

# Web-based access to experimental and predicted data for environmental fate, transport and toxicity data

*Antony Williams<sup>1</sup>, Todd Martin<sup>2</sup>, Valery Tkachenko<sup>3</sup>,  
Kamel Mansouri<sup>4</sup> and Chris Grulke<sup>1</sup>*

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

*2) National Risk Management Research Laboratory, U.S. Environmental Protection Agency, Cincinnati, OH*

*3) Science Data Software, LLC, Rockville, MD 20850*

*4) Integrated Laboratory Systems, Research Triangle Park, North Carolina, United States*

*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*

- The EPA CompTox Chemistry Dashboard provides access to data associated with ~760,000 chemical substances. The available data includes experimental and predicted physicochemical properties, environmental fate and transport data, *in vivo* and *in silico* toxicity data, *in vitro* bioassay data, exposure data and a variety of other types of information. The data are under continuous expansion and curation and the experimental data have been used to develop QSAR and QSPR models. A number of these models are available via a web interface so that users can submit a chemical structure and predict properties in real time. The dashboard also provides access to pre-compiled chemical lists and categories, including pesticides, and chemicals detected in the environment via non-targeted mass spectrometry analysis. The data are searchable using chemical identifiers (systematic names, trade names, CAS Registry Numbers), by structure, mass and formula. Batch searches allow for data associated with thousands of chemicals to be obtained in a few seconds, with just a few button clicks, and downloaded to the desktop. This presentation will provide an overview of the Dashboard and its applications to accessing source data associated with agriculturally related chemicals.


# Who is NCCT?

- National Center for Computational Toxicology – part of EPA's Office of Research and Development
- Research driven by EPA's *Chemical Safety for Sustainability Research Program*
  - Develop new approaches to **evaluate** the **safety** of chemicals
  - Integrate advances in biology, biotechnology, chemistry, exposure science and computer science
- Goal - To identify **chemical exposures** that may disrupt biological processes and cause adverse outcomes.
- Prediction models and predicted data are some of our major outputs

- A publicly accessible website delivering access:
  - ~762,000 chemicals with related property data
  - Searchable by chemical, product use, gene and assay
  - Experimental and **predicted** physicochemical property data
  - “Bioactivity data” for the ToxCast/Tox21 project – plus derived **models**
  - Generalized **Read-Across** (GenRA) module
  - “Batch searching” of **predicted** data for thousands of chemicals


# CompTox Dashboard

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

 United States  
Environmental Protection  
Agency

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Share ▼



762 Thousand Chemicals

Chemicals Product/Use Categories Assay/Gene

Search for chemical by systematic name, synonym, CAS number, DTXSID 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**

March 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


Share ▼

762 Thousand Chemicals


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




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



Bisphenol A propoxylate diglycidyl ether  
DTXSID10399098



Bisphenol A propoxylate glycerolate diacrylate  
DTXSID40400126

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

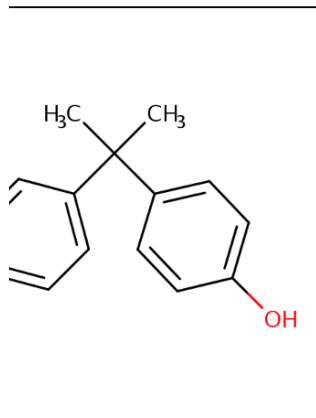
LINKS

COMMENTS

Bisphenol A

DTXSID7020182

3STox Substance Id.



Batch Search Lists Predictions Downloads

Copy

Share

Submit Comment

Search all data

### 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. It has been in commercial use since 1957.

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

# Physicochemical properties

## Property



Summary



Summary

LogP: Octanol-Water

Melting Point

Boiling Point

Water Solubility

Vapor Pressure

Flash Point

Surface Tension

Index of Refraction

Molar Refractivity

Polarizability

Density

Molar Volume

Thermal Conductivity

Viscosity

Henry's Law

LogKoa: Octanol-Air

## Summary

Search query

d average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
		3.43	3.32	2.40 to 3.64	
	156	138	153 to 156	125 to 157	°C
		360	200	343 to 401	°C
		1.00e-3	5.26e-4	5.44e-4 to 1.31e-3	mol/L
		3.43e-7	-	6.83e-8 to 2.59e-6	mmHg
		190	-	188 to 192	°C
			-	46.0	dyn/cm
			-	1.60	
			-	68.2	cm^3
			-	27.0	Å^3
		1.17	-	1.14 to 1.20	g/cm^3
			-	200	cm^3
			-	150	mW/(m*K)



## An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling

K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson & A. J. Williams

To cite this article: K. Mansouri, C. M. Grulke, A. M. Richard, R. S. Judson & A. J. Williams (2016)

An automate  
datasets use  
DOI: [10.1081](https://doi.org/10.1081)

Mansouri et al. *J Cheminform* (2018) 10:10  
<https://doi.org/10.1186/s13321-018-0263-1>

 Journal of Cheminformatics

To link to th

RESEARCH ARTICLE

Open Access



## OPERA models for predicting physicochemical properties and environmental fate endpoints

Kamel Mansouri<sup>1,2,3\*</sup> , Chris M. Grulke<sup>1</sup>, Richard S. Judson<sup>1</sup> and Antony J. Williams<sup>1</sup>

# Curation to QSAR Ready Files

Property	Initial file	Curated Data	Curated QSAR ready
AOP	818	818	745
BCF	685	618	608
BioHC	175	151	150
Biowin	1265	1196	1171
BP	5890	5591	5436
HL	1829	1758	1711
KM	631	548	541
KOA	308	277	270
LogP	15809	14544	14041
MP	10051	9120	8656
PC	788	750	735
VP	3037	2840	2716
WF	5764	5076	4836
WS	2348	2046	2010

# Detailed OPERA Prediction Reports

Source

Result

Calculation Details

Experimental Values

PhysPropNCCT

Predicted Values

EPISUITE

NICEATM

ACD/Labs Conse

ACD/Labs

OPERA

OPERA Models: LogP: Octanol-Water

Bisphenol A

80-05-7 | DTXSID7020182

Cc1ccc(cc1)C(C)(C)c2ccc(O)cc2

Model Results

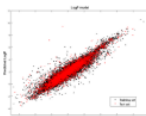
Predicted value: 3.35

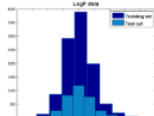
Global applicability domain: Inside

Local applicability domain index: 0.88

Confidence level: 0.75

Model Performance





Weighted KNN model

QMRP

5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.85	0.89	0.85	0.87	0.88	0.78

Nearest Neighbors from the Training Set

Cc1ccc(cc1)C(C)(C)c2ccc(O)cc2

Bisphenol A  
Measured: 3.32  
Predicted: 3.35

CC(O)(C(=O)O)c1ccc(cc1)-c2ccccc2

BUTANOIC ACID,2-(4-BIPHENYL)-3-HYDROXY-  
Measured: 3.25  
Predicted: 3.45

CC(=O)C(O)c1ccc(cc1)-c2ccccc2

Flurbiprofen  
Measured: 4.18  
Predicted: 3.83

Cc1ccc(cc1)C(C)(C)C(=O)O

2,2-Diphenylpropionic acid  
Measured: 2.89  
Predicted: 2.93

CCC(O)C(=O)Oc1ccc(cc1)-c2ccccc2

3-OH-2-(4-BIPHENYL)HEXANOIC ACID  
Measured: 3.75  
Predicted: 3.88

10

# Prediction Details and QMRF Report

## Model Results

**Predicted value:** 144 °C

**Global applicability domain:** Inside ⓘ

**Local applicability domain index:** 0.91 ⓘ

**Confidence level:** 0.65 ⓘ


Applicability domain using the leverage approach. All training set space considered. More details in QMRF.

QMRF\_NCCT\_MP\_08212016 - Adobe Acrobat Pro

File Edit View Window Help

Create [Icons]

1 / 10 [Navigation Icons] 143% [Zoom] [Tools] Fill & Sign Comment


	<b><i>QMRF identifier (JRC Inventory):</i></b> To be entered by JRC
	<b><i>QMRF Title:</i></b> MP: Melting point prediction from the NCCT Models Suite.
	<b><i>Printing Date:</i></b> May 4, 2016

**1.QSAR identifier**

**1.1.QSAR identifier (title):**  
MP: Melting point prediction from the NCCT\_Models Suite.

**1.2.Other related models:**  
No related models

**1.3.Software coding the model:**  
NCCT\_models V1.02  
Suite of QSAR models to predict physicochemical properties and environmental fate of organic chemicals


**EPA**
US Environmental Protection Agency

[Español](#) | [中文: 繁體版](#) | [中文: 简体版](#) | [Tiếng Việt](#) | [한국어](#)

[Learn the Issues](#)
[Science & Technology](#)
[Laws & Regulations](#)
[About EPA](#)

[Related Topics: Safer Chemicals Research](#)
[Contact Us](#)
[Share](#)

## Toxicity Estimation Software Tool (TEST)

On this page:

- [QSAR Methodologies](#)
- [What's New in Version 4.2?](#)
- [Prior Version History](#)
- [System Requirements](#)
- [Installation Instructions](#)
- [Publications](#)
- [Get Email Alerts](#)

---

The Toxicity Estimation Software Tool (TEST) was developed to allow users to easily estimate the toxicity of chemicals using Quantitative Structure Activity Relationships (QSARs) methodologies. QSARs are mathematical models used to predict measures of toxicity from the physical characteristics of the structure of chemicals (known as molecular descriptors). Simple QSAR models calculate the toxicity of chemicals using a simple linear function of molecular descriptors:

### Ask a Technical Expert

Got a question about our research model? Want to give us feedback? Contact a technical expert about [TEST](#).

# Physical properties in TEST

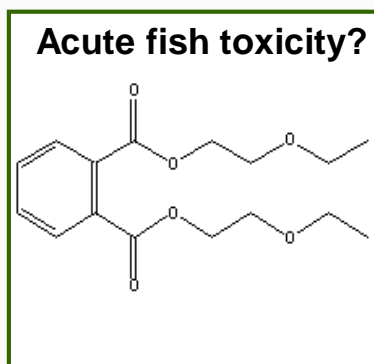
Endpoint	Definition
Viscosity	A measure of the resistance of a fluid to flow (cP) defined as the proportionality constant between shear rate and shear stress
Surface tension	A property of the surface of a liquid (dyn/cm) that allows it to resist an external force
Water solubility	The amount of a chemical (mg/L) that will dissolve in liquid water to form a homogeneous solution

- Predictions and models expand outside of simply physicochemical and environmental fate and transport
- Examples
  - Read-across for Toxicity Endpoints
  - Quantitative Structure–Use Relationship (QSUR) models
  - High-Throughput Toxicokinetics (HTTK)
  - Models based on high throughput bioactivity data

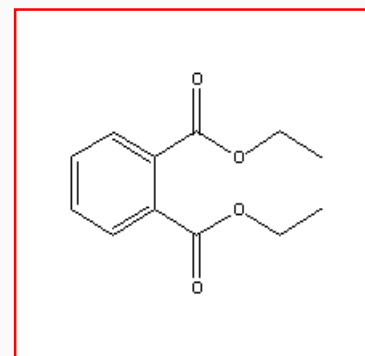
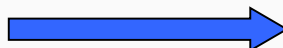
# Definitions: Read-Across

- Known information on the property of a substance (source) is used to make a prediction of the same property for another substance (target) that is considered “similar”

	Source chemical	Target chemical	
Property	● → ○		● Reliable data ○ Missing data



**Known to be harmful**

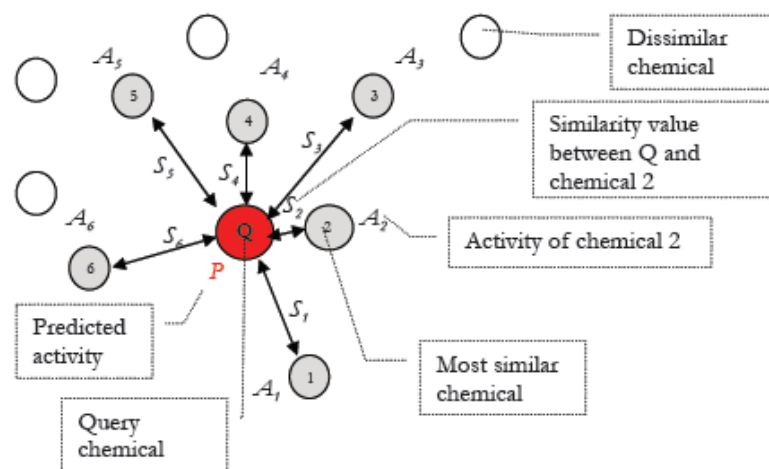


**Predicted to be harmful**



# GenRA (Generalised Read-Across)

- Predicting toxicity as a similarity-weighted activity of nearest neighbors based on chemistry and/or bioactivity descriptors
- Goal: to systematically evaluate read-across performance and uncertainty using available data
- The approach enabled a performance baseline for read-across predictions of toxicity effects within specific study outcomes to be established



# GenRA (Generalised Read-Across)

## Fluconazole

86386-73-4 | DTXSID3020627

Searched by DSSTox Substance Id.

### DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

► EXPOSURE

► BIOACTIVITY

SIMILAR COMPOUNDS

GENRA

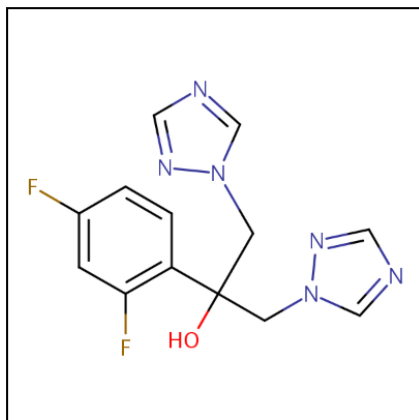
RELATED SUBSTANCES

SYNONYMS

► LITERATURE

LINKS

COMMENTS





### Wikipedia

Fluconazole is an antifungal medication used for a number of fungal infections. This includes candidiasis, blastomycosis, coccidioidomycosis, cryptococcosis, histoplasmosis, dermatophytosis, and pityriasis versicolor. It is also used to prevent candidiasis in those who are at high risk such as following organ transplantation, low birth weight babies, and those with low blood neutrophil counts. It is given either by mouth or by injection into a vein.


Common side effects include vomiting

[Read more](#)

### Intrinsic Properties

 Molecular Formula: C<sub>13</sub>H<sub>12</sub>F<sub>2</sub>N<sub>6</sub>O  Mol File

 Find All Chemicals

 Average Mass: 306.277 g/mol  Isotope Mass Distribution

 Monoisotopic Mass: 306.104065 g/mol

### Structural Identifiers

### Linked Substances

### Presence in Lists

### Record Information

### Quality Control Notes

# GenRA (Generalised Read-Across)

## Fluconazole

86386-73-4 | DTXSID3020627

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

► EXPOSURE

► BIOACTIVITY

SIMILAR COMPOUNDS

**GENRA**

RELATED SUBSTANCES

SYNONYMS

► LITERATURE

LINKS

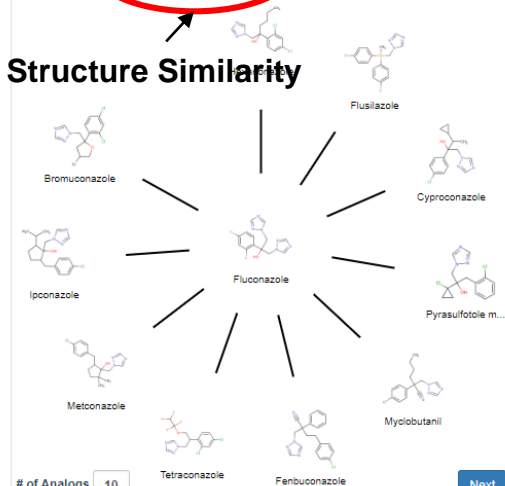
COMMENTS

### Step One: Analog Identification and Evaluation

Neighbors by Chem: Morgan Fgrprts

Filter by: invivo data

### Structure Similarity



# of Analogs

10

Next

# GenRA (Generalised Read-Across)

GenRA

## Step Two: Data Gap Analysis & Generate Data Matrix

Neighbors by: Chem: Morgan Fgrpts

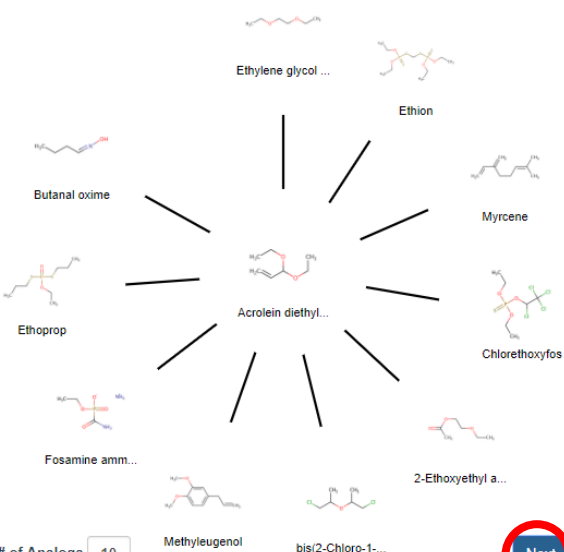
Filter by: invivo data

Summary Data Gap Analysis

Group: ToxRef

By: Tox Fingerprint

Generate Data Matrix



Next

		bio h21	bio hct	chm_ct	tox brf
Fluconazole	3	714	15	0	
Hexaconazole	43	819	18	345	
Flusilazole	28	819	9	345	
Cyproconazole	14	819	16	408	
Pyrasulfotole metabolite ...	0	0	18	234	
Myclobutanil	15	818	15	345	
Fenbuconazole	34	819	17	345	
Tetraconazole	35	819	20	345	
Metconazole	35	215	15	82	
Ipconazole	46	232	16	180	
Bromuconazole	24	277	13	345	

	Fluconazole	Hexaconazole	Flusilazole	Cyproconazole	Pyrasulfotole metab...	Myclobutanil	Fenbuconazole	Tetraconazole	Metconazole	Ipconazole	Bromuconazole
CHR:Abdominal Cavity											
CHR:Adrenal Gland											
CHR:Artery (General)											
CHR:Auditory Startle Re...											
CHR:Bile duct											
CHR:Blood											
CHR:Blood vessel											
CHR:Body Weight											
CHR:Bone											
CHR:Bone Marrow											
CHR:Brain											
CHR:Chus											

Data gap analysis

# GenRA (Generalised Read-Across)

GenRA

Step Three: Run GenRA Prediction

Neighbors by: Chem: Morgan Fgrprts Filter by: invivo data Summary Data Gap Analysis Group: ToxRef By: Tox Fingerprint Run Read-Across

Chemical structures shown: Ethylene glycol diethyl acetal, Ethion, Butanal oxime, Myrcene.

Summary Data Gap Analysis table:

	bio_tox21	bio_tox27	chem_ct	tox_bcr
Acrolein diethylacetal	14	0	4	0
Ethylene glycol diethyl acetal	7	0	4	95

Similarity Weight: 0.39 ✓, 0.31 ✓, 0.21 ✓, 0.21 ✓, 0.20 ✓

Target: Fluconazole

Source analogues: Hexaconazole, Flusilazole, Cyproconazole, Pyrasulfotole m..., Myclobutanil, Fenbuconazole, Tetraconazole, Metoconazole, Ipconazole, Bromuconazole

Run GenRA

Run Read-Across

Download: Filetype

	Fluconazole	Hexaconazole	Flusilazole	Cyproconazole	Pyrasulfotole m...	Myclobutanil	Fenbuconazole	Tetraconazole	Metoconazole	Ipconazole	Bromuconazole
CHR:Abdominal Cavity											
CHR:Adrenal Gland											
CHR:Artery (General)											
CHR:Auditory Startle Re...											
CHR:Bile duct											
CHR:Blood											
CHR:Blood vessel											
CHR:Body Weight											
CHR:Bone											

- Singleton searches are useful but we work with thousands of chemicals!
- Typical questions
  - What are the SMILES strings for a list of 1000 chemicals?
  - Do any of this list of chemicals have XXX type of data?
  - What are the predicted logP values for a list of chemicals?
  - Can I get lists of predicted properties in Excel files? In SDF files?

# Batch Searching








## 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 

☐ 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



# Excel Output

	A	B	C	D	E	F
1	DTXSID	PREFERRED_NAME	VAPOR_PRESSURE	WATER_SOLUBILITY	ATMOSPHERIC_HYDRO	BIOCONCENTRATION
2	<a href="#">DTXSID0020523</a>	2,4-Dinitrophenol	0.000058479	0.0055847	8.57649E-13	5.31642
3	<a href="#">DTXSID0021256</a>	Sulfasalazine	6.85488E-14	1.07895E-06	1.05999E-11	13.1817
4	<a href="#">DTXSID0021333</a>	Tetramethylthiuram mon-		0.0076913	2.89566E-11	5.1961
5	<a href="#">DTXSID0021337</a>	Thiabendazole	7.06318E-07	0.000156675	4.3347E-11	7.89835
6	<a href="#">DTXSID0022436</a>	Diphenolic acid	3.40408E-08	0.00125893	1.70869E-11	2.94716
7	<a href="#">DTXSID0023745</a>	D-Xylose	7.60326E-08	3.52371	3.57198E-11	1.66292
8	<a href="#">DTXSID0023901</a>	Bentazone	-	5.40754E-05	1.67809E-11	1.4173
9	<a href="#">DTXSID0023951</a>	Carboxin	1.06905E-07	0.000304789	1.79673E-11	10.2339
10	<a href="#">DTXSID0024216</a>	Nitrapyrin	0.0107895	0.000162181	2.28302E-11	661.452
11	<a href="#">DTXSID0024345</a>	Triasulfuron	1.47231E-11	0.00022856	1.3866E-11	11.1975
12	<a href="#">DTXSID0027983</a>	Dipropylene glycol mono-		-	-	-
13	<a href="#">DTXSID0032493</a>	Triadimenol	1.58489E-07	0.000267301	1.65682E-11	111.018
14	<a href="#">DTXSID0032520</a>	Azoxystrobin	9.4189E-11	1.53815E-06	1.6811E-11	12.6717
15	<a href="#">DTXSID0032601</a>	Cyproconazole	5.24807E-08	0.000289734	1.66415E-11	43.1754
16	<a href="#">DTXSID0032655</a>	Triticonazole	9.88553E-09	0.000014388	2.98831E-11	102.292
17	<a href="#">DTXSID0039223</a>	Aldicarb	0.000138676	0.0254683	1.5544E-11	5.90478
18	<a href="#">DTXSID0040707</a>	4-Pentylaniline	0.00272898	0.000567545	2.73005E-11	94.115
19	<a href="#">DTXSID0041270</a>	3,5,5-Trimethylhexyl ace	0.21727	0.000521195	1.32332E-11	43.3424
20	<a href="#">DTXSID0044361</a>	Hexanedihydrazide	-	0.341193	5.54119E-12	1.78013
21	<a href="#">DTXSID0044494</a>	2-Phenylethyl phenylace	0.000101859	0.000628058	1.43965E-11	342.896

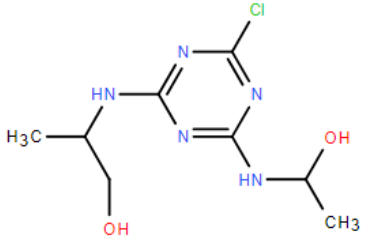
# Real-Time Predictions

 United States  
Environmental Protection  
Agency

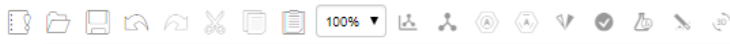
Home Advanced Search Batch Search Lists ▼ Predictions Downloads


Share Search all data


Atrazine

CN1C=NC2=C(N1)N(C)C(=N2)C3=NC=CC(=C3)N(C)C


100%







Chiral

 Select properties to predict

H  
C  
N  
O  
S  
P  
F  
Cl  
Br  
I  
PT

T.E.S.T.

☒ Toxicological properties

☒ 96 hour fathead minnow LC50

☒ 48 hour D. magna LC50

☒ 48 hour T. pyriformis IGC50

☒ Oral rat LD50

☒ Bioaccumulation factor

☒ Developmental toxicity

☒ Ames mutagenicity

☒ Estrogen Receptor RBA

☒ Estrogen Receptor Binding

☒ Physical properties

☒ Normal boiling point

☒ Melting point

☒ Flash point

☒ Vapor pressure

☒ Density

☒ Surface tension

☒ Thermal conductivity

☒ Viscosity

Calculate

# Real-Time Predictions



Provider: T.E.S.T.

[Calculate](#)

[Download Summary](#)

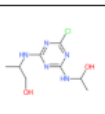
Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		3.358 -Log10(mol/L) 108.625 mg/L	3.378 -Log10(mol/L) 103.737 mg/L	3.235 -Log10(mol/L) 144.209 mg/L	2.755 -Log10(mol/L) 435.012 mg/L	4.064 -Log10(mol/L) 21.394 mg/L
48 hour D. magna LC50		3.276 -Log10(mol/L) 131.155 mg/L	3.091 -Log10(mol/L) 201.017 mg/L	3.021 -Log10(mol/L) 235.873 mg/L	2.533 -Log10(mol/L) 725.991 mg/L	4.460 -Log10(mol/L) 8.596 mg/L
48 hour T. pyriformis IGC50			2.537 -Log10(mol/L) 718.880 mg/L			
Oral rat LD50		1.993 -Log10(mol/kg) 2514.643 mg/kg	1.888 -Log10(mol/kg) 3203.747 mg/kg			2.099 -Log10(mol/kg) 1973.760 mg/kg
Bioaccumulation factor		0.386 Log10 2.429	0.473 Log10 2.971	0.386 Log10 2.432	-0.290 Log10 0.512	0.974 Log10 9.409
Developmental toxicity		true	true	true		
Ames mutagenicity		false	true			false
Estrogen Receptor RBA						
Estrogen Receptor Binding		false	false	false	false	
Normal boiling point		371.7 °C	360.4 °C		470.4 °C	284.3 °C
Melting point		162.3 °C	163.0 °C		158.5 °C	165.3 °C
Flash point		225.4 °C	250.7 °C		249.9 °C	175.4 °C
Vapor pressure		-7.853 Log10(mmHg) 1.403*10 <sup>-8</sup> mmHg	-7.672 Log10(mmHg) 2.126*10 <sup>-8</sup> mmHg		-9.429 Log10(mmHg) 3.728*10 <sup>-10</sup> mmHg	-6.458 Log10(mmHg) 3.486*10 <sup>-7</sup> mmHg

# Real-Time Predictions

**Predicted Oral rat LD50 for C1C=1N=C(N=C(N1)NC(C)CO)NC(O)C from Consensus method**

Prediction results		
Endpoint	Experimental value	Predicted value
Oral rat LD <sub>50</sub> -Log10(mol/kg)	N/A	1.99
Oral rat LD <sub>50</sub> mg/kg	N/A	2514.64

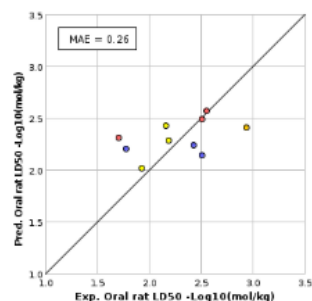
Individual Predictions	
Method	Predicted value -Log10(mol/kg)
Hierarchical clustering	1.89
Nearest neighbor	2.10



**Predictions for the test chemical and for the most similar chemicals in the external test set**

If the predicted value matches the experimental values for similar chemicals in the test set (and the similar chemicals were predicted well), one has greater confidence in the predicted value.

Prediction results (colors defined in table below)



Chemicals	MAE*
Entire set	0.43
Similarity coefficient $\geq 0.5$	0.26

\*Mean absolute error in -Log10(mol/kg)

CAS	Structure	Similarity Coefficient	Experimental value -Log10(mol/kg)	Predicted value -Log10(mol/kg)
<chem>C1C=1N=C(N=C(N1)NC(C)CO)NC(O)C</chem> (test chemical)			N/A	1.99
<a href="#">1912-24-9</a>		0.84	2.51	2.15
<a href="#">3004-71-5</a>		0.83	2.43	2.24
<a href="#">139-40-2</a>		0.82	1.78	2.20
<a href="#">1610-17-9</a>		0.76	2.16	2.43
<a href="#">1014-69-3</a>		0.74	2.19	2.28
<a href="#">22936-75-0</a>		0.71	1.93	2.02

# API in development

## Prototype services available

<https://comptox.epa.gov/dashboard/web-test/WS?smiles=CCO&method=hc>

JSONRaw DataHeaders

SaveCopy

uuid:

"55547f4f-f966-48e8-b831-a0d217998064"

predictionTime:

1520539090089

software:

"T.E.S.T (Toxicity Estimation Software Tool)"

softwareVersion:

"5.01"

condition:

"25°C"

endpoint:

"Water solubility at 25°C"

method:

"Hierarchical clustering"

▼ predictions:

▼ 0:

id:

"C\_1520539090089"

smiles:

"OCC"

expValMolarLog:

"-1.337"

expValMass:

"1001180.703"

predValMolarLog:

"-1.338"

predValMass:

"1002625.241"

molarLogUnits:

"-Log10(mol/L)"

massUnits:

"mg/L"

# Our support for FAIR Data

F<sub>indable</sub>



A<sub>ccessible</sub>



I<sub>nteroperable</sub>



R<sub>eusable</sub>



# Downloadable Data



## DSSTox Mapping File

Posted: 12/14/2016

The DSSTOX mapping file contains mappings between the DSSTox substance identifier (DTXSID) and the associated InChI String and InChI Key. The file is made available as a Tab Separated Value (TSV) file with each entry represented as shown:

DTXSID7020001 InChI=1S/C11H9N3/c12-10-6-5-8-7-3-1-2-4-9(7)13-11(8)14-10/h1-6H,(H3,12,13,14) FJTNLJLPLJDTRM-UHFFFAOYSA-N



## DSSTox Predicted Property Data

Posted: 12/14/2016

A number of property prediction models were developed using curated data as described in the publication "[An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling](#)". These property prediction models include logP, water solubility, bioconcentration factor and many others. The files include DTXSIDs, names and the predicted properties where possible. The models cannot predict properties for all chemicals contained in the database (for example, inorganics, organometallics and elements cannot be handled).

## DSSTox Synonyms File

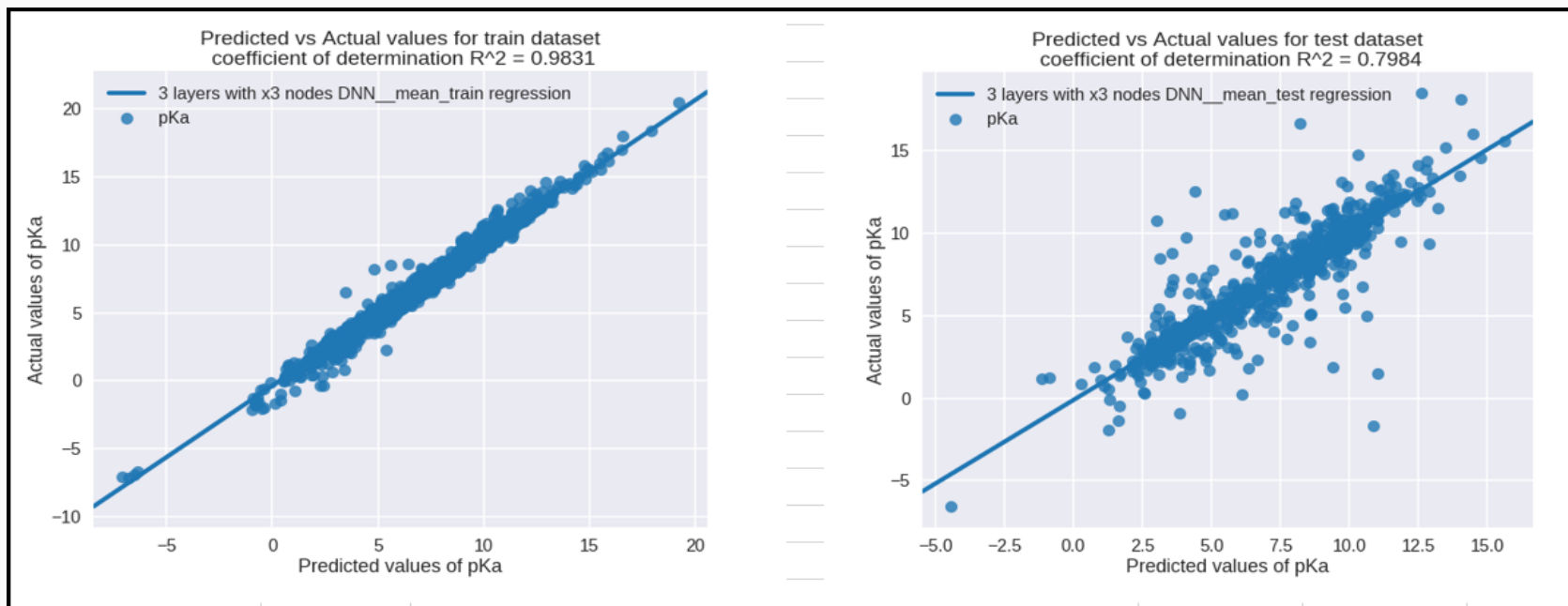
Posted: 12/14/2016

The DSSTox synonyms file is in SDF format and includes the DSSTox substance identifier (DTXSID). The preferred name, the CAS Registry Number and the list of associated synonyms for over 720,000 chemicals. 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.

- Present work in development
  - Development of OPERA model web services
  - Development of pKa and logD prediction models
  - Display of TEST Toxicity endpoints predicted data
  - Analytical Data support
    - Spectral searching against predicted Mass Spectra



- pKa prediction models based on Open Data Set of 8000 chemicals – acidic, basic and amphoteric chemicals



# Toxicity Endpoints in TEST

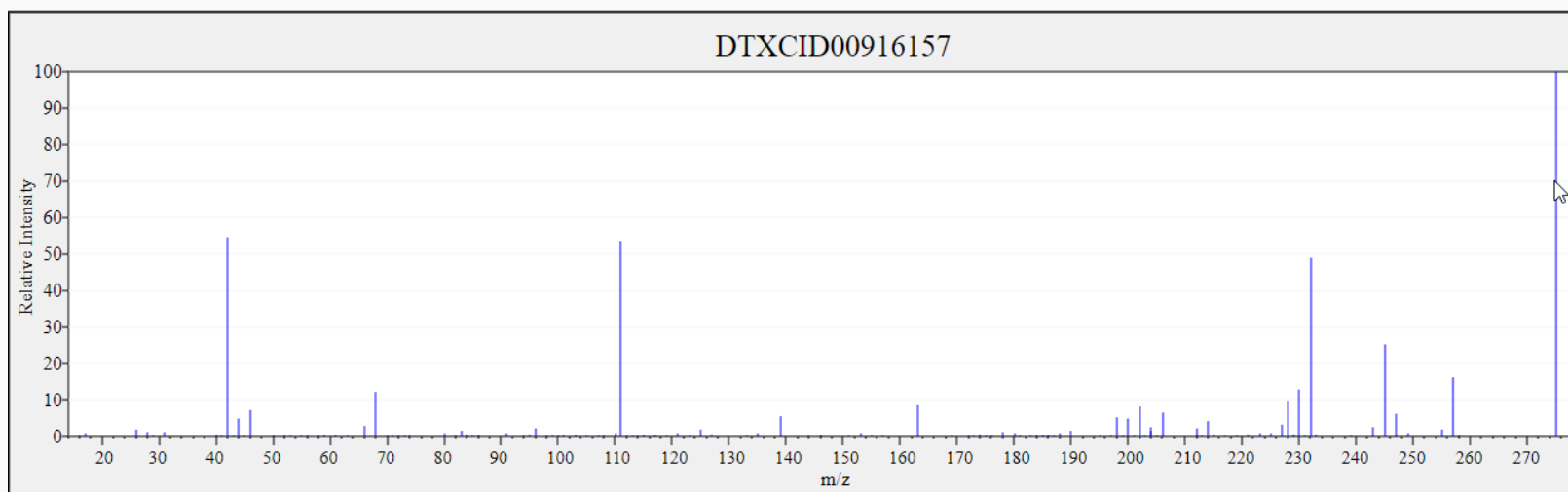
Endpoint	Definition
96 hour fathead minnow LC <sub>50</sub>	Concentration in mg/L that causes 50% of fathead minnow to die after 96 hours
48 hour <i>Daphnia magna</i> LC <sub>50</sub>	Concentration in mg/L that causes 50% of <i>Daphnia magna</i> to die after 48 hours
48 hour <i>T. pyriformis</i> IGC <sub>50</sub>	Concentration in mg/L that causes 50% growth inhibition to <i>T. pyriformis</i> after 48 hours
Oral rat LD <sub>50</sub>	Amount of chemical in mg/kg body weight that causes 50% of rats to die after oral ingestion
Developmental toxicity	Whether or not a chemical causes developmental toxicity effects to humans or animals
Ames mutagenicity	A compound is positive for mutagenicity if it induces revertant colony growth in any strain of <i>Salmonella typhimurium</i>

# Predicted Mass Spectra

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

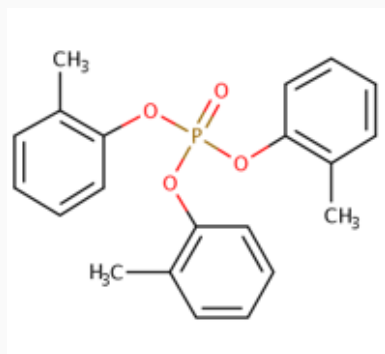


- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >700,000 structures, to be accessible via Dashboard

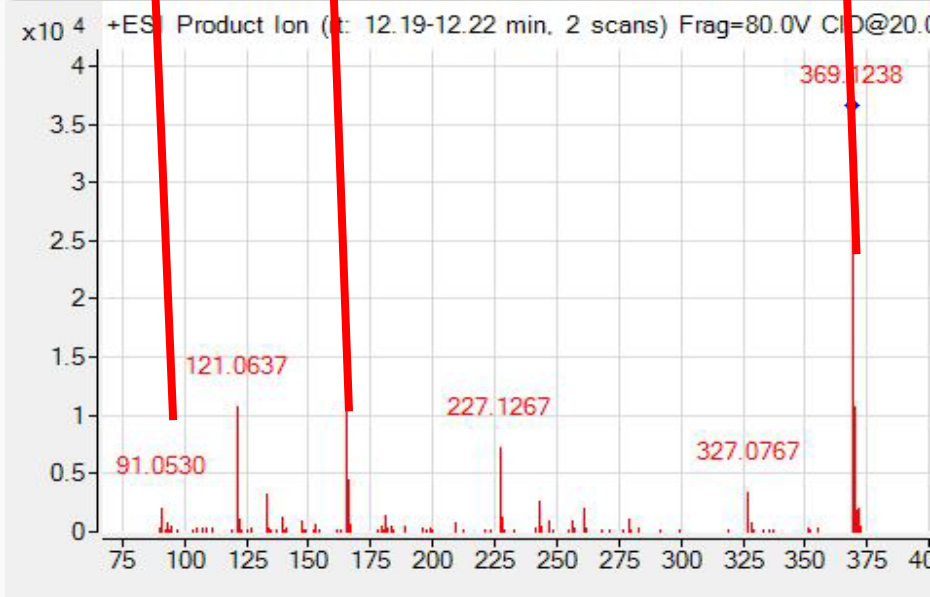
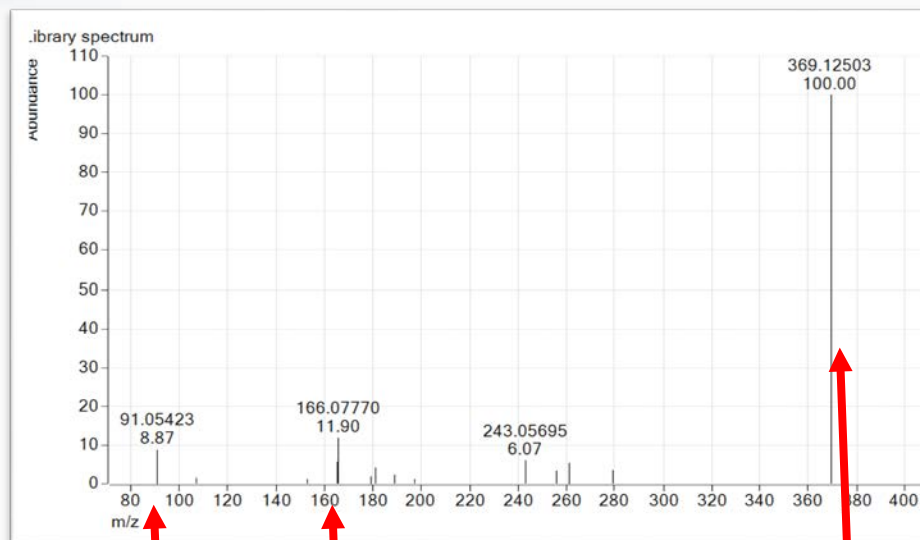


# Predicted Mass Spectra

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

- The CompTox Dashboard provides access to data for ~762,000 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 delivered
- Real time prediction models expanding in number
- Web services available for some physchem and toxicity endpoints

## **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>