTXSID | PubChem | CPDAT | Isotopes | Filter by Name or CASRN

 <p><b>Benzocaine</b> DTXSID: DTXSID8021804 PubChem: 184 CPDAT: 42</p>	 <p><b>Anesthesine succinate</b> DTXSID: DTXSID60148336 PubChem: 10 CPDAT: 0</p>	 <p><b>Anesthesine oxalate</b> DTXSID: DTXSID20148337 PubChem: 6 CPDAT: 0</p>	 <p><b>Progaron</b> DTXSID: DTXSID20148579 PubChem: 5 CPDAT: 0</p>	 <p><b>Benzocaine hydrochloride</b> DTXSID: DTXSID50177812 PubChem: 33 CPDAT: 0</p>	 <p><b>Antipyrine mixture with benzocaine</b> DTXSID: DTXSID80212866 PubChem: 9 CPDAT: 0</p>
 <p><b>Amagel A-neo</b> DTXSID: DTXSID60227559 PubChem: 9 CPDAT: 0</p>	 <p><b>Amagel</b> DTXSID: DTXSID70227560 PubChem: 9 CPDAT: 0</p>	 <p><b>Ethyl 4-aminobenzoate--2,4,6-trinitroph...</b> DTXSID: DTXSID70787033 PubChem: 5 CPDAT: 0</p>	 <p><b>1-Hexadecylpyridin-1-ium 2-butoxy-N-[...]</b> DTXSID: DTXSID50997335 PubChem: 0 CPDAT: 0</p>		

# Mass and Formula Searches Supporting Mass Spectrometry

## Advanced Search

### Mass Search

  Min/Max

Select Adduct: Neutral 

Mass Da

±

Error Da

Da ppm

Search Q

### Molecular Formula Search

Molecular Formula

MS Ready Formula 

Exact Formula 

Search Q

### Generate Molecular Formula(e)

  Min/Max

Mass Da

±

Error

Da ppm

Search Q

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

$\pm$  Min/Max

Adduct

Neutral 

All Adducts

Choose adduct from dropdown

191.131 Da  $\pm$  5 Da ppm

Search 

# Advanced Searches

## Mass Based 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

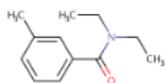
CASRN

TOXCAST

Mass Diff

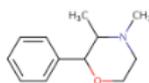
Multicomponent Chemicals

Filter by Name or CASRN



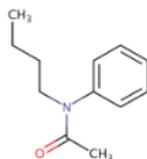
DEET

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



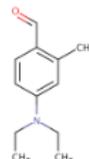
Phendimetrazine

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



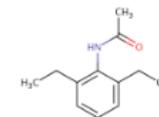
N-Butylacetanilide

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



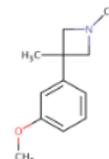
Benzaldehyde, 4-(diethylamino)-2-meth...

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



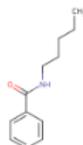
Acetanilide, 2',6'-diethyl-

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



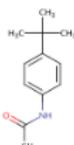
Azetidine, 1,3-dimethyl-3-(m-methoxyph...

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



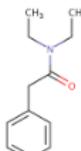
Benzamide, N-pentyl-

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



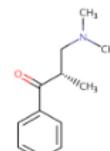
p-t-Butylacetanilide

DTXSID: DTXSID80174238  
CASRN: 20330-45-4  
TOXCAST: -  
Mass Diff: 0.000014



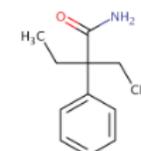
N,N-Diethylphenylacetamide

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



3-(Dimethylamino)-2-methylpropionophen...

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

# Advanced Searches

## Mass Based Search

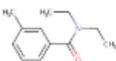
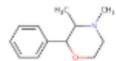
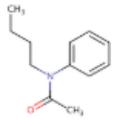
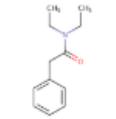
### Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

329 of 329 chemicals visible

Select all Download Send to Batch Search Sources

Multicomponent Chemicals Filter by Name or CASRN

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	134-62-3	Level 1	111	135	155	753	191.131014	0.000014
	<a href="#">DTXSID1023447</a>	Phendimetrazine	634-03-7	Level 2	12	34	35	50	191.131014	0.000014
	<a href="#">DTXSID2042197</a>	N-Butylacetanilide	91-49-6	Level 2	1	32	50	1	191.131014	0.000014
	<a href="#">DTXSID00179048</a>	N,N-Diethylphenylacetamide	2431-96-1	Level 4	0	21	52	34	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 Searches

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

## Batch Search

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

Step Six: Click "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

Display All Chemicals Download Chemical Data

**Select Output Format:** Excel

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

107-02-8  
79-06-1  
107-13-1  
309-00-2  
107-18-6  
62-53-3  
7631-89-2  
1327-53-3  
7784-46-5  
26628-22-8

**Download**

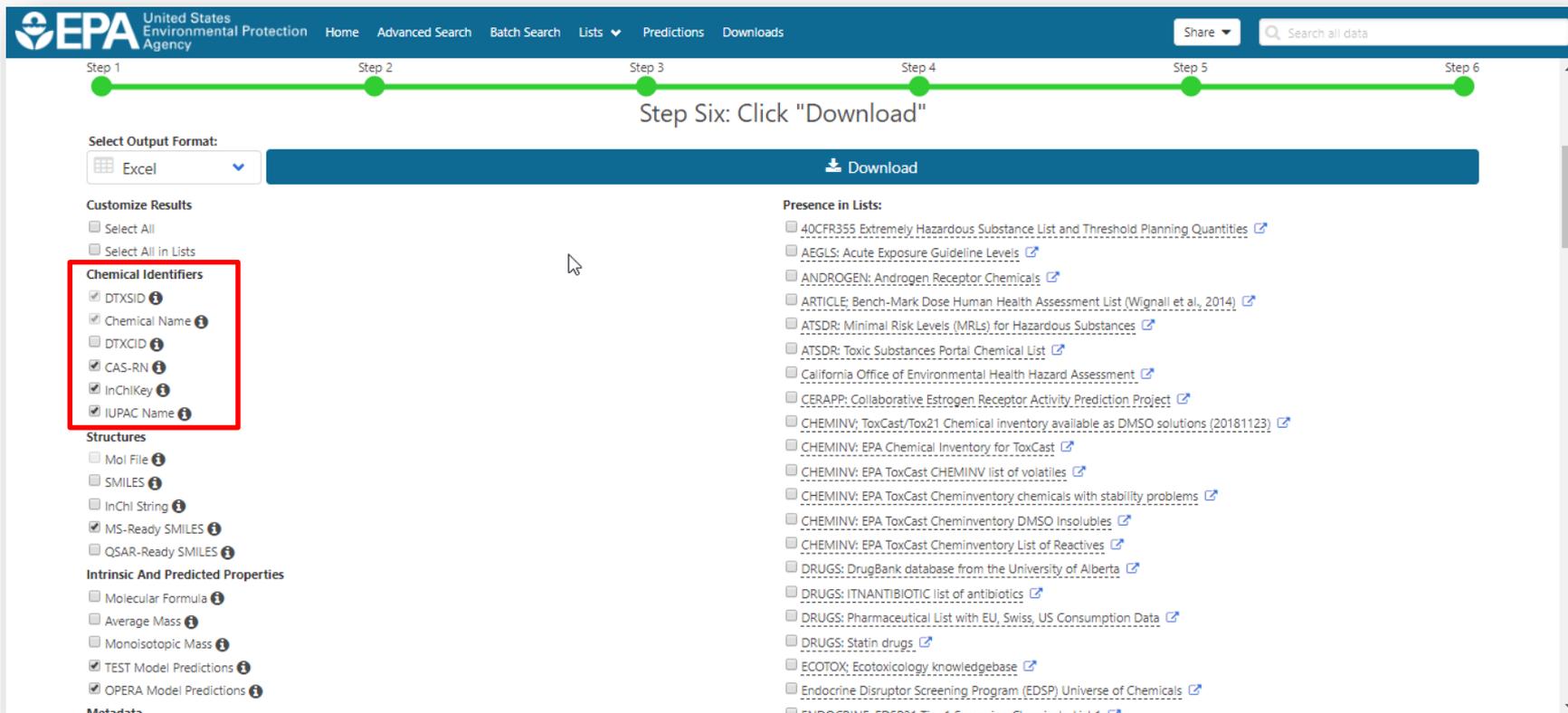
**Customize Results**

- Select All
- Select All in Lists

**Presence in Lists:**

- 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities
- AEGLS: Acute Exposure Guideline Levels

# Batch Searches



The screenshot shows the EPA Batch Search interface at Step Six. At the top, a progress bar indicates steps 1 through 6, with Step Six highlighted. Below the progress bar, the text "Step Six: Click 'Download'" is displayed. The interface includes a navigation menu with options like Home, Advanced Search, Batch Search, Lists, Predictions, and Downloads. A search bar is present with the text "Search all data".

Under the "Select Output Format:" section, "Excel" is selected in a dropdown menu, and a large blue "Download" button is visible. To the left, the "Customize Results" section is expanded, showing a list of checkboxes for various identifiers and structures. The "Chemical Identifiers" section is highlighted with a red box and includes the following checked items:

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

Other sections include "Structures" (with "MS-Ready SMILES" checked) and "Intrinsic And Predicted Properties" (with "TEST Model Predictions" and "OPERA Model Predictions" checked). On the right side, the "Presence in Lists:" section contains a list of various chemical lists and databases, such as "40CFR355 Extremely Hazardous Substance List" and "EPA Chemical Inventory for ToxCast".

# Batch Searching Formula/Mass

## Batch Search



### Step Five: Choose Data Fields to Download

Please enter one identifier per line 

+/-  ppm

#### 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 

 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
```

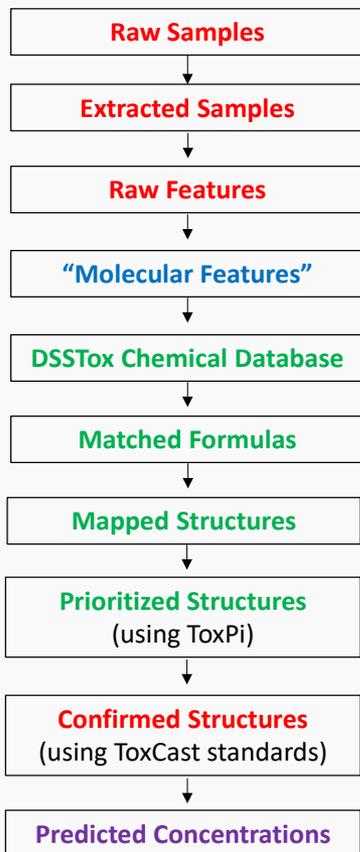
This search is based on what we refer to as "Mass Spec (MS) Ready" structures. All chemicals within the database are treated in a manner such that all are desalted, mixtures are separated, and stereochemistry is removed as Mass Spectrometry detects the major components of a salt or mixture and is insensitive 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>.

# Excel Output

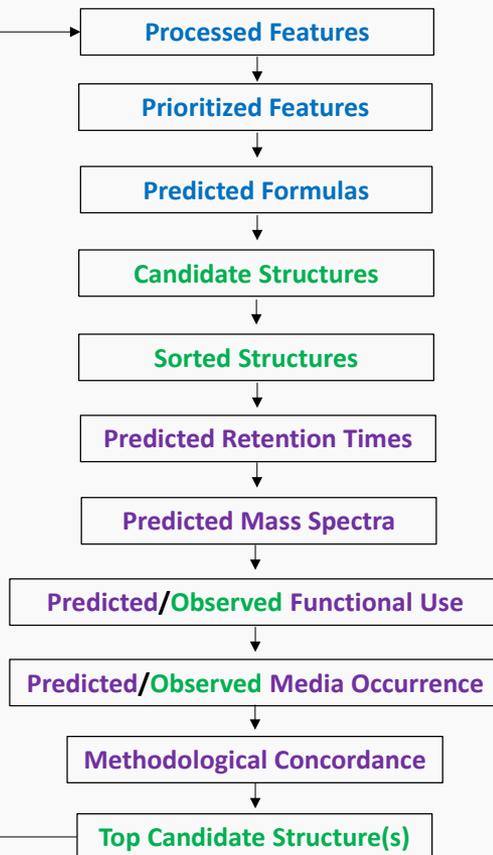
PREFERRE	CASRN	DATA_SOURC	TOXVAL_D	TOXCAST	TOXCAST	NUMBER_OF_PUBMED	PUBCHEM_DATA_S	CPDAT_COUNT
Aniline	62-53-3	148	Y	0.47	1/211	1932	284	80
3-Methylpyr	108-99-6	93	Y	2.8	16/571	16	120	16
2-Methylpyr	109-06-8	88	Y	0.0	0/211	1395	112	12
4-Methylpyr	108-89-4	82	Y	0.52	2/381	1	118	4
Aniline hydr	142-04-1	54	Y	2.23	9/404	1932	66	5
Anilinium ni	542-15-4	13	-	2.37	5/211	1932	4	-
Benzenami	542-16-5	13	-	-	-	-	-	5
1,3,5-Trinitr	3101-79-9	10	-	-	-	-	10	1
Di-2-propyn	6921-28-4	10	-	-	-	-	43	-
Benzenami	542-14-3	8	-	-	-	-	15	-
Butanal--an	68411-20-1	8	-	-	-	-	-	6
Benzenami	37832-42-1	7	-	-	-	1932	27	-
2,4-Hexadie	1516-01-4	6	-	-	-	-	25	-
Cyclobutan	15760-35-7	6	-	-	-	-	79	2
Aniline hydr	542-13-2	6	-	-	-	1932	6	-
Benzenami	542-11-0	6	-	-	-	-	-	-
Tris(4-amin	68389-46-8	6	-	-	-	-	-	3
4-[(2-Chloro	71566-74-0	6	-	-	-	-	-	4
Heptanal--a	9003-50-3	6	-	-	-	-	-	1

# Suspect Screening and Non-Targeted Analysis Workflow

## Suspect Screening



## Non-Targeted Analysis



## Color Key

**Red** = Analytical Chemistry

**Blue** = Data Processing & Analysis

**Purple** = Mathematical & QSPR Modeling

**Green** = Informatics & Web Services



# MS-Ready Structures Underpin Analysis

Mass

Search 

Molecular Formula Search

MS Ready Formula 

Formula

Please use the format of C1=CC=C(O)C=C1

Search 

Generate Molecular Formula

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 formula 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>

C1=CC=C(O)C=C1

# MS-Ready Structures Underpin Analysis

## 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 

 Display All Chemicals

+/-  ppm

Enter Identifiers to Search [\(search\)](#)

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>.

# 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] S[0-20] P[0-20] Cl[0-20] Br[0-20] I[0-20] F[0-20]

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

Options ▾

**Select Input Type(s)**

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

**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

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

C14H22N2O3  
C10H12N2O  
C14H18N4O3  
C12H11N7  
C8H9NO2

Display All Chemicals Download Chemical Data

# MS-Ready Mappings

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

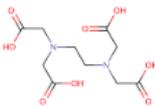
MS Ready Formula 
 Exact Formula 

Formula

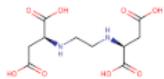


3 of 3 chemicals visible

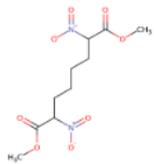
Select all Download Send to Batch Search Default     Filter by Name or CASR 



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

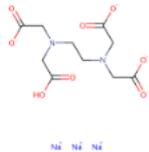
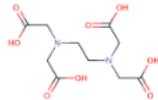
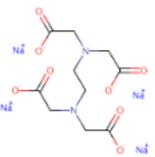
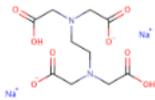
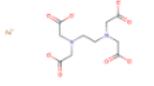
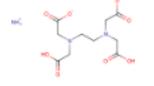
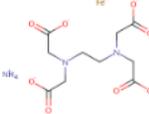
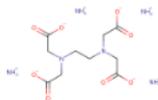
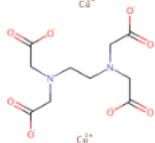
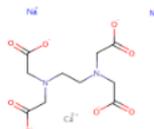
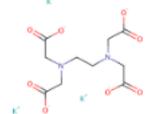
# MS-Ready Mappings

- **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 CASRN

 <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

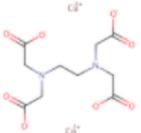
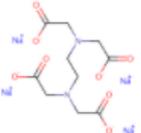
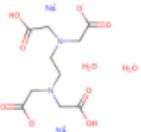
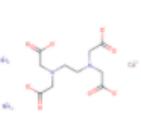
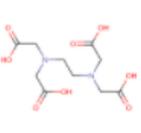
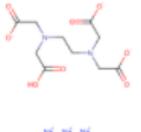
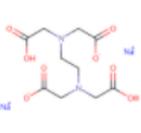
# Complexity to Simplicity

## 93 Chemicals – 7 in EPAHFR

INPUT	DTXCID_INDIVID	FORMULA	SMILES	DTXSID	CASRN	EXPOCAST	MEXPOCAST	DATA_SOURCE	TOXVAL	TOXCAST	TOXCAST	# OF PUBMED	PUBCHEM	EPAHFR	
2	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID6022977	60-00-4	7.96e-05	Y		71	Y	2.65	3/113	25251	158	Y
3	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID9027073	139-33-3	-	-		41	Y	-	-	25251	56	Y
4	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID3026350	64-02-8	-	-		37	Y	-	-	-	57	Y
5	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID7020556	150-38-9	-	-		30	Y	-	-	-	33	Y
6	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID5049609	67989-88-2	-	-		20	Y	-	-	-	8	Y
7	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID5049576	6381-92-6	-	-		19	Y	-	-	25251	31	Y
8	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID0034564	12276-01-6	-	-		11	-	-	-	-	8	Y
9	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID5027774	15708-41-5	-	-		48	Y	1.98	6/303	241	53	-
10	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID2036409	62-33-9	4.64e-06	Y		37	Y	0.0	0/64	25251	42	-
11	C10H16N2O8	DTXCID00197424	C10H16N2C(OC(=O)C	DTXSID1051852	20846-91-7	-	-		36	Y	-	-	89	25	-
12	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID6042107	15375-84-5	-	-		25	Y	-	-	97	25	-
13	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID3036442	5964-35-2	-	-		23	Y	-	-	25251	25	-
14	C10H16N2O8	DTXCID00197424	C10H16N2C(OC(=O)C	DTXSID1051806	178949-82-1	-	-		22	Y	-	-	-	5	-
15	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID0065696	14025-21-9	-	-		22	Y	-	-	-	43	-
16	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID9027813	20824-56-0	-	-		21	Y	-	-	-	12	-
17	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID9027815	21265-50-9	-	-		20	Y	-	-	241	24	-
18	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID5058272	17421-79-3	-	-		19	Y	-	-	25251	25	-
19	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID3058612	2001-94-7	-	-		18	Y	-	-	25251	19	-
20	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID8027820	22473-78-5	-	-		16	Y	-	-	-	11	-
21	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID8058324	17572-97-3	-	-		15	-	-	-	-	36	-
22	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID8028343	67859-51-2	-	-		14	Y	-	-	-	5	-
23	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID4051328	13235-36-4	-	-		14	-	-	-	-	18	-
24	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID6070980	68015-77-0	-	-		14	Y	-	-	-	13	-
25	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID9058317	15934-01-7	-	-		11	-	-	-	-	5	-
26	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID0066163	17099-81-9	-	-		11	-	-	-	241	14	-
27	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID1068988	54959-35-2	-	-		11	-	-	-	241	14	-
28	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID5074266	60816-63-9	-	-		11	-	-	-	1	10	-
29	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID4048197	39208-15-6	-	-		10	-	-	-	-	28	-
30	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID2065830	14931-83-0	-	-		10	-	-	-	47	9	-
31	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID70189997	36499-65-7	-	-		10	-	-	-	25298	26	-
32	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID7051420	61916-40-3	-	-		9	-	-	-	-	4	-
33	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID2051425	73513-47-0	-	-		8	Y	-	-	-	3	-
34	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID7051426	73637-19-1	-	-		8	Y	-	-	-	5	-
35	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID2051427	73637-20-4	-	-		8	Y	-	-	-	-	-
36	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID3058741	10378-23-1	-	-		8	Y	-	-	-	31	-
37	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID6065925	15708-48-2	-	-		8	-	-	-	-	19	-
38	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID20217976	6766-87-6	-	-		8	-	-	-	-	13	-
39	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID5065807	14689-29-3	-	-		7	-	-	-	-	12	-
40	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID6069408	60544-70-9	-	-		7	-	-	-	-	12	-
41	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID00153984	123354-94-9	-	-		7	-	-	-	2	6	-
42	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	DTXSID70190705	37209-61-3	-	-		7	-	-	-	6	9	-
43	C10H16N2O8	DTXCID902977	C10H16N2C(OC(=O)C	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	
	DTXSID0034564	EDTA, copper salt	12276-01-6	Level 1	10	11	8	0	413.918561	<input type="radio"/>
	DTXSID3026350	Ethylenediaminetetraacetic acid tetrasodium salt	64-02-8	Level 1	1227	37	57	0	380.018442	<input type="radio"/>
	DTXSID5049576	Disodium ethylenediaminetetraacetate dihydrate	6381-92-6	Level 1	93	19	31	25251	372.075683	<input type="radio"/>
	DTXSID5049609	Ethylenediaminetetraacetic acid, diammonium copper salt	67989-88-2	Level 2	9	20	8	0	387.057712	<input type="radio"/>
	DTXSID6022977	Ethylenediaminetetraacetic acid	60-00-4	Level 1	346	71	158	25251	292.090665	<input type="radio"/>
	DTXSID7020556	Trisodium ethylenediaminetetraacetate	150-38-9	Level 1	85	30	33	0	358.036498	<input type="radio"/>
	DTXSID9027073	Ethylenediaminetetraacetic acid, disodium salt	139-33-3	Level 1	1358	41	56	25251	336.054554	<input type="radio"/>

# 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-phenoxyimidine	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-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 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

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

Posted: 11/14/2016

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

## [DSSTox SDF File](#)

Posted: 12/14/2016

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

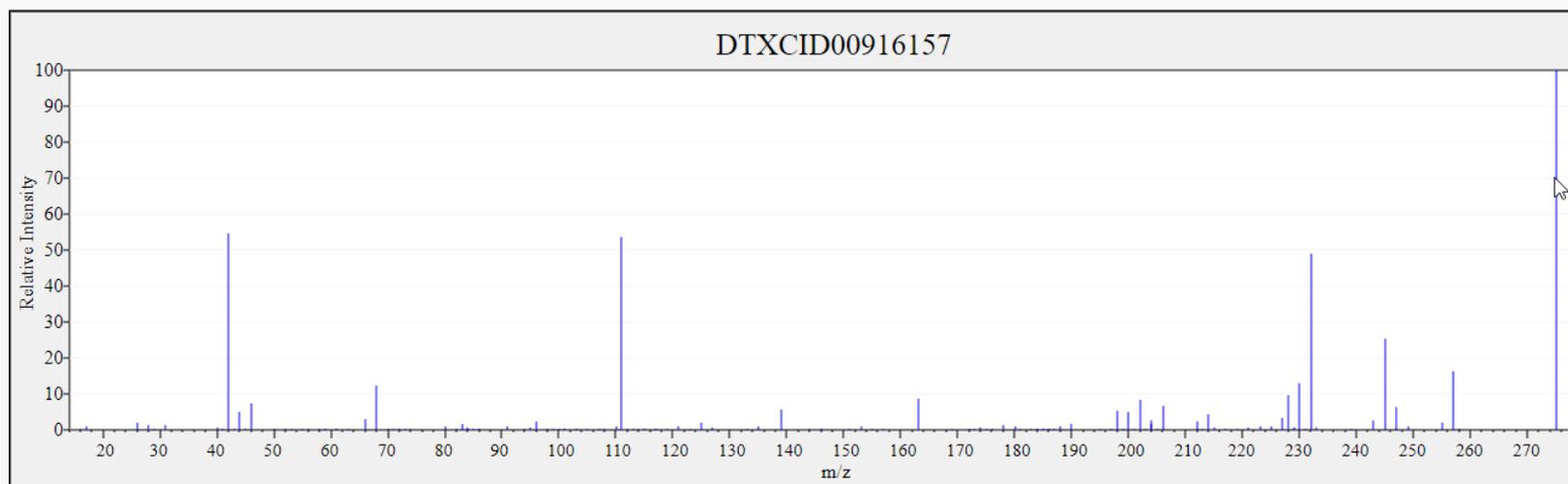
- **CFM-ID**
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- **Retention Time Index Prediction**
- **Structure/substructure/similarity search**
- **Generation of MS-ready structures:**
  - Upload file, download results
  - Service based generation

# Predicted Mass Spectra

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

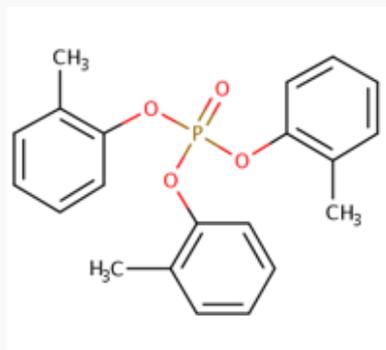


- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >700,000 structures, to be accessible via Dashboard

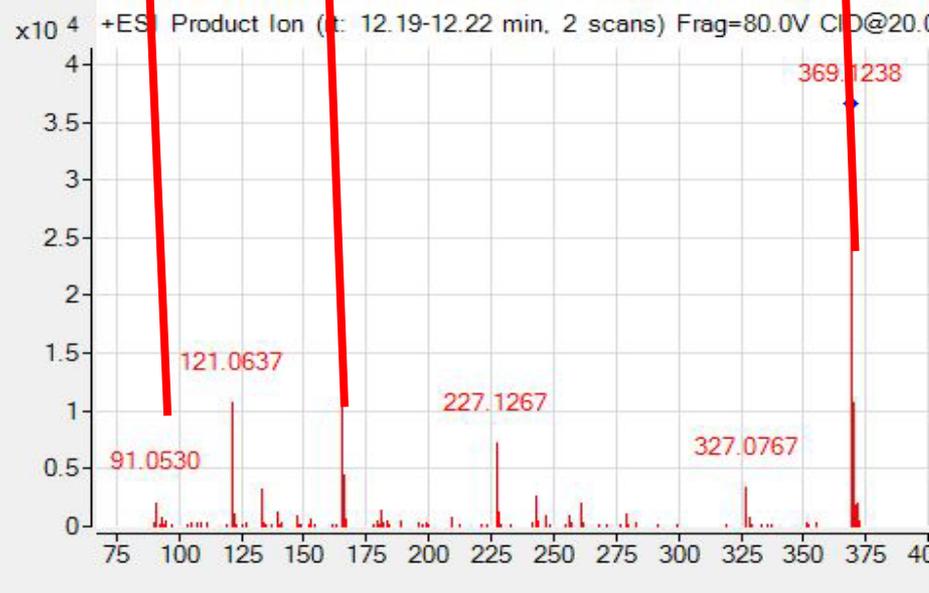
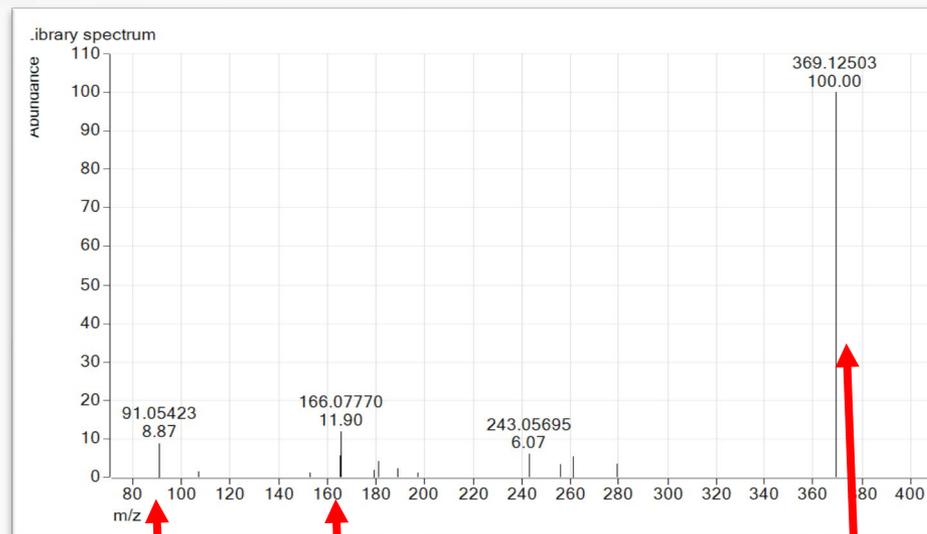


# Predicted Mass Spectra

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:  Da ppm

# Prototype Development

## AADashboard

atrazine

Search



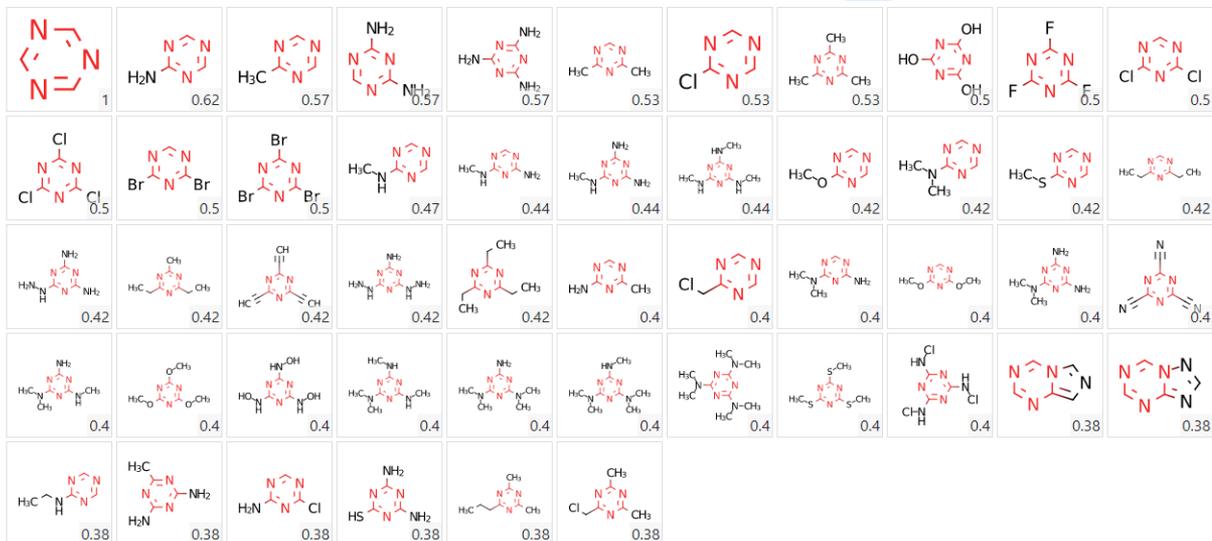
### Select properties to predict

- H  
C  
N  
O
- T.E.S.T. 18 OPERA Search
- Exact
  - Substructure

Search result 2540

Show  Isotopically Labeled  Charged  Salts or Mixtures

Sort Similarity



# Prototype Development

atrazine Search

100%

Select properties to predict

H T.E.S.T. 18 OPERA Search

C

N

O

S

P

F

Cl

Br

Exact

Substructure

Similarity

Molecular Formula

Molecular Weight

Input formula (e.g. C6 H6):

C15H16O2

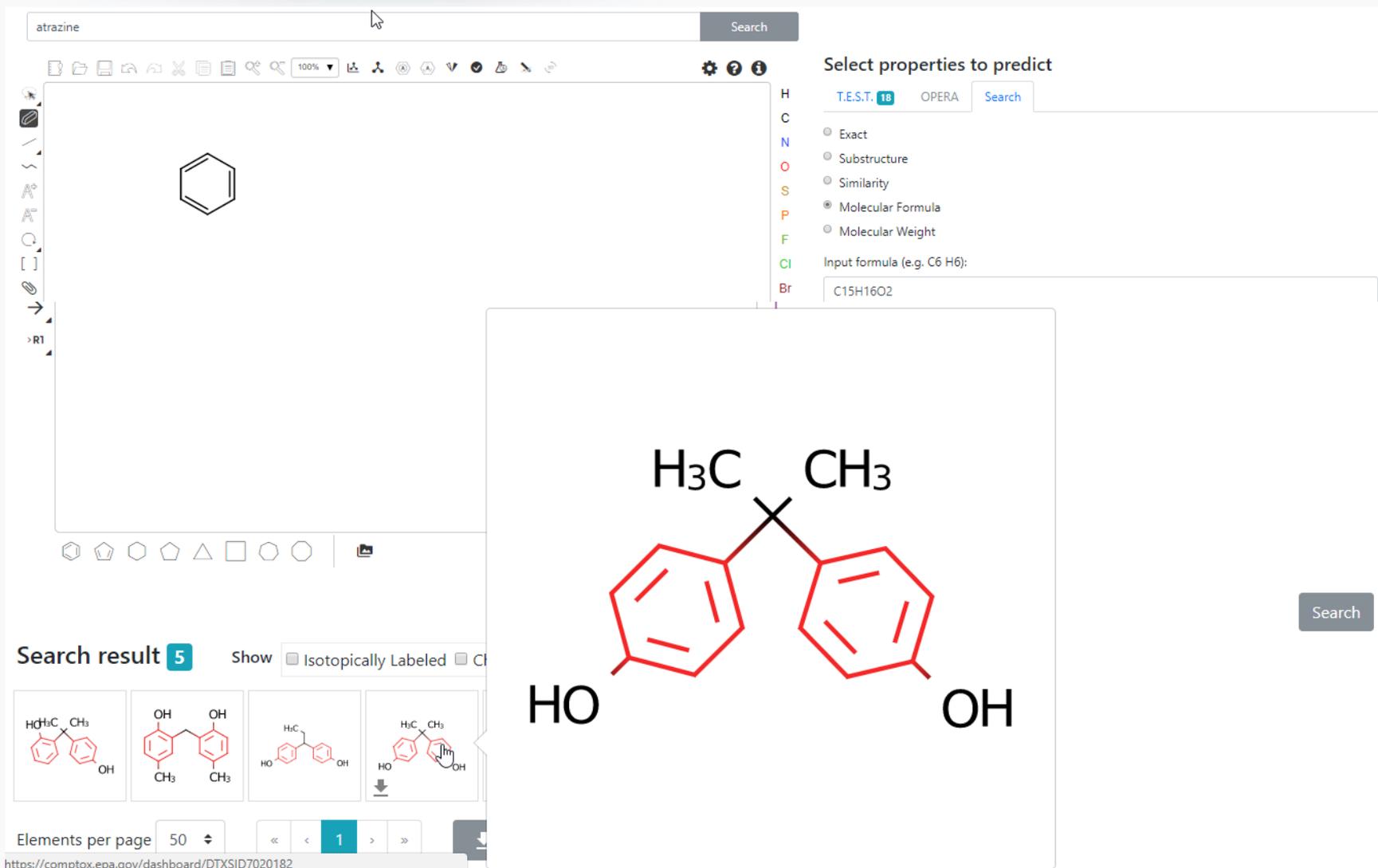
Search

Search result 5

Show  Isotopically Labeled  Cl

Elements per page 50

1



- The CompTox Chemicals Dashboard provides access to data for ~875,000 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
- 2 years development as a CompTox Integration Hub

- IT Development team – especially Jeff Edwards and Jeremy Dunne
- Chris Grulke for the ChemReg system
- NERL colleagues – Jon Sobus, Elin Ulrich, Mark Strynar, Seth Newton

## **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>