

# CompTox Chemicals Dashboard providing access to experimental and predicted environmental fate and transport data

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

1) NCCT, U.S. Environmental Protection Agency, RTP, NC

2) Integrated Laboratory Systems, Research Triangle Park, NC.

3) NRMRL, U.S. Environmental Protection Agency, Cincinnati, OH


*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*

*Fall 2019  
ACS Fall Meeting, San Diego*

- The CompTox Chemicals Dashboard - web-based database of 875k substances
- Associated data including:
  - *In vivo* hazard data
  - *In vitro* bioactivity screening data
  - Link farm to tens of public resources
- Includes experimental and predicted physchem and experimental fate and transport data
- Access to real-time predictions
- A quick overview of capabilities...


# CompTox Chemicals Dashboard

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

 EPA  
United States  
Environmental Protection  
Agency

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Share ▼



875 Thousand Chemicals

Chemicals

Product/Use Categories

Assay/Gene

☐ Identifier substring search

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


Latest News

[Read more news](#)

**Journal of Cheminformatics article regarding "MS-Ready structures"**


March 9th, 2019 at 1:09:45 PM


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](#).




# BASIC Search

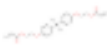
**Chemicals** **Product/Use Categories** **Assay/Gene**

 Bisphenol

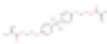


Bisphenol A  
DTXSID7020182

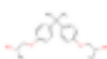




Bisphenol A bis(2-hydroxyethyl ether) diacrylate  
DTXSID6066991

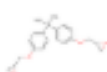


Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate  
DTXSID1066992

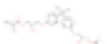


Bisphenol A bis(2-hydroxypropyl) ether  
DTXSID8051592

Bisphenol A carbonate polymer  
DTXSID6027840



Bisphenol A diglycidyl ether  
DTXSID6024624




Bisphenol A glycidyl methacrylate  
DTXSID7044841

# Detailed Chemical Pages

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

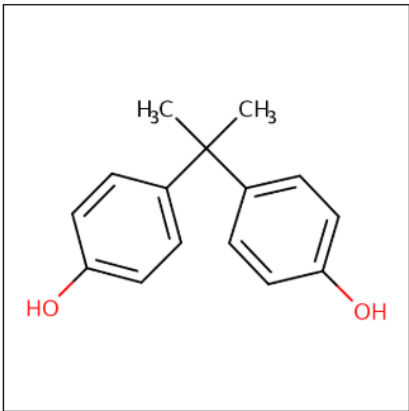
Copy Share Submit Comment Search all data



## Bisphenol A

80-05-7 | DTXSID7020182

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


**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**


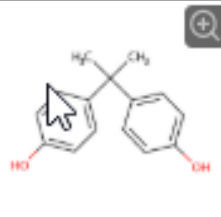
**Linked Substances**

**Presence in Lists**

**Record Information**

**Quality Control Notes**

# Experimental and Predicted Data









## Bisphenol A




80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Property

 Summary 

 Download   Columns 

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

# Prediction models and transparency



Predicted

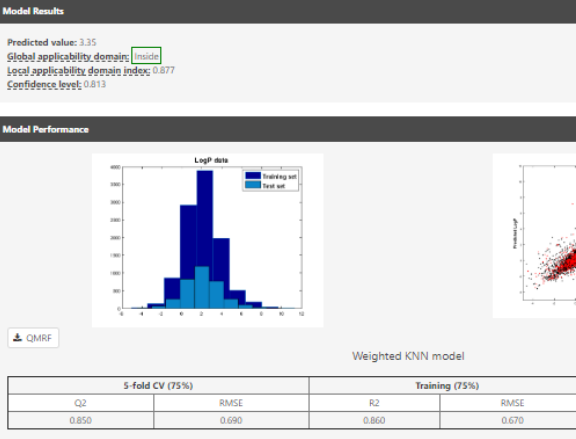
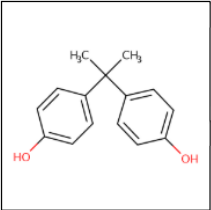
Download Predicted Data

Source	Result	Calculation Details	QMRF
<a href="#">EPISUITE</a>	3.64	Not Available	Not Available
<a href="#">NICEATM</a>	2.40	Not Available	<a href="#">Available</a>
<a href="#">ACD/Labs Consensus</a>	3.63	Not Available	Not Available
<a href="#">ACD/Labs</a>	3.43	Not Available	Not Available
<a href="#">OPERA</a>	3.35	<a href="#">OPERA Model Report [Inside AD]</a>	<a href="#">Available</a>

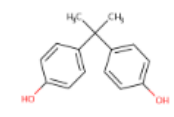
OPERA Models: LogP: Octanol-Water

Bisphenol A  
80-05-7 | DTXSID7020182

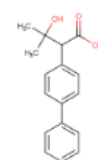
Print PDF



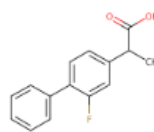
Nearest Neighbors from the Training Set



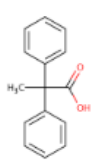
[Bisphenol A](#)  
Measured: 3.32  
Predicted: 3.35076



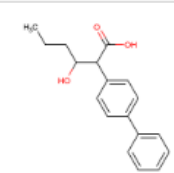
[BUTANOIC ACID 2-\(4-BIPHENYL\)-3-HYDROXY-3-METHYL-](#)  
Measured: 3.25  
Predicted: 3.39062



[Flurbiprofen](#)  
Measured: 4.16  
Predicted: 3.94445



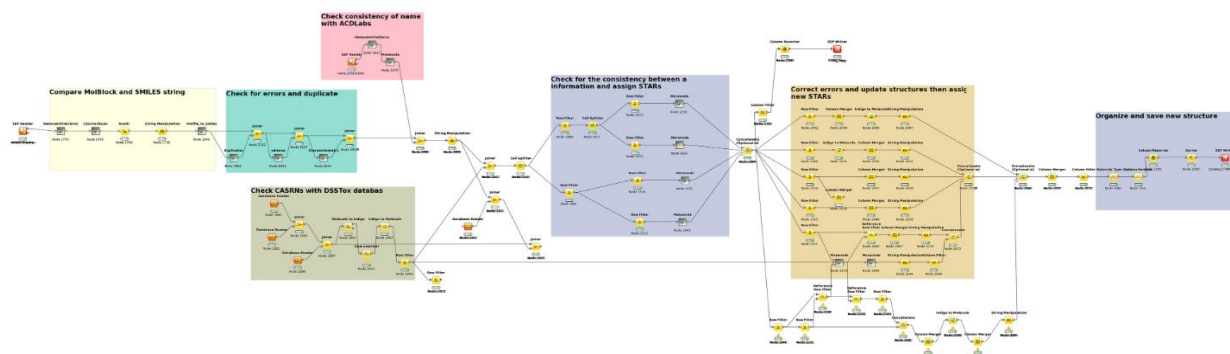
[2,2-Diphenylpropanoic acid](#)  
Measured: 2.69  
Predicted: 2.84603



[3-OH-2-\(4-BIPHENYL\)-L-HEXANOIC ACID](#)  
Measured: 3.75  
Predicted: 3.70322

To link to this article: <https://doi.org/10.1080/1062936X.2016.1253611>

Figure 8. The KNIME curation workflow developed based on the log *P* PHYSPROP dataset and generalized for application to other datasets.





# Public PHYSPROP Dataset

## Valence Errors

Mol Block	CAS	NAME	Smiles
	000276-67-3	TETRAZOLO[1,5-A]PYRIDINE	
	000442-85-8	ETHYL ISOTHIOCYANATE	
	000707-98-2	9-PROPYL ADENINE	
	000735-48-0	6-METHYL-4-NITROQUINOLINE-1-OXIDE	

## Mismatching structures

Mol Block	CAS	NAME	Smiles
	000076-43-7	PLUCKINESTERONE	
	000077-99-4	1,1,1-TRIS(HYDROXYMETHYL)PROPANE	
	000076-60-7	CORTISONE-6A-FLUORIDE	
	000082-38-2	DISPERSE RED 0	

## Duplicate Structures

Structure	Formula	PUB	CAS	NAME	MP	ExtMP	ErrMP
	C <sub>3</sub> H <sub>5</sub> O <sub>3</sub>	99-8779	000050-21-5	LACTIC ACID	1.6500000000000000e+001	2.2600000000000000e+001	5.9000000000000000e+000
	C <sub>3</sub> H <sub>5</sub> O <sub>3</sub>	99-8779	000079-31-4	L-LACTIC ACID	5.3000000000000000e+001	2.2600000000000000e+001	-3.0340000000000000e+001
	C <sub>3</sub> H <sub>5</sub> O <sub>3</sub>	99-8779	000048-82-3	2-HYDROXYPROPANOIC ACID	1.0000000000000000e+001	2.2600000000000000e+001	4.0000000000000000e+000
	C <sub>3</sub> H <sub>5</sub> O <sub>3</sub>	99-8779	010320-41-7	D-LACTIC ACID	5.2000000000000000e+001	2.2600000000000000e+001	-3.0340000000000000e+001

## Covalent Halogens

Mol Block	CAS	NAME	Smiles
	000056-93-9	BENZYL TRIMETHYL AMMONIUM CHLORIDE	
	000098-05-3	TETRAETHYL AMMONIUM IODIDE	
	000071-91-0	TETRAETHYL AMMONIUM BROMIDE	

# LogP dataset: 15,809 structures

- CAS Checksum: 12163 valid, 3646 invalid (**>23%**)
- Invalid names: 555
- Invalid SMILES 133
- Valence errors: 322 Molfile, 3782 SMILES (**>24%**)
- Duplicates check:
  - 31 DUPLICATE MOLFILES
  - 626 DUPLICATE SMILES
  - 531 DUPLICATE NAMES
- SMILES vs. Molfiles (structure check)
  - 1279 differ in stereochemistry (**~8%**)
  - 362 “Covalent Halogens”
  - 191 differ as tautomers
  - 436 are different compounds (**~3%**)

# OPERA Predicted Properties


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

 Journal of Cheminformatics

**RESEARCH ARTICLE** **Open Access**

 CrossMark

## 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>

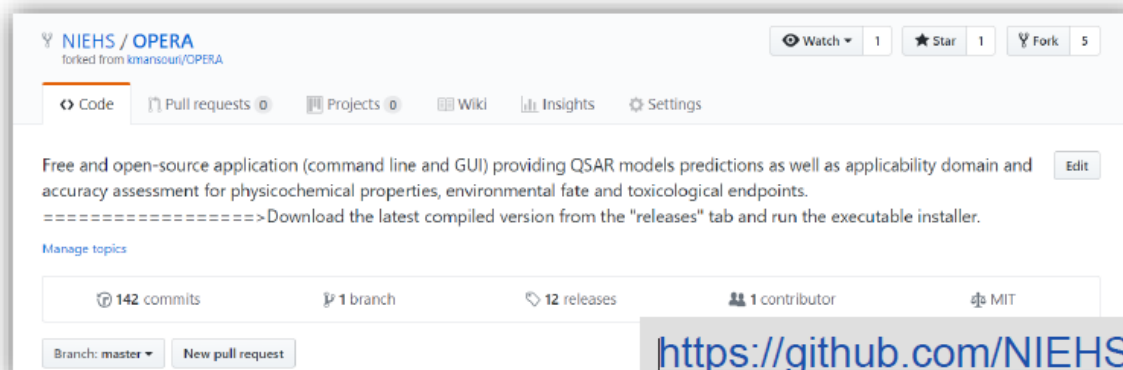
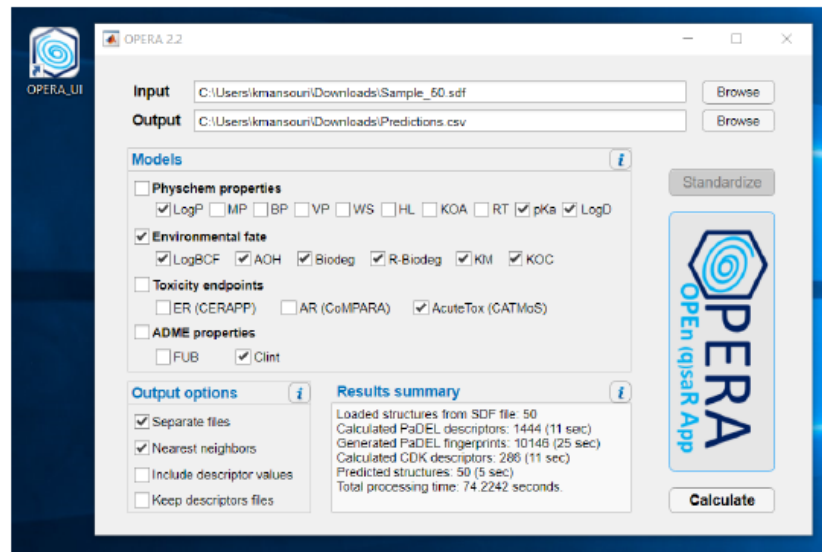
OPERA Models: <https://github.com/kmansouri/OPERA>

# OPERA Standalone Application

## Command line



## Graphical User Interface



<https://github.com/NIEHS/OPERA>

## README.md

# OPERA

OPERA is a free and open-source/open-data suite of QSAR models providing predictions on physicochemical properties, environmental fate and toxicity endpoints as well as additional information including applicability domain and accuracy assessment. All models were built on curated data and standardized QSAR-ready chemical structures. OPERA is available in command line and user-friendly graphical interface for Windows and Linux operating systems. It can be installed as a standalone desktop application or embedded in a different tool/workflow.

## References:

- [1] Mansouri K. et al. J Cheminform (2018) <https://doi.org/10.1186/s13321-018-0263-1>.
- [2] Mansouri, K. et al. SAR and QSAR in Env. Res. (2016). <https://doi.org/10.1080/1062936X.2016.1253611>
- [3] Williams A. J. et al. J Cheminform (2017) <https://doi.org/10.1186/s13321-017-0247-6>
- [4] The CompTox Chemistry Dashboard (<https://comptox.epa.gov/dashboard>)
- [5] JRC QSAR Model Database <https://qsardb.jrc.ec.europa.eu/qmrf/endpoint>

## Models:

- \* Latest version OPERA v2.2:
  - + Molecular descriptors:
    - PaDEL (2.21) (<https://doi.org/10.1002/jcc.21707> )
    - CDK (2.0) (<https://doi.org/10.1186/s13321-017-0220-4>)

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

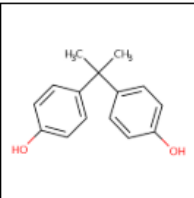
▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES



## Bisphenol A

80-05-7 | DTXSID7020182

Searched by Expert Validated Synonym.

Property

Summary

Download Columns

Property	Experimental average	Predicted average	Ex
Bioaccumulation Factor	-	173	
Bioconcentration Factor	133 (93)	93.5	15
Soil Adsorp. Coeff. (logKoc)	-	1.34e+3	
Atmos. Hydroxylation Rate	-	1.64e-11	
Biodeg. Half-Life	-	15.1	
Fish Biotrans. Half-Life (Km)	1.86 (1)	1.63	

# Environmental Fate and Transport

Source	Result	Experimental Details
ECOTOX: aquatic	150	Species: Navicula incerta; Response Site: Not reported
ECOTOX: aquatic	150	Species: Navicula incerta; Response Site: Not reported
ECOTOX: aquatic	100	Species: Navicula incerta; Response Site: Not reported
ECOTOX: aquatic	100	Species: Navicula incerta; Response Site: Not reported
PhysPropNCCT	43.7	
ECOTOX: aquatic	38.4	Species: Oncorhynchus mykiss; Response Site: Liver
ECOTOX: aquatic	25.0	Species: Oncorhynchus mykiss; Response Site: Liver
ECOTOX: aquatic	22.0	Species: Oncorhynchus mykiss; Response Site: Liver
ECOTOX: aquatic	10.8	Species: Oncorhynchus mykiss; Response Site: Liver
ECOTOX: aquatic	8.70	Species: Oncorhynchus mykiss; Response Site: Liver

## ECOTOX Knowledgebase

[Home](#)[Search](#)[Explore](#)[Help](#)[Contact Us](#)

Data last updated

June 13,  
2019

[See update totals](#)

Recent chemicals with full searches and coding completed

2-Phenylphenol

Amicarbazone

Fluazifop-p-butyl

Flutolanil

Per- and Polyfluoroalkyl Substances...

Total in database

11,722

Chemicals

12,775

Species

48,683

References

939,392

Results

**WELCOME TO ECOTOX VERSION 5!**

[Please click here to provide feedback so that we can continue to improve your experience.](#)



## Toxicity Estimation Software Tool (TEST)

On this page:

- [QSAR Methodologies](#)
- [What's New in Version 4.2.1?](#)
- [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](#).

# ***Real-Time Predictions Based on TEST***

# Real-Time Predictions

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Search: Atrazine

100%

Select properties to predict

**H** T.E.S.T.

**C**

**N**

**O**

**S**

**P**

**F**

**Cl**

**Br**

**I**

**PT**

- ☒ 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
  - ☒ Water solubility

Calculate

Chemical structure of Atrazine: CC1=NC2=C(N1)N=CN=C2N(C)CC(C)Cl


Chiral

# TEST Predictions

## Detailed calculation reports



Provider: T.E.S.T.

 Download Summary ▼

Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		4.951 -Log10(mol/L) 3.425 mg/L	5.198 -Log10(mol/L) 1.943 mg/L	5.257 -Log10(mol/L) 1.693 mg/L	5.287 -Log10(mol/L) 1.581 mg/L	4.064 -Log10(mol/L) 26.452 mg/L
48 hour D. magna LC50		4.430 -Log10(mol/L) 11.374 mg/L	4.764 -Log10(mol/L) 5.269 mg/L	5.006 -Log10(mol/L) 3.020 mg/L	4.430 -Log10(mol/L) 11.386 mg/L	3.521 -Log10(mol/L) 92.353 mg/L
48 hour T. pyriformis IGC50			5.272 -Log10(mol/L) 1.639 mg/L			
Oral rat LD50		1.989 -Log10(mol/kg) 3141.571 mg/kg	1.867 -Log10(mol/kg) 4157.591 mg/kg			2.111 -Log10(mol/kg) 2373.843 mg/kg
Bioaccumulation factor		1.321 Log10 20.956	1.209 Log10 16.192	1.585 Log10 38.452	1.517 Log10 32.923	0.974 Log10 9.409
Developmental toxicity		true	true	true		
Ames mutagenicity		false	false			false
Estrogen Receptor RBA						
Estrogen Receptor Binding		false	false	false	false	
Normal boiling point		357.4 °C	334.0 °C		432.8 °C	305.5 °C
Melting point		111.3 °C	98.3 °C		99.1 °C	136.7 °C
Flash point		219.9 °C	272.7 °C		211.4 °C	175.7 °C
Vapor pressure		-6.849 Log10(mmHg) 1.417*10 <sup>-7</sup> mmHg	-6.471 Log10(mmHg) 3.382*10 <sup>-7</sup> mmHg		-7.617 Log10(mmHg) 2.415*10 <sup>-8</sup> mmHg	-6.458 Log10(mmHg) 3.486*10 <sup>-7</sup> mmHg
Density		1.211 g/cm <sup>3</sup>	1.157 g/cm <sup>3</sup>		1.278 g/cm <sup>3</sup>	1.197 g/cm <sup>3</sup>

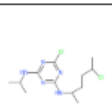
# TEST Predictions

## Detailed calculation reports

Predicted Vapor pressure at 25°C for C1C=1N=C(N=C(N1)NC(C)CCC(Cl)C)NC(C)C from Consensus method

Prediction results		
Endpoint	Experimental value	Predicted value
Vapor pressure at 25°C Log10(mmHg)	N/A	-6.85
Vapor pressure at 25°C mmHg	N/A	1.42E-07

Individual Predictions	
Method	Predicted value Log10(mmHg)
Hierarchical clustering	-6.47
Group contribution	-7.62
Nearest neighbor	-6.46

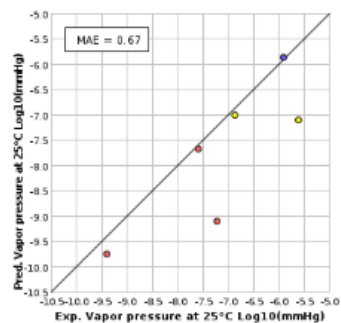


Predictions for the test chemical and for the most similar chemicals in the external test set

If the predicted value matches the experimental values for similar chemicals in the test set (and the similar chemicals are in the test set)

CAS	Structure	Similarity Coefficient	Experimental value Log10(mmHg)	Predicted value Log10(mmHg)
<chem>C1C=1N=C(N=C(N1)NC(C)CCC(Cl)C)NC(C)C</chem> (test chemical)			N/A	-6.85
<a href="#">7287-19-6</a>		0.83	-5.91	-5.86
<a href="#">130339-07-0</a>		0.77	-5.62	-7.11
<a href="#">21725-46-2</a>		0.76	-6.86	-7.01
<a href="#">120928-09-8</a>		0.58	-7.59	-7.67
<a href="#">101200-48-0</a>		0.56	-9.41	-9.76
<a href="#">119738-06-6</a>		0.55	-7.23	-9.11

Prediction results (colors defined in table below)



Chemicals	MAE*
Entire set	0.47
Similarity coefficient $\geq 0.5$	0.67

\*Mean absolute error in Log10(mmHg)

# Built on TEST Web Services

[https://www.epa.gov/sites/production/files/2018-08/documents/webtest\\_users\\_guide.pdf](https://www.epa.gov/sites/production/files/2018-08/documents/webtest_users_guide.pdf)



<https://comptox.epa.gov/dashboard/web-test/>

## User's Guide for WebTEST (version 1.0) (Web-services Toxicity Estimation Software Tool)

*A Web-Service  
from Mole*

5.	USING THE WEB SERVICE.....	43
5.1.	Downloading the Web-TEST project .....	43
5.2.	Importing the project into Eclipse .....	43
5.3.	Building WebTEST in Maven .....	43
5.4.	Starting a local server for WebTEST .....	43
5.5.	Running a client to perform batch calculations .....	44
5.6.	Run batch calculations without using a server .....	47
5.7.	“GET” API call.....	48
5.8.	“POST” API call.....	49
5.9.	Docker .....	50

- Given a set of chemicals how can data be harvested?
- OPERA and TEST predictions have been generated for structures in the database
- How can the dashboard be used to harvest data for hundreds to thousands of chemicals...

# Example list of chemicals - Opioids



## Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059



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

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

 [Show more](#)

<https://doi.org/10.1016/j.teac.2018.e00059>

[Get rights and content](#)



# Batch Search Names

Buprenorphine  
 Codeine  
 Dextromethorphan  
 Dihydrocodeine  
 Dihydromorphine  
 Ethylmorphine  
 Fentanyl  
 Heroin  
 Hydrocodone  
 Hydromorphone  
 Ketamine  
 Meperidine  
 Methadone  
 Morphine  
 Morphinone  
 Naloxone  
 Naltriben  
 Oxycodone  
 Oxymorphone  
 Propoxyphene  
 Sufentanil  
 Tramadol

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

Step Five: Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

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

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







Buprenorphine  
 Codeine  
 Dextromethorphan  
 Dihydrocodeine  
 Dihydromorphine  
 Ethylmorphine  
 Fentanyl  
 Heroin  
 Hydrocodone  
 Hydromorphone

Excel  
Download






INPUT	FOUND_BY	DTXSID
Buprenorphine	Approved Name	DTXSID2022705
Codeine	Approved Name	DTXSID2020341
Dextromethorphan	Approved Name	DTXSID3022908
Dihydrocodeine	Approved Name	DTXSID5022936
Dihydromorphine	Approved Name	DTXSID7048908
Ethylmorphine	Approved Name	DTXSID1046760
Fentanyl	Approved Name	DTXSID9023049
Heroin	Synonym	DTXSID6046761
Hydrocodone	Approved Name	DTXSID8023131
Hydromorphone	Approved Name	DTXSID8023133
Ketamine	Approved Name	DTXSID8023187
Meperidine	Approved Name	DTXSID9023253
Methadone	Approved Name	DTXSID7023273
Morphine	Approved Name	DTXSID9023336

# Include Other Data of Interest






## Chemical Identifiers

- ☒ DTXSID 
- ☒ Chemical Name 
- ☐ DTXCID 
- ☒ 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 

INPUT	DTXSID	CASRN	MOLECULAR_FORMULA	MONOISOTOPIC	MS_READY_SMILES
Buprenorphine	<a href="#">DTXSID202</a>	52485-79-7	C29H41NO4	467.3035588	[H]C12CC3=C4C
Codeine	<a href="#">DTXSID202</a>	76-57-3	C18H21NO3	299.1521435	[H]C12CC3=C4C
Dextromethamphetamine	<a href="#">DTXSID302</a>	125-71-3	C18H25NO	271.1936144	[H]C12CC3=C(C=
Dihydrocodone	<a href="#">DTXSID502</a>	125-28-0	C18H23NO3	301.1677936	[H]C12CC3=C4C
Dihydromorphine	<a href="#">DTXSID704</a>	509-60-4	C17H21NO3	287.1521435	[H]C12CC3=C4C
Ethylmorphine	<a href="#">DTXSID104</a>	76-58-4	C19H23NO3	313.1677936	[H]C12CC3=C4C
Fentanyl	<a href="#">DTXSID902</a>	437-38-7	C22H28N2O	336.2201635	CCC(=O)N(C1CC
Heroin	<a href="#">DTXSID604</a>	561-27-3	C21H23NO5	369.1576228	[H]C12CC3=C4C
Hydrocodone	<a href="#">DTXSID802</a>	125-29-1	C18H21NO3	299.1521435	[H]C12CC3=C4C
Hydromorphone	<a href="#">DTXSID802</a>	466-99-9	C17H19NO3	285.1364935	[H]C12CC3=C4C
Ketamine	<a href="#">DTXSID802</a>	6740-88-1	C13H16ClNO	237.0920418	CNC1(CCCCC1=
Meperidine	<a href="#">DTXSID902</a>	57-42-1	C15H21NO2	247.1572289	CCOC(=O)C1(CC
Methadone	<a href="#">DTXSID702</a>	76-99-3	C21H27NO	309.2092645	CCC(=O)C(CC(C)
Morphine	<a href="#">DTXSID902</a>	57-27-2	C17H19NO3	285.1364935	[H]C12CC3=C4C
Morphinone	<a href="#">DTXSID501</a>	467-02-7	C17H17NO3	283.1208434	[H]C12CC3=C4C
Naloxone	<a href="#">DTXSID802</a>	465-65-6	C19H21NO4	327.1470582	[H]C12CC3=C4C
Naltrexone	-	-	-	-	-
Oxycodone	<a href="#">DTXSID502</a>	76-42-6	C18H21NO4	315.1470582	[H]C12CC3=C4C
Oxymorphone	<a href="#">DTXSID502</a>	76-41-5	C17H19NO4	301.1314081	[H]C12CC3=C4C
Propoxyphene	<a href="#">DTXSID102</a>	469-62-5	C22H29NO2	339.2198292	CCC(=O)OC(CC1
Sufentanil	<a href="#">DTXSID602</a>	56030-54-7	C22H30N2O2S	386.2027994	CCC(=O)N(C1=C
Tramadol	<a href="#">DTXSID908</a>	27203-92-5	C16H25NO2	263.188529	COC1=CC=CC(=

# Batch search results - SDF

Database View Record Search Lists Plates Options ACD/Labs Help

Tile Table Default (One Record) Screen Form 2 Screen Form 1

a→a a→a a→a

CCCCC1=CC=C(C=C1)[C@H](C/C=C/C)O

INPUT: DTXSID6021327  
FOUND BY: DSSTox\_Substance\_Id  
DTXSID: DTXSID6021327  
PREFERRED\_NAME: Tetrahydrocannabinol  
BIOCONCENTRATION\_FACTOR\_TEST\_PRED: 295.6  
BOILING\_POINT\_DEGC\_TEST\_PRED: 372.841  
48HR\_DAPHNIA\_LC50\_MOL/L\_TEST\_PRED: 1.156  
DENSITY\_G/CM^3\_TEST\_PRED: 1.066  
DEVTOX\_TEST\_PRED: 0.863  
96HR\_FATHEAD\_MINNOW\_MOL/L\_TEST\_PRED: 6.2  
FLASH\_POINT\_DEGC\_TEST\_PRED: 181.634  
MELTING\_POINT\_DEGC\_TEST\_PRED: 101.426  
AMES\_MUTAGENICITY\_TEST\_PRED: 0.08  
ORAL\_RAT\_LD50\_MOL/KG\_TEST\_PRED: 0.00186  
SURFACE\_TENSION\_DYN/CM\_TEST\_PRED: 39.68

ID: 1 A: 1/39 B: 39 Last Upd

Database View Record Search Lists Plates Options ACD/Labs Help

Tile Table Default (One Record) Screen Form 2 Screen Form 1

1 (ID:1)	2 (ID:2)	3 (ID:3)	4 (ID:4)	5 (ID:5)	6 (ID:6)
7 (ID:7)	8 (ID:8)	9 (ID:9)	10 (ID:10)	11 (ID:11)	12 (ID:12)
13 (ID:13)	14 (ID:14)	15 (ID:15)	16 (ID:16)	17 (ID:17)	18 (ID:18)

ID: 1 A: 1/39 B: 39 Last Updated: 26/08/2019 02:16 Single DB

1-ChemSketch 2-Database 3-Processor

# ***Work in Progress***

# Prototype Development Structure/substructure search

## AADashboard

atrazine

Search



### Select properties to predict

T.E.S.T. 18

OPERA

Search

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

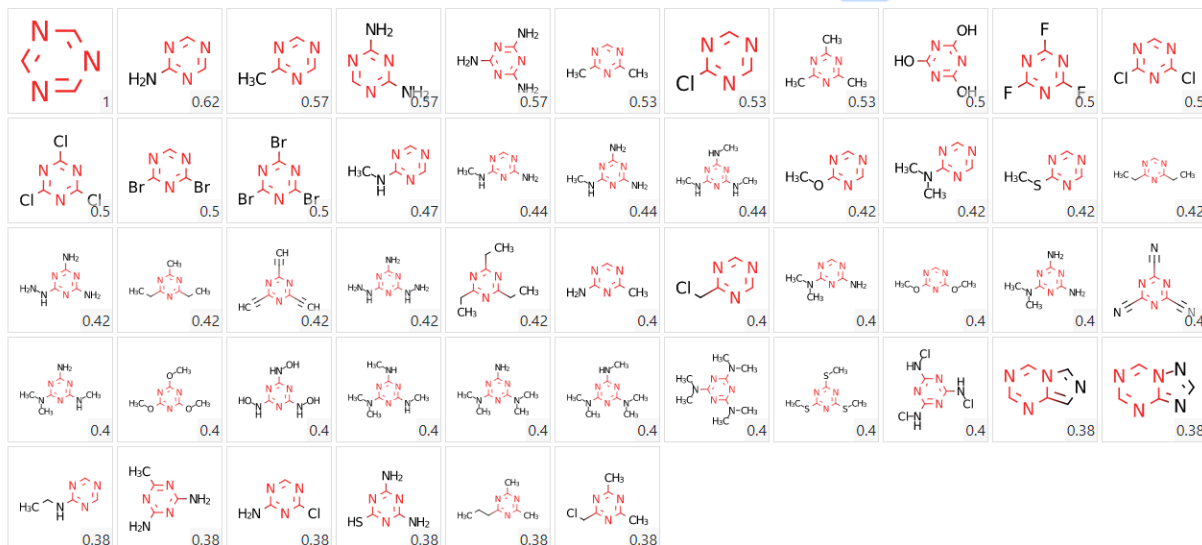
Filter by elements (enter comma separated list e.g. C,F,H) include

Filter by elements (enter comma separated list e.g. C,F,H) exclude

### Search result 2540

Show ☐ Isotopically Labeled ☐ Charged ☐ Salts or Mixtures

Sort Similarity



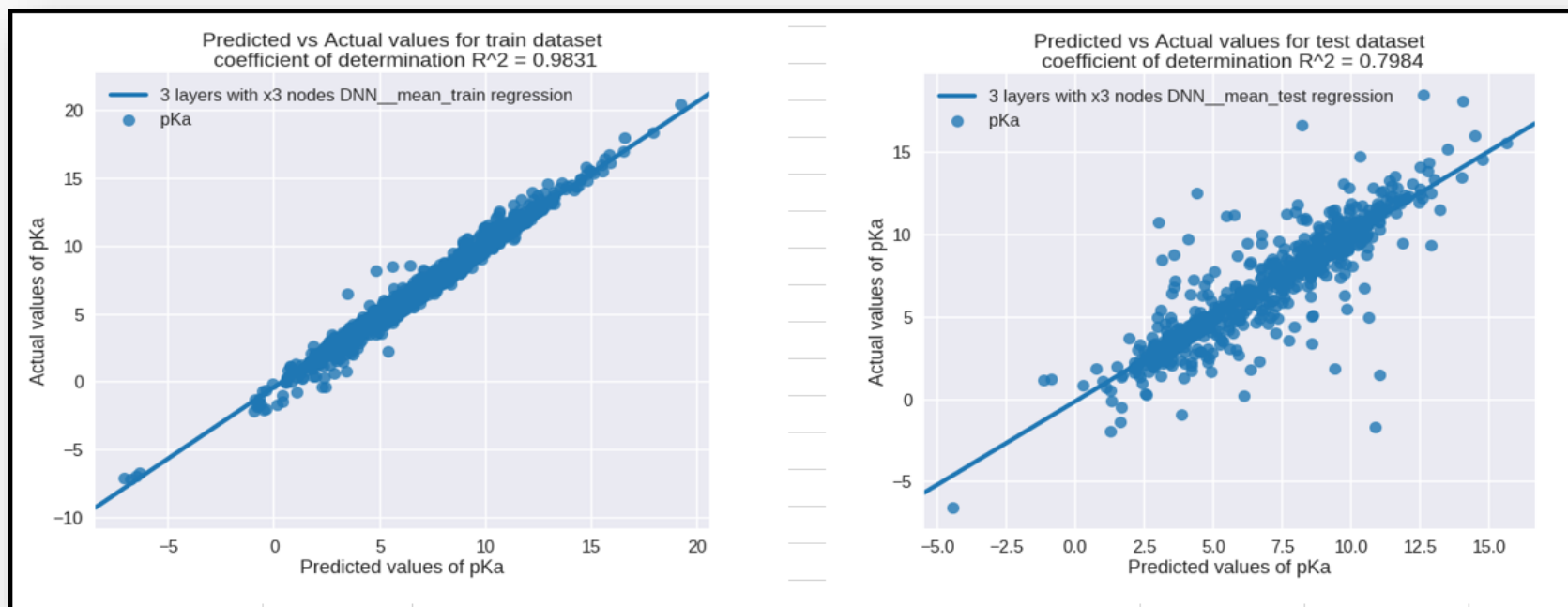
### Search result 2540

Show ☐ Isotopically Labeled ☐ Charged



# OPERA pKa Prediction Model

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



- *Accepted for publication to Journal of Cheminformatics*

# “Chemical Transformation Simulator”

- Chemical Transformation Simulator has public web services already available
  - (1) Abiotic Hydrolysis
  - (2) Abiotic Reduction
  - (3) Phase 1 Metabolism

**Mock-up**

United States Environmental Protection Agency

Share Search all data

Physical properties

- ☒ Normal boiling point
- ☒ Melting point
- ☒ Flash point
- ☒ Vapor pressure
- ☒ Density
- ☒ Surface tension
- ☒ Thermal conductivity
- ☒ Viscosity
- ☒ Water solubility

☒ Run CTS (Chemical Transformation Simulator)

- ☐ Abiotic Hydrolysis
- ☐ Abiotic Reduction
- ☒ Human Phase 1 Metabolism

Provider: [CTS](#) Transformation products (human phase 1 metabolism):

Triphenyl phosphate  
115-86-6

Diphenyl phosphate  
838-85-7

Phenol  
108-95-2

Provider: T.E.S.T.

Download Summary

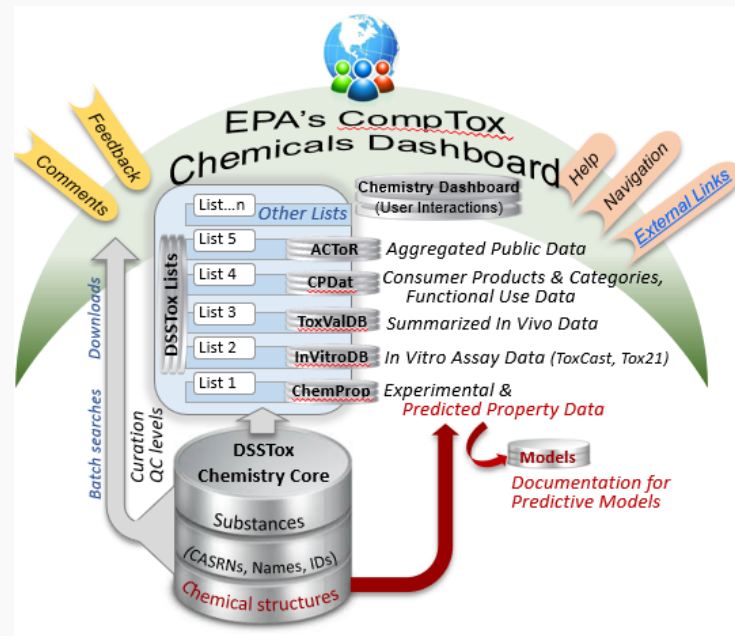
Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50	3.445 -Log10(mol/L) 28.039 mg/L	3.249 -Log10(mol/L) 44.040 mg/L	3.460 -Log10(mol/L) 27.100 mg/L	3.191 -Log10(mol/L) 50.282 mg/L	2.954 -Log10(mol/L) 86.948 mg/L	3.391 -Log10(mol/L) 31.751 mg/L
48 hour D. magna LC50	2.480 -Log10(mol/L) 20.000 mg/L	3.473 -Log10(mol/L) 86.000 mg/L	2.925 -Log10(mol/L) 23.000 mg/L	3.688 -Log10(mol/L) 11.000 mg/L	3.519 -Log10(mol/L) 33.000 mg/L	3.761 -Log10(mol/L) 40.000 mg/L

- Data are extracted from literature based on agency priorities
- Specific data sets of interest at present include those available for PFAS chemicals
- Data releases are every 6 months at present



# Conclusion

- Dashboard access to data for ~875,000 chemicals
- Ongoing aggregation of physicochemical property and environmental fate and transport data
- Retraining and rebuilding of models will occur as new data are assembled
- Web services already available for TEST predictions with services for OPERA next
- Future developments include integration of chemical transformation simulator



# Acknowledgements



Credit: the Research Triangle Foundation

## EPA-RTP

- *An enormous team of contributors from NCCT, especially the IT software development team and data curation team*
- *US-EPA ECOTOX for sharing data from their database*
- *Valery Tkachenko for development of TEST web services*

## Antony Williams

NCCT, US EPA Office of Research and Development,

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

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

Williams et al. *J Cheminform* (2017) 9:61  
DOI 10.1186/s13321-017-0247-6


 Journal of Cheminformatics

**DATABASE**

**Open Access**

### The CompTox Chemistry Dashboard: a community data resource for environmental chemistry



Antony J. Williams<sup>1\*</sup> , Christopher M. Grulke<sup>1</sup>, Jeff Edwards<sup>1</sup>, Andrew D. McEachran<sup>2</sup>, Kamel Mansouri<sup>1,2,4</sup>, Nancy C. Baker<sup>3</sup>, Grace Patlewicz<sup>1</sup>, Imran Shah<sup>1</sup>, John F. Wambaugh<sup>1</sup>, Richard S. Judson<sup>1</sup> and Ann M. Richard<sup>1</sup>

<https://doi.org/10.1186/s13321-017-0247-6>