

Using open data, open services, and open source software to deliver the EPA CompTox Chemicals Dashboard

***Antony Williams¹, Chris Grulke¹, Kamel Mansouri², Jeremy Dunne¹
and Jeff Edwards¹***

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

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

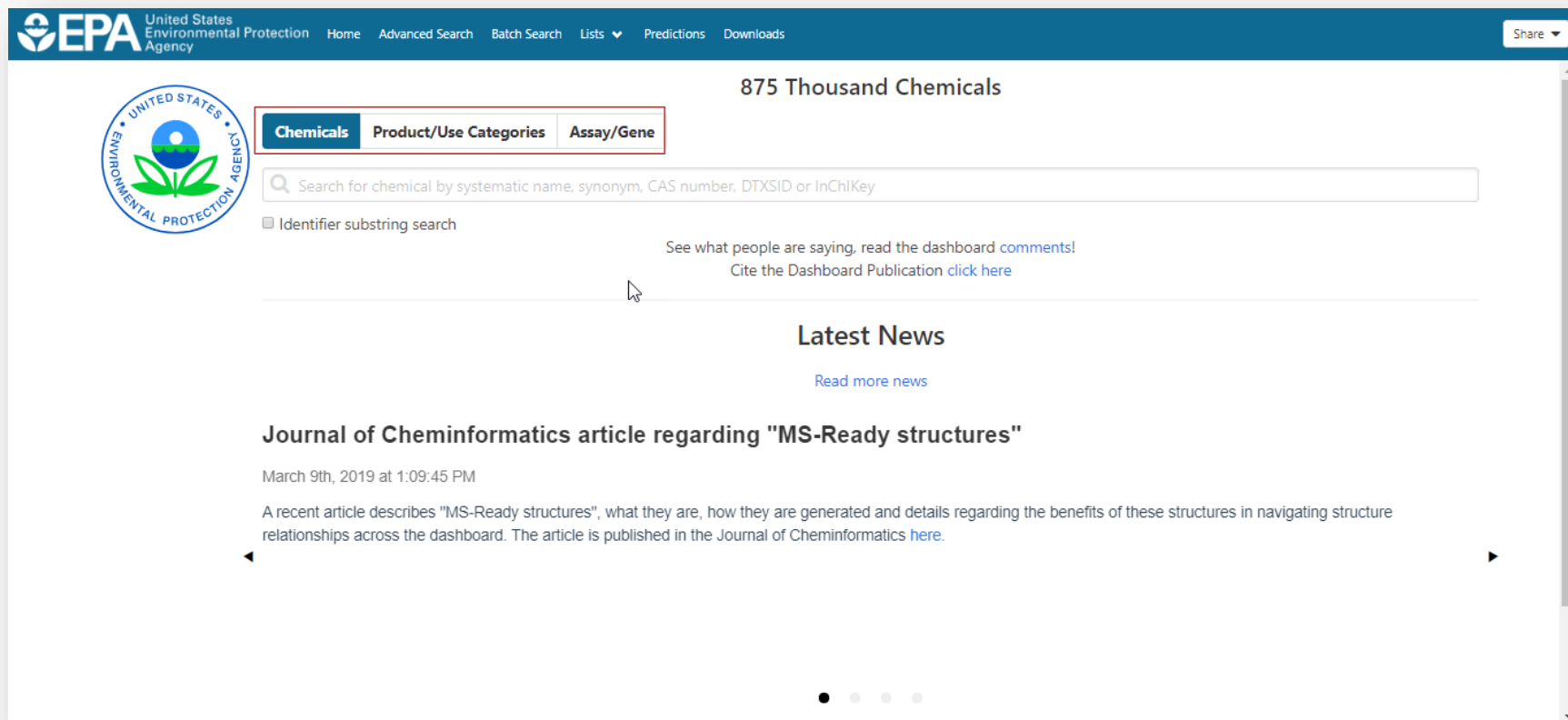
The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

*Fall 2019
ACS Fall Meeting, San Diego*

CompTox Chemicals Dashboard

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

875k Chemical Substances



The screenshot shows the EPA CompTox Chemicals Dashboard. At the top is a blue navigation bar with the EPA logo, the text "United States Environmental Protection Agency", and links for Home, Advanced Search, Batch Search, Lists, Predictions, and Downloads. A "Share" button is in the top right. Below the navigation bar is a white header area with the EPA seal on the left and the text "875 Thousand Chemicals" on the right. Under the seal are three tabs: "Chemicals" (selected), "Product/Use Categories", and "Assay/Gene". Below the tabs 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 for "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 first news item is titled "Journal of Cheminformatics article regarding 'MS-Ready structures'" and is dated "March 9th, 2019 at 1:09:45 PM". The text of the article states: "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 dashboard are four small circular indicators.

United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share

875 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](#)

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

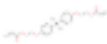
Chemicals **Product/Use Categories** **Assay/Gene**

 Bisphenol

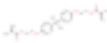


Bisphenol A
DTXSID7020182

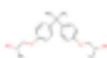




Bisphenol A bis(2-hydroxyethyl ether) diacrylate
DTXSID6066991

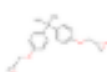


Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate
DTXSID1066992

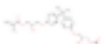


Bisphenol A bis(2-hydroxypropyl) ether
DTXSID8051592

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

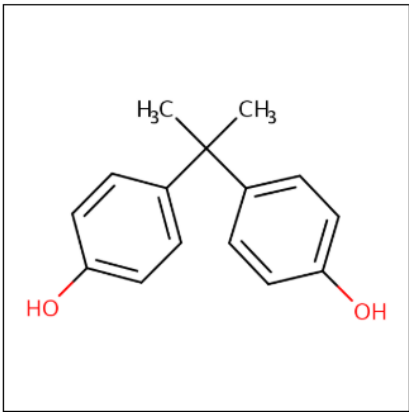
Copy Share Submit Comment Search all data



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

- Total data landscape includes:
 - ~875,000 chemical substances
 - Experimental & predicted physchem property data
 - Experimental Human and Ecological hazard data
 - Bioactivity data for 1000s of chemicals
 - Consumer products containing chemicals
 - “Literature” searches for chemicals using PubMed
 - Real time prediction of physchem/toxicity endpoints

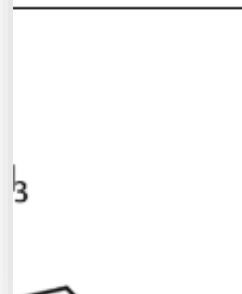
- To make this happen we **CONSUME** open data and open source software
- We also **PRODUCE** open data and both free and open source software
- This presentation is an overview of what we consume and what we produce...

CONSUMER

Integrated Wikipedia Snippet

XSID7020182

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](#)

- Integrated Wikipedia snippet linked out to full article

PhysChem Data and Predictions

CONSUMER: Available Data 2010 files underlying EPI Suite

→ ↻ ⓘ Not secure esc.syrres.com/interkow/EpiSuiteData_ISIS_SDF.htm

Apps Travel Voucher | OR... ChemReg_v0.9.2a Altmetric it! Science Data Expert... Jo

EPI Suite Data - ISIS/Base & SDF

The downloaded files are provided in "zip" format ... the downloaded file must be "un-zipped" with common utility programs such as [WinZip](#).

... **Updated September 15, 2010**

Basic Instructions:

- (1) Download the zip file
- (2) Un-Zip the file

NOTE ... zipped files extract to **Folders** containing the individual data files ... Folders named **EPI_ISIS_Data** and **EPI_SDF_Data**

Substructure Searching Files:

ISISTM/Base & SD Files of the EPI Suite Program Experimental Data Files are now available ... The ISISTM/Base files require the commercial program for use ... The SD Files can be imported into other commercial chemical structure programs (such as ChemFinder).

... [Click here to download EPI_ISIS_Data.zip](#) ... (about 11 MB)

... [Click here to download EPI_SDF_Data.zip](#) ... (about 10 MB)

NOTE ... EPI Suite Data Files (some in Excel, Text, Word format) available at:

<http://esc.syrres.com/interkow/EpiSuiteData.htm>

Data required thorough curation

Valence Errors

Mol Block	CAS	NAME	Smiles
	000274-87-3	TETRAZOLO(1,5-A)PYRIDINE	
	000342-85-8	ETHYL ISOTHIOCYANATE	
	000707-98-2	9-PROPYL ADENINE	
	000735-48-0	6-METHYL-4-NITROQUINOLINE-1-OXIDE	

Mismatching structures

Mol Block	CAS	NAME	Smiles
	000075-43-7	FLUCYMESTERONE	
	000077-99-6	1,1,1-TRIS(2-HYDROXYETHYL)PROPANE	
	000079-60-7	CORTISONE-6A-FLUORIDE	
	000082-38-2	DISPERSE RED 1	

Duplicate Structures

Structure	Formula	PW	CAS	NAME	BP	TaBP	ErrorBP
	C ₃ H ₅ O ₃	99.8779	000050-21-5	LACTIC ACID	1.6900000000000000e+001	2.2800000000000000e+001	5.9600000000000000e+000
	C ₃ H ₅ O ₃	99.8779	000079-31-4	L-LACTIC ACID	5.2600000000000000e+001	2.2800000000000000e+001	-3.8240000000000000e+001
	C ₃ H ₅ O ₃	99.8779	000040-82-3	2-HYDROXYPROANOIC ACID	1.6900000000000000e+001	2.2800000000000000e+001	4.9600000000000000e+000
	C ₃ H ₅ O ₃	99.8779	010330-41-7	D-LACTIC ACID	5.2600000000000000e+001	2.2800000000000000e+001	-3.8240000000000000e+001

Covalent Halogens

Mol Block	CAS	NAME	Smiles
	000056-93-9	BENZYL TRIMETHYL AMMONIUM CHLORIDE	
	000088-05-3	TETRAETHYL AMMONIUM IODIDE	
	000071-91-0	TETRAETHYL AMMONIUM BROMIDE	

CONSUMER KNIME workflows for curation

1,363

Views

32

CrossRef citations
to date

15

Altmetric

Listen
Articles

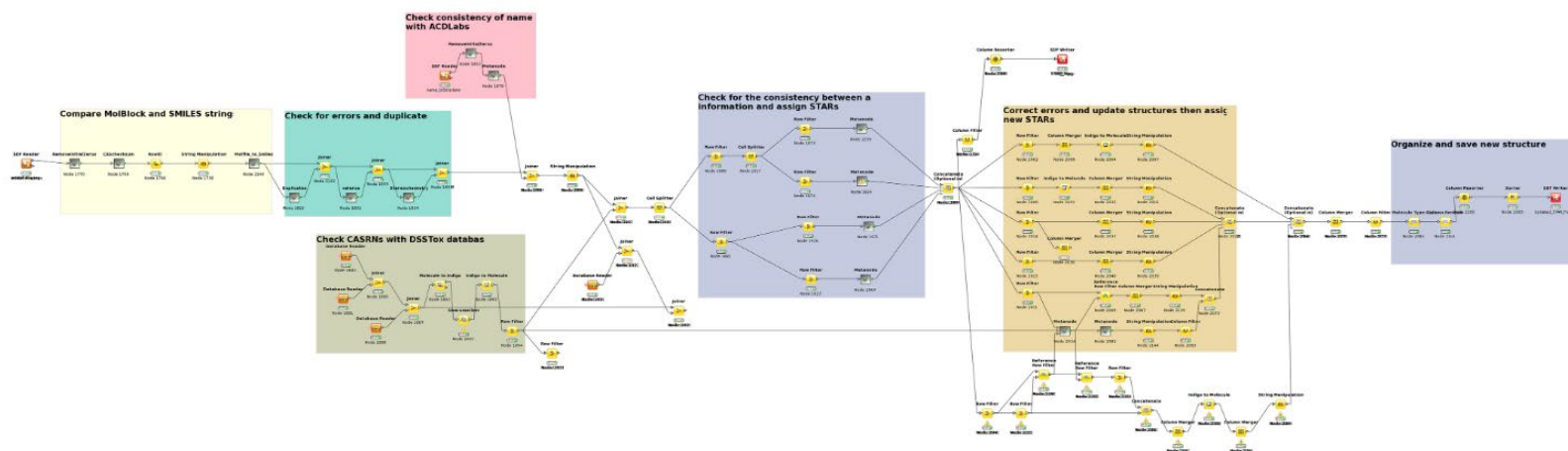
An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modelling^{\$}

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

Pages 911-937 | Received 03 Sep 2016, Accepted 24 Oct 2016, Published online: 25 Nov 2016

Figure 8 of 14

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



PRODUCER

All curated data available

[Cite](#)[Share](#)[Embed](#)[+ Collect \(you need to log in first\)](#)

Chemistry Dashboard Data: Physprop Analysis

Version 2  Dataset posted on 10.05.2018, 14:18 by EPA's National Center for Computational Toxicology

The data associated with the publication “An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modeling” represents the curated data associated with the OPERA models used to predicted properties for the CompTox Chemistry Data. The data include the training and test data sets as well as the KNIME workflows used to perform the curation of the data. For a full understanding of the data and workflows we recommend accessing the publication also.

TIMELINE



First online date
09.11.2017



Posted date
10.05.2018

335
views

3
downloads

0
citations



CATEGORIES

• Toxicology

KEYWORD(S)

PHYSPROP

NCCT

Computational Toxicology

Chemistry Dashboard

LICENCE



CC0

PRODUCER TEST and OPERA Predictions

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

COMMENTS

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

 Journal of Cheminformatics

RESEARCH ARTICLE

Open Access



OPERA models for predicting physicochemical properties and environmental fate endpoints

Kamel Mansouri^{1,2,3*}, Chris M. Grulke¹, Richard S. Judson¹ and Antony J. Williams¹

Predicted

 Download Predicted Data ▼

Source	Result	Calculation Details	QMRF
EPISUITE	9.93e-4	Not Available	Not Available
NICEATM	2.16e-4	Not Available	Available
TEST	4.41e-4	TEST Report	Not Available
OPERA	1.42e-4	OPERA Model Report [Inside AD]	Available

Transparency for prediction models



Predicted

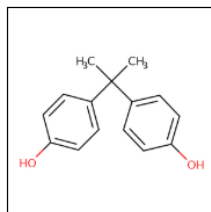
Download Predicted Data ▼

Source	Result	Calculation Details	QMRF
EPISUITE	3.64	Not Available	Not Available
NICEATM	2.40	Not Available	Available
ACD/Labs Consensus	3.63	Not Available	Not Available
ACD/Labs	3.43	Not Available	Not Available
OPERA	3.35	OPERA Model Report [Inside AD]	Available

OPERA Models: LogP: Octanol-Water

Bisphenol A
80-05-7 | DTXSID7020182

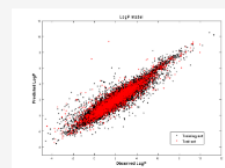
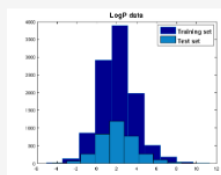
Print PDF



Model Results

Predicted value: 3.35
Global applicability domain: [Inside](#)
Local applicability domain index: 0.877
Confidence level: 0.813

Model Performance



QMRF

Weighted KNN model

5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.850	0.690	0.860	0.670	0.860	0.780

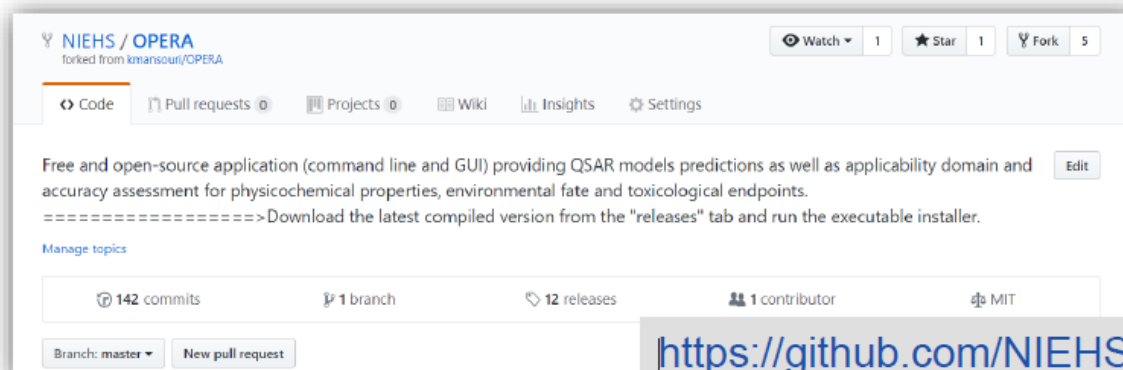
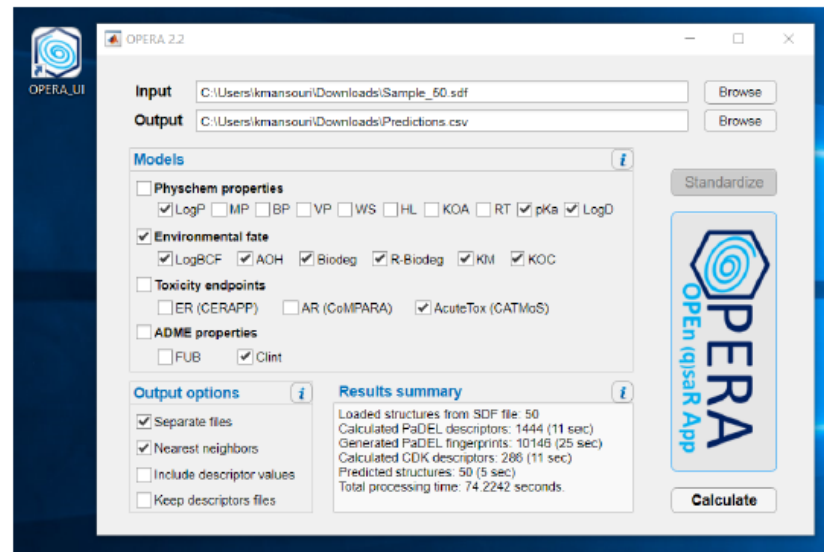
PRODUCER

OPERA Standalone Application

Command line



Graphical User Interface



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

PRODUCER: Open Source

<https://github.com/kmansouri/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>)

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](#).

Bioactivity Data: Tox21 and Toxcast

PRODUCER

In Vitro Bioassay Screening



Bisphenol A
80-05-7 | DTXSID7020182
Searched by Expert Validated Synonym.



ToxCast/Tox21

QC Data ID	Grade	Description
Tox21_202992	Pass	Purity>90% and MW confirmed
Tox21_400088	Pass	Purity>90% and MW confirmed

Assay Selection 136 Selected

☒ Active ☐ Inactive ☐ All

A Single Assay Can Have Multiple Charts

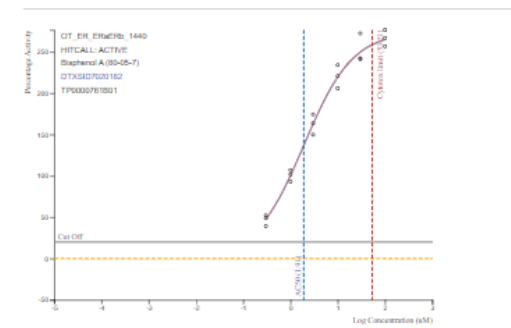
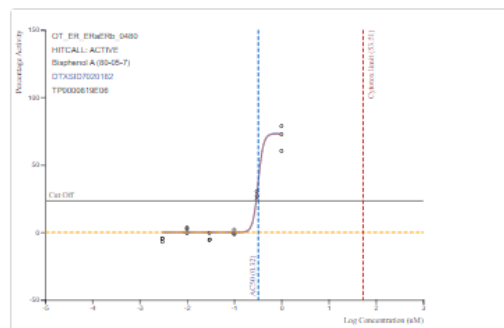
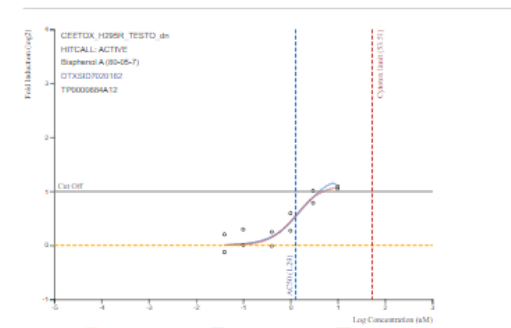
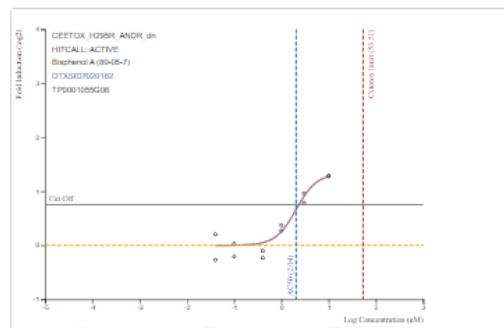
☒ Representative Samples Only

Bioactivity Summary

Number of Charts: 136

Filter assays

- Ceetox/OpAns (2 of 24 selected)
- Odyssey Thera (6 of 17 selected)
- Attagene (4 of 165 selected)
- Tox21/NCGC (44 of 211 selected)
- CellDirect (3 of 48 selected)
- Bioseek (4 of 174 selected)
- Apradica (8 of 107 selected)
- NHEERL Padilla Lab (1 of 1 selected)
- Novascreen (46 of 167 selected)
- NHEERL's Hunter Lab (0 of 4 selected)
- NCCT's Lab (4 of 4 selected)
- ACEA Biosciences (4 of 6 selected)
- Tanguay Lab (9 of 19 selected)
- NHEERL Stoker & Laws Lab (1 of 2 selected)



PRODUCER

Bioactivity: Downloadable Data

Exploring ToxCast Data: Downloadable Data

The results after processing through the Pipeline are available on the [ToxCast Dashboard](#), and for most users EPA recommends accessing the data there.

- [ToxCast Chemicals](#)
- [ToxCast Assays](#)

ToxCast Data and Information

- **ToxCast & Tox21 Summary Files.** Data for a single chemical endpoint pair for thousands of chemicals and assay endpoints for 20 variables such as the activity or hit call, activity concentrations, whether the chemical was tested in a specific assay, etc.
 - [Download ToxCast Summary Information](#)
 - [Download ReadMe](#)
- **ToxCast & Tox21 Data Spreadsheet.** A spreadsheet of EPA's analysis of the chemicals screened through ToxCast and the Tox21 collaboration which includes EPA's activity calls from the screening of over 1,800 chemicals.
 - [Download Data](#)
 - [Download ReadMe](#)
- **ToxCast Data Pipeline R Package.** The R computer programming package used to process and model all EPA ToxCast and Tox21 chemical screening data. The files include the R programming package as well as documents that provide overviews of the data analysis pipeline used and the R package. Users will need experience with R to use these files.
 - [Download Package](#)
 - [TCPL Overview](#)

Resources

- [Toxicity Forecaster \(ToxCast\) Fact Sheet](#)
- [ToxCast Publications](#)
- [ToxCast Citation](#)
- [About ToxCast](#)

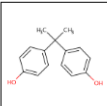
CONSUMER

PubChem Widgets - Bioactivities

United States
Environmental Protection
Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

Copy▼Share▼Submit CommentSearch all data



Bisphenol A

80-05-7 | DTXSID7020182

Searched by Expert Validated Synonym.

PubChem Biological Activities

[PUBCHEM](#) > [BISPHENOL A](#) > [BIOASSAY RESULTS](#)

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

TOXCAST: SUMMARY

EDSP21

TOXCAST/TOX21

PUBCHEM

TOXCAST: MODELS

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

BioAssay Results

2,261 items View More Rows & Details

Download

SORT BY Activity Value						
Activity	Activity Value, μM	Activity Type	Target Name	BioAssay Name	BioAssay AID	Substance SID
Inconclusive	0.0014	Potency		qHTS assay to identify small molecule agonists of the endoplasmic reticulum stress response signaling pathway - cell viability counter screen	1159517	144214049
Active	0.0055	Kd	Chain A, Crystal Structure Of Human Estrogen-Related Receptor Gamma Ligand Binding Domain Complex With Bisphenol A (human)	Experimentally measured binding affinity data (Kd) for protein-ligand complexes derived from PDB	977611	87557090
Active	0.0055	Kd	ESRRG - estrogen related receptor gamma (human)	Binding affinity to human ERR gamma	1121409	103308477
Inconclusive	0.0126	Potency	THRB - thyroid hormone receptor beta (human)	qHTS assay for small molecule antagonists of thyroid hormone receptor beta signaling	588547	26752849
Inconclusive	0.1364	Potency		qHTS assay to identify small molecule antagonists of the peroxisome proliferator-activated receptor delta (PPARd) signaling pathway - cell viability counter screen	743213	144210190

1 2 3 ... 453 Next >

Literature Data

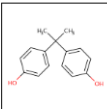
CONSUMER

PubChem Widgets - Articles

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 Expert Validated Synonym.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

LITERATURE

GOOGLE SCHOLAR

PUBMED ABSTRACT SIFTER

PUBCHEM ARTICLES

PubChem Articles

PUBCHEM > BISPHENOL A > DEPOSITOR PROVIDED PUBMED CITATIONS

Depositor Provided PubMed Citations

2,754 items View More Rows & Details Download

SORT BY Publication Date

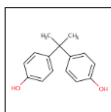
PMID	Publication Date	Title	Journal
30099086	2018-12-01	Effects of bisphenol analogs on thyroid endocrine system and possible interaction with 17 β -estradiol using GH3 cells.	Toxicology in vitro : an international journal published in association with BIBRA
30228064	2018-12-01	Mutual promotion of apoptosis and autophagy in prepubertal rat testes induced by joint exposure of bisphenol A and nonylphenol.	Environmental pollution (Barking, Essex : 1987)
30248606	2018-12-01	Dose-dependent transcriptomic responses of zebrafish eleutheroembryos to Bisphenol A.	Environmental pollution (Barking, Essex : 1987)
30075455	2018-11-15	Maternal Bisphenol A exposure impaired endochondral ossification in craniofacial cartilage of rare minnow (Gobiocypris rarus) offspring.	Ecotoxicology and environmental safety
30196065	2018-11-15	Cellular, transcriptomic and methylome effects of individual and combined exposure to BPA, BPF, BPS on mouse spermatocyte GC-2 cell line.	Toxicology and applied pharmacology

1 2 3 ... 551 Next >

from PubChem

CONSUMER

PubChem Widgets – Patents



Bisphenol A

80-05-7 | DTXSID7020182

Searched by Expert Validated Synonym.

PubChem Patents

[PUBCHEM](#) > [BISPHENOL A](#) > [DEPOSITOR-SUPPLIED PATENT IDENTIFIERS](#)

Depositor-Supplied Patent Identifiers



75,385 items [View More Rows & Details](#)

[Download](#)

SORT BY Relevance			
Patent ID	Title	Submitted Date	Granted Date
US2019169189	TRIAZA-SPIRODECANONES AS DDR1 INHIBITORS	2019-02-04	
US2019105237	HIGH REFRACTIVE INDEX ADDITION-FRAGMENTATION AGENTS	2018-12-10	
US2019029972	SULFUR(VI) FLUORIDE COMPOUNDS AND METHODS FOR THE PREPARATION THEREOF	2018-10-12	
US2019015301	Dental Materials Based On Low-Viscosity Radically Polymerizable Monomers With A High Refractive Index	2018-07-11	
US2019010172	Polymer Materials With Silane-Based Transfer Reagents	2018-07-02	
1 2 3 ... 15,077 Next >			

[from PubChem](#)

[Link to all deposited patent identifiers](#)

[from PubChem](#)

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

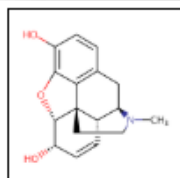
▼ LITERATURE

GOOGLE SCHOLAR

PUBMED ABSTRACT SIFTER

PUBCHEM ARTICLES

CONSUMER: Pubmed Services Literature Searching



Morphine

57-27-2 | DTXSID9023336

Searched by Approved Name.

Abstract Sifter

1) Select PubMed starting point query then 2) click on Retrieve. 

Select a Query Term



Retrieve Articles

Select a Query Term

- Hazard
- Fate and Transport
- Metabolism/PK/PD
- Chemical Properties
- Exposure
- Mixtures
- Male Reproduction
- Androgen Disruption
- Female Reproduction
- GeneTox
- Cancer
- Clinical Trials
- Embryo and embryonic development
- Child (infant through adolescent)
- Dust and Exposure
- Food and Exposure
- Water and Exposure
- Algae
- Disaster / Emergency

Optionally, edit the query before retrieving.

"57-27-2" OR "Morphine"

CONSUMER: Pubmed Services

Literature Searching

Child (infant through adolescent)

Dust and Exposure

Food and Exposure

Water and Exposure

Algae

Disaster / Emergency



Optionally, edit the query before retrieving.

("57-27-2" OR "Morphine") AND ((water OR groundwater OR drinking water) AND Environmental Exposure)



CONSUMER: Pubmed Services Literature Searching

37 of 37 articles loaded...

To find articles quickly, enter terms to sift abstracts. 

☐	wastewater	Spectrometry ↓	EPA	Total	PMID	Year	Title	Authors	Journal	Rev
☐	4	2	0	6	29274731	2017	Simultaneous analysis of opioid analgesics and thei...	Krizman-Matic; Kostanjevecki; Ahel; Terzic	Journal of chromatography. A	
☐	0	1	0	1	25768972	2015	Evaluating external contamination of polybrominate...	Poon; Aleksa; Carnevale; Kapur; Goodyer; Koren	Therapeutic drug monitoring	
☐	0	1	0	1	22544551	2012	Spatial distribution of illicit drugs in surface waters o...	Vazquez-Roig; Andreu; Blasco; Morillas; Picó	Environmental science and pollution research inter...	
☐	1	1	0	2	20801487	2010	Analysis of illicit and illicit drugs in waste, surface an...	Berset; Brenneisen; Mathieu	Chemosphere	
☐	1	1	0	2	17935751	2007	Illicit drugs, a novel group of environmental contami...	Zuccato; Castiglioni; Bagnati; Chiabrando; Grassi; ...	Water research	
☐	2	1	1	4	17607391	2007	Using environmental analytical data to estimate lev...	Bones; Thomas; Paull	Journal of environmental monitoring : JEM	
☐	3	1	2	6	17180984	2006	Simultaneous determination of psychoactive drugs ...	Hummel; Löffler; Fink; Ternes	Environmental science & technology	
☐	6	0	0	6	30583189	2018	Assessment of drugs of abuse in a wastewater trea...	Kumar; Tschärke; O'Brien; Mueller; Wilkins; Padhye	The Science of the total environment	
☐	0	0	3	3	30488421	2018	Effect of enriched environment during adolescence ...	Mohammadian; Najafi; Miladi-Gorji	Developmental psychobiology	
☐	3	0	0	3	29574368	2018	Estimation of the consumption of illicit drugs during ...	Foppe; Hammond-Weinberger; Subedi	The Science of the total environment	
☐	1	0	0	1	28787791	2017	Evaluation of in-sewer transformation of selected illi...	Gao; Banks; Li; Jiang; Lai; Mueller; Thai	The Science of the total environment	
☐	9	0	0	9	28472697	2017	Occurrence and fate of illicit drugs and pharmaceuti...	Causanilles; Ruepert; Ibáñez; Emke; Hernández; d...	The Science of the total environment	
☐	0	0	0	0	28010888	2016	Dose-dependent effects of morphine on lipopolysac...	Mottaz; Schönenberger; Fischer; Eggen; Schirmer; ...	Environmental pollution (Barking, Essex : 1987)	
☐	0	0	0	0	27746311	2016	Effects of voluntary exercise on the viability, prolifer...	Haydari; Safari; Zarbakhsh; Bandegi; Miladi-Gorji	Neuroscience letters	
☐	0	0	0	0	27261879	2016	Genotoxic effects induced by the exposure to an en...	Parolini; Magni; Castiglioni; Binelli	Ecotoxicology and environmental safety	
☐	3	0	0	3	27179320	2016	Temporal trends in drug use in Adelaide, South Aus...	Tschärke; Chen; Gerber; White	The Science of the total environment	

PRODUCER

Abstract Sifter for Excel


Open for Science

Search

BROWSE GATEWAYS HOW TO PUBLISH ▾ ABOUT ▾ BLOG

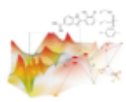


SOFTWARE TOOL ARTICLE

Abstract Sifter: a comprehensive front-end system to PubMed [version 1; referees: 2 approved]

 Nancy Baker ¹, Thomas Knudsen², Antony Williams ²

 Author details

 This article is included in the [Chemical Information Science](#) gateway.

Abstract

The Abstract Sifter is a Microsoft Excel based application that enhances existing search capabilities of PubMed. The Abstract Sifter assists researchers to search effectively, triage results, and keep track of articles of interest. The tool implements an innovative "sifter" functionality for relevance ranking, giving the researcher a way to find articles of interest quickly. The tool also gives


METRICS

629

 VIEWS

118

 DOWNLOADS

 Get PDF

 Get XML

 Cite

 Export

 Track

 Email

Real-Time Predictions

<epam> | LIFE SCIENCES OPEN SOURCE | PRODUCTS | RESOURCES | DOWNLOADS | CONTACT INFO

Home / Ketcher

Overview

Features

Supported browsers

Third-party code

Ketcher 2.0

Download

License

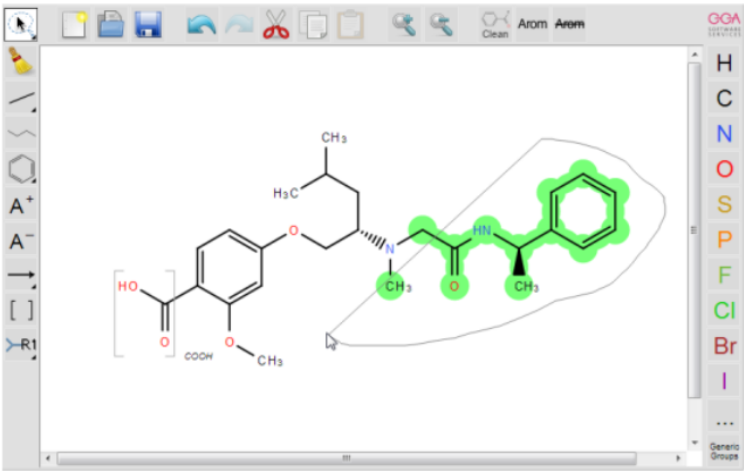
Feedback

Commercial Availability

Ketcher

Overview

Ketcher is a web-based chemical structure editor.



Since Ketcher is written in pure Javascript, it incorporates high performance, good portability and light weight. You will not need any Java or Flash plugins to use it in your browser. Ketcher is completely free and open-source, while also available on a commercial basis.

CONSUMER: Ketcher Drawing

TEST Real Time Predictions

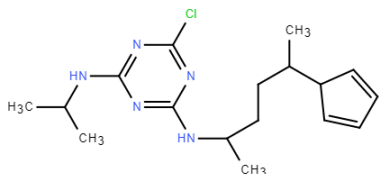
 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Predictions

Atrazine



100%

Select properties to predict

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
 - ☒ Water solubility

Calculate

Discover. [About/Disclaimer](#)

Connect. [ACToR](#)

Ask. [Contact](#)

Ketcher Drawing TEST Real Time Predictions

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

Provider: T.E.S.T.

Download Summary

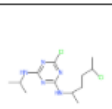
Property	Experimental Value	Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		5.665 -Log10(mol/L) 0.727 mg/L	5.978 -Log10(mol/L) 0.353 mg/L	6.256 -Log10(mol/L) 0.186 mg/L	6.362 -Log10(mol/L) 0.146 mg/L	4.064 -Log10(mol/L) 29.012 mg/L
48 hour D. magna LC50		5.210 -Log10(mol/L) 2.073 mg/L	5.134 -Log10(mol/L) 2.468 mg/L	5.356 -Log10(mol/L) 1.480 mg/L	5.504 -Log10(mol/L) 1.052 mg/L	4.844 -Log10(mol/L) 4.807 mg/L
48 hour T. pyriformis IGC50		5.342 -Log10(mol/L) 1.530 mg/L	5.737 -Log10(mol/L) 0.616 mg/L			4.947 -Log10(mol/L) 3.798 mg/L
Oral rat LD50		1.944 -Log10(mol/kg) 3819.151 mg/kg	2.077 -Log10(mol/kg) 2814.178 mg/kg			1.812 -Log10(mol/kg) 5183.011 mg/kg
Bioaccumulation factor		1.547 Log10 35.264	1.221 Log10 16.650	1.753 Log10 56.677	2.709 Log10 511.865	0.505 Log10 3.201
Developmental toxicity		true	true	true		
Ames mutagenicity		false	false			false
Estrogen Receptor RBA						
Estrogen Receptor Binding		false	false	false		
Normal boiling point		373.6 °C	315.5 °C		500.0 °C	305.5 °C
Melting point		125.9 °C	114.5 °C		126.6 °C	136.7 °C
Flash point		250.3 °C	291.2 °C		239.7 °C	220.1 °C
Vapor pressure		-7.296 Log10(mmHg) 5.061*10 ⁻⁸ mmHg	-6.837 Log10(mmHg) 1.456*10 ⁻⁹ mmHg		-8.873 Log10(mmHg) 1.34*10 ⁻⁹ mmHg	-6.178 Log10(mmHg) 6.643*10 ⁻⁷ mmHg
Density		1.177 g/cm ³	1.123 g/cm ³		1.197 g/cm ³	1.212 g/cm ³
Surface tension						
Thermal conductivity						
Viscosity						
Water solubility		4.683 -Log10(mol/L) 6.964 mg/L	4.512 -Log10(mol/L) 10.328 mg/L		4.885 -Log10(mol/L) 4.378 mg/L	4.653 -Log10(mol/L) 7.469 mg/L

TEST 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

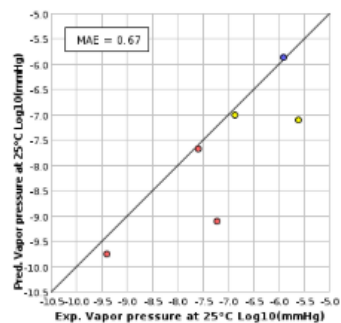


Predictions for the test chemical and for the most similar chemicals in the external test set

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
7287-19-6		0.83	-5.91	-5.86
130339-07-0		0.77	-5.62	-7.11
21725-46-2		0.76	-6.86	-7.01
120928-09-8		0.58	-7.59	-7.67
101200-48-0		0.56	-9.41	-9.76
119738-06-6		0.55	-7.23	-9.11

Prediction results (colors defined in table below)



Chemicals	MAE*
Entire set	0.47
Similarity coefficient ≥ 0.5	0.67

*Mean absolute error in Log10(mmHg)

PRODUCER: TEST Software

<https://www.epa.gov/chemical-research/toxicity-estimation-software-tool-test>



Environmental Topics

Laws & Regulations

About EPA

Search EPA.gov



Related Topics: [Safer Chemicals Research](#)

CONTACT US

SHARE



Toxicity Estimation Software Tool (TEST)

On this page:

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

PRODUCER: TEST Web Services

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

PRODUCER

Web Services

- Dozens of web services to provide access to data
- Data in UI, JSON and XML format

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=80-05-7>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=80-05-7>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=80-05-7>

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=DTXCID60513>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=DTXCID60513>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=DTXCID60513>

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

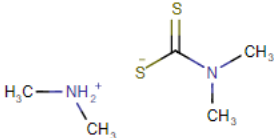
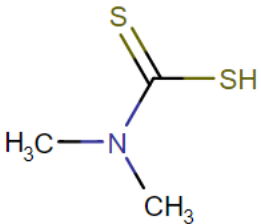
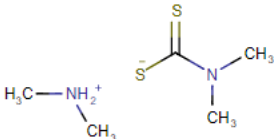
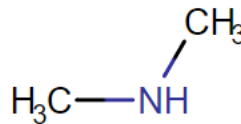
<https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

<https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

PRODUCER

InChIKey to DTXCIDs

<https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=UVOFGKIRTCCNKG-UHFFFAOYSA-N>

Image	DTXCID	Smiles	Image	MsReady DTXCID	MsReady SMILES
	DTXCID60513	<chem>C[NH2+]C.CN(C)C([S-])=S</chem>		DTXCID0023797	<chem>CN(C)C(S)=S</chem>
	DTXCID60513	<chem>C[NH2+]C.CN(C)C([S-])=S</chem>		DTXCID704057	<chem>CNC</chem>

CONSUMER of our services

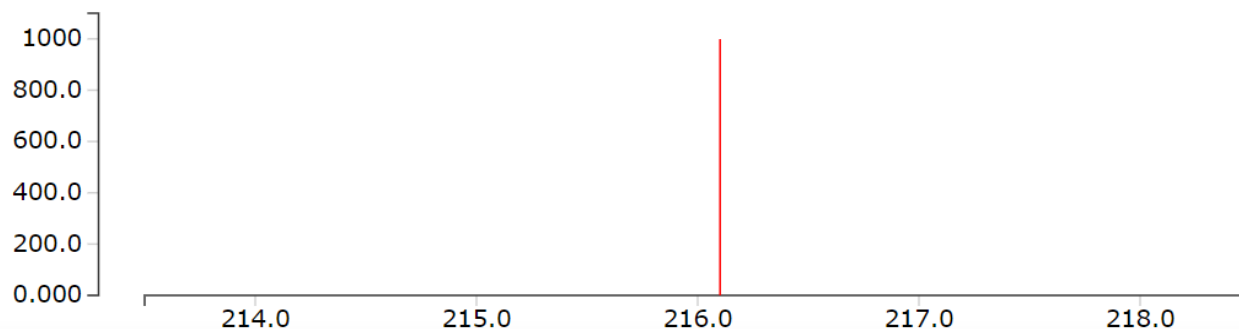
MassBank mapping to Dashboard

MassBank Record: EA028808

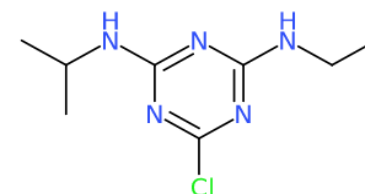
[Home](#) | [Search](#) | [Record Index](#) | [Data Privacy](#) | [Imprint](#) | MassBank ID:

Atrazine; LC-ESI-ITFT; MS2; CE: 15%; R=15000; [M+H]⁺

Mass Spectrum



Chemical Structure



Options
● Labels

CH\$NAME: Atrazine
CH\$NAME: 6-chloro-N-ethyl-N'-isopropyl-1,3,5-triazine-2,4-diamine
CH\$NAME: 6-chloranyl-N4-ethyl-N2-propan-2-yl-1,3,5-triazine-2,4-diamine
CH\$COMPOUND_CLASS: N/A; Environmental Standard
CH\$FORMULA: [C8H14ClN5](#)
CH\$EXACT_MASS: 215.0932
CH\$SMILES: c1(nc(nc(n1)Cl)NCC)NC(C)C
CH\$IUPAC: InChI=1S/C8H14ClN5/c1-4-10-7-12-6(9)13-8(14-7)11-5(2)3/h5H,4H2,1-3H3,(H2,10,11,12,13,14)
CH\$LINK: CAS [1912-24-9](#)
CH\$LINK: CHEBI [15930](#)
CH\$LINK: KEGG [C06551](#)
CH\$LINK: PUBCHEM [CID:2256](#)
CH\$LINK: INCHIKEY [MXWJVTOOROXGIU-UHFFFAOYSA-N](#)
CH\$LINK: CHEMSPIDER [2162](#)
CH\$LINK: COMPTOX [DTXSID9020112](#)

Open Data Sharing

CONSUMER NORMAN Suspect List Exchange



Wastewater Suspect List based on Swedish Product Data	Wastewater Suspect List Original File with Mapped DTXSIDs (12/02/2019)	KEMIWSUS InChIKeys (12/02/2019)	A prioritized list of 1,123 substances relevant for wastewater based on Swedish product registry data, including scores. Provided by Stellan Fischer, KEMI.
Algal toxins list from CompTox	ALGALTOX XLSX, CSV (14/02/2019) CompTox ALGALTOX List	ALGALTOX InChIKeys (14/02/2019)	List of algal toxins (generated during blooms) from the CompTox Chemicals Dashboard.
CCL 4 Chemical Candidate List	CCL4 XLSX, CSV (14/02/2019) CompTox CCL4 List	CCL4 InChIKeys (14/02/2019)	Contaminants that are not (yet) regulated in the USA but are known or anticipated to occur in public water systems; from CompTox.
Hydrogen Deuterium Exchange (HDX) Standard Set	HDXNOEX XLSX, CSV (14/02/2019) CompTox HDXNOEX List CompTox HDXEXCH List	HDXNOEX InChIKeys (14/02/2019)	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule HRMS (Ruttkies et al. submitted). HDXEXCH list also contains observed deuterated species.
Neurotoxicants Collection from Public Resources	NEUROTOXINS XLSX, CSV (14/02/2019) CompTox NEUROTOXINS List	NEUROTOXINS InChIKeys (14/02/2019)	A list of neurotoxicants compiled from public resources, details on CompTox and Schymanski <i>et al.</i> (submitted).
Statins Collection from Public Resources	STATINS XLSX, CSV (14/02/2019) CompTox STATINS List	STATINS InChIKeys (14/02/2019)	A list of statins (lipid-lowering medications) compiled from public resources, details on CompTox.
Synthetic Cannabinoids and Psychoactive Compounds	SYNTHCANNAB XLSX, CSV (14/02/2019) CompTox SYNTHCANNAB List	SYNTHCANNAB InChIKeys (14/02/2019)	A list of synthetic cannabinoids and psychoactive compounds assembled from public resources, from CompTox.

PRODUCER

Curated Chemical Lists

Home Advanced Search Batch Search Lists ▼ Predictions Downloads

Lists of Chemicals

List of Assays

Select List

Download ▼

Columns ▼

PFAS

 Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	PFAS[EPA: List of 75 Test Samples (Set 1)]	2018-06-29	74	PFAS list corresponds to 75 samples (Set 1) submitted for initial testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFAS75S2	PFAS[EPA: List of 75 Test Samples (Set 2)]	2019-02-21	75	PFAS list corresponds to a second set of 75 samples (Set 2) submitted for testing screens conducted by EPA researchers in collaboration with researchers at the National Toxicology Program.
EPAPFASCAT	PFAS[EPA Structure-based Categories]	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	PFAS[EPA: Chemical Inventory Insoluble in DMSO]	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	PFAS[EPA: ToxCast Chemical Inventory]	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	PFAS[EPA: Cross-Agency Research List]	2017-11-16	199	EPAPFASRL is a manually curated listing of mainly straight-chain and branched PFAS (Per- & Poly-fluorinated alkyl substances) compiled from various internal, literature and public sources by EPA researchers and program office representatives.
PFASKEMI	PFAS: List from the Swedish Chemicals Agency (KEMI) Report	2017-02-09	2416	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	PFAS Master List of PFAS Substances	2018-07-26	5061	PFASMASTER is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.
PFASOECD	PFAS: Listed in OECD Global Database	2018-05-16	4729	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing more than 4700 new PFAS
PFASTRIER	PFAS Community-Compiled List (Trier et al, 2015)	2017-07-16	597	PFASTRIER community-compiled public listing of PFAS (Trier et al, 2015)

United States
Environmental Protection
Agency

☐ Identifier substring search

1



PRODUCER

Publishing Open Data: CPDat

SCIENTIFIC DATA

Data Descriptor | [OPEN ACCESS](#) | Published: 10 July 2018

The Chemical and Products Database, a resource for exposure-relevant data on chemicals in consumer products

Kathie L. Dionisio, Katherine Phillips, Paul S. Price, Christopher M. Grulke, Antony Williams, Derya Biryol, Tao Hong & Kristin K. Isaacs 

Scientific Data **5**, Article number: 180125 (2018) | [Download Citation](#) 

Data Citations

1. Williams, A. Figshare
<http://dx.doi.org/10.23645/epacomptox.5352997> (2017)

...and then reused in PubChem

PubChem Atrazine (Compound)

10.4 Uses



EPA CPDat Chemical and Product Categories

20 items [View More](#)

[Download](#)

SORT BY Category		
Category	Category Description	Categorization Type
Agrochemical, crop	Products used on crops, or related to the growing of crops	CPCat Cassette
Agrochemical, crop, flowers	Related to the growing of flowers as crops	CPCat Cassette
Agrochemical, crop, fruit	Related to fruit crops, or the processing or preserving of fruit	CPCat Cassette
Agrochemical, crop, wheat	Wheat crops	CPCat Cassette
Drinking_water_contaminant, pesticide, detected	Chemicals detected in substances or products (note that these chemicals may be absent from an 'ingredient list' for the product and thus unexpected, but have been detected in product testing studies)	CPCat Cassette
1 2 3 4 Next >		

► from EPA Chemical and Products Database (CPDat)

PRODUCER

Downloadable Data

Downloads

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

Posted: 11/14/2016

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

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

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

Posted: 11/14/2016

The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID), Preferred Name, DTXCID, Standard InChI String and Standard InChIKey (UPDATED APRIL 2019).

[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. (UPDATED APRIL 2019)

[DSSTox SDF File](#)

Posted: 12/14/2016

This zip file contains the entire chemical structure collection of over 850,000 chemicals from the DSSTox database contained in one large SDF file. The file contains the structure, The DSSTox Structure Identifier (DTXSID), The DSSTOX Substance Identifier (DTXSID listed as PubChem External Data Source), the associated Dashboard URL, associated synonyms and Quality Control Level details. In order to view an SDF file you will need to have access to the appropriate piece of software to open an SDF files. Examples include ChemAxon JChem, ACD/ChemFolder or ChemDraw. (UPDATED APRIL 2019)

[PHYSPROP Analysis File](#)

Posted: 12/14/2016

The data associated with the publication "[An automated curation procedure for addressing chemical errors and inconsistencies in public datasets used in QSAR modeling](#)" represents the curated data associated with the OPERA models used to predicted properties for the CompTox Chemistry Data. The data include the training and test data sets as well as the KNIME workflows used to perform the curation of the data. For a full understanding of the data and workflows we recommend accessing the

- Multiple projects in progress that use Open Source software
 - Structure, substructure and similarity searching
 - MS-Ready and QSAR-Ready data preparation
 - WebTEST “Batch” predictions
 - OPERA model predictions
 - Prediction of MS fragmentation patterns and matching to searched experimental spectra
- New services and Full API will be openly available (in future releases)

CONSUMER: Prototype Development epam Ketcher + Bingo NoSQL

AADashboard

atrazine Search

100%

Select properties to predict

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

C

N ☐ Exact

O ☒ Substructure

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

<td><td><td><td></td></td></td></td>	<td><td><td></td></td></td>	<td><td></td></td>	<td></td>	
<td><td><td><td></td></td></td></td>	<td><td><td></td></td></td>	<td><td></td></td>	<td></td>	
<td><td><td><td></td></td></td></td>	<td><td><td></td></td></td>	<td><td></td></td>	<td></td>	
<td><td><td><td></td></td></td></td>	<td><td><td></td></td></td>	<td><td></td></td>	<td></td>	
<td><td><td><td></td></td></td></td>	<td><td><td></td></td></td>	<td><td></td></td>	<td></td>	

Search result **2540** Show ☐ Isotopically Labeled

CONSUMER

Epam Bingo NoSQL

<epam>

LIFE SCIENCES OPEN SOURCE

PRODUCTS

RESOURCES

DOWNLOADS

CONTACT INFO

Search

Overview >

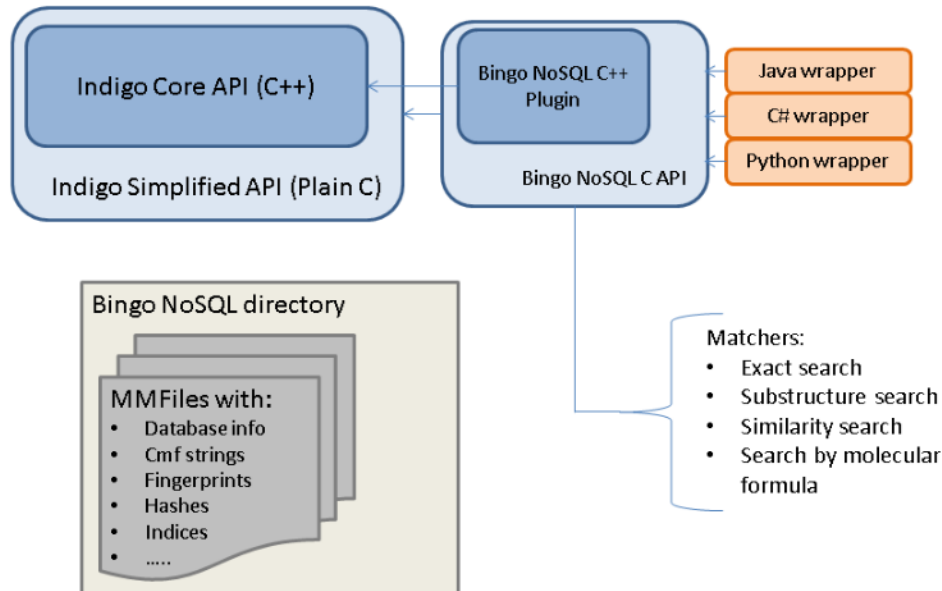
Manuals

Bingo NoSQL

Overview

Bingo NoSQL is a Indigo plugin and non-relational database management system for storing chemical information and searching through it. With this plugin you can create databases which will be located on the hard drive of your local machine or some remote server. Bingo NoSQL uses only own and OS functionality for creating and accessing the databases, so there is no need in installing any additional third-party software. For storing chemical structures and other extra information memory-mapped files technology was used. This technology shows better performance than using direct read and write operations, so I/O delays have no significant effect on the Bingo NoSQL speed.

Structure



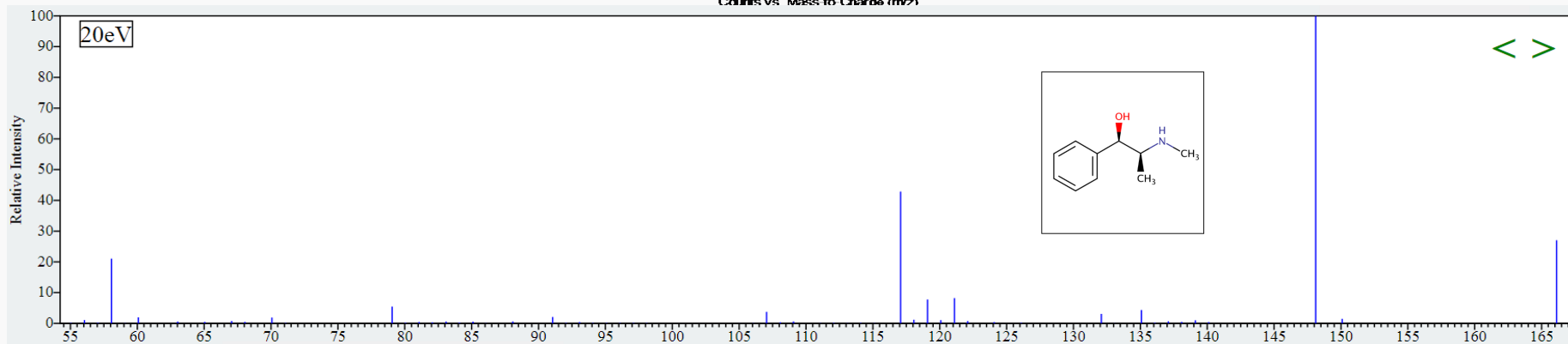
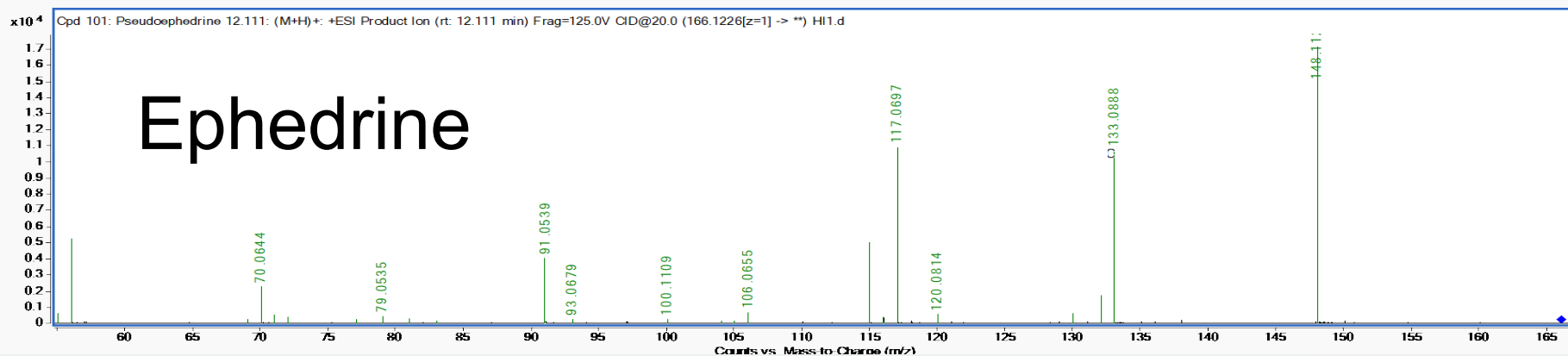
CONSUMER

CFM-ID Fragmentation Prediction

CFM-ID

Competitive Fragmentation Modeling for Metabolite Identification

Ephedrine



PRODUCER

CFM-ID Predicted Library

- Predictions generated and stored for >700,000 structures
- Python code to score experimental vs predicted spectra
- Cosine dot product match score calculation

Data Descriptor | [OPEN](#) | Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

Andrew D. McEachran , Ilya Balabin, Tommy Cathey, Thomas R. Transue, Hussein Al-Ghoul, Chris Grulke, Jon R. Sobus & Antony J. Williams 

Scientific Data **6**, Article number: 141 (2019) | [Download Citation](#) 

PRODUCER

Published data is on FigShare

Data Descriptor | [OPEN](#) | Published: 02 August 2019

Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns

Andrew D. McEachran[✉], Ilya Balabin, Tomm Grulke, Jon R. Sobus & Antony J. Williams[✉]

Scientific Data **6**, Article number: 141 (2019)

CFM-ID Paper Data

Dataset posted on 01.03.2019, 08:38 by EPA's National Center for Computational Toxicology

This upload is a zip containing the following files:

Predicted EI-MS Spectra of CompTox Chemicals Dashboard Structures:

Predicted EI-MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (<https://doi.org/10.1021/acs.analchem.6b01622>). These data are provided in .dat ASCII format.

Predicted MS/MS Spectra in ESI-positive mode of CompTox Chemicals Dashboard Structures:

Predicted MS/MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (<https://doi.org/10.1007/s11306-014-0676-4>) in ESI-positive mode. These data are provided in .dat ASCII format.

Predicted MS/MS Spectra in ESI-negative mode of CompTox Chemicals Dashboard Structures:

Predicted MS/MS spectra of ~700,000 chemical structures from the CompTox Chemicals Dashboard were generated using the CFM-ID model developed by Allen, et al. (<https://doi.org/10.1007/s11306-014-0676-4>) in ESI-negative mode. These data are provided in .dat ASCII format.

88
views

17
downloads

0
citations



CATEGORIES

• Toxicology

KEYWORD(S)

Computational Toxicology

DSSTox Chemical Database

Chemicals Dashboard

Non-targeted analysis

CFM-ID

LICENCE



CC0

EXPORT

RefWorks

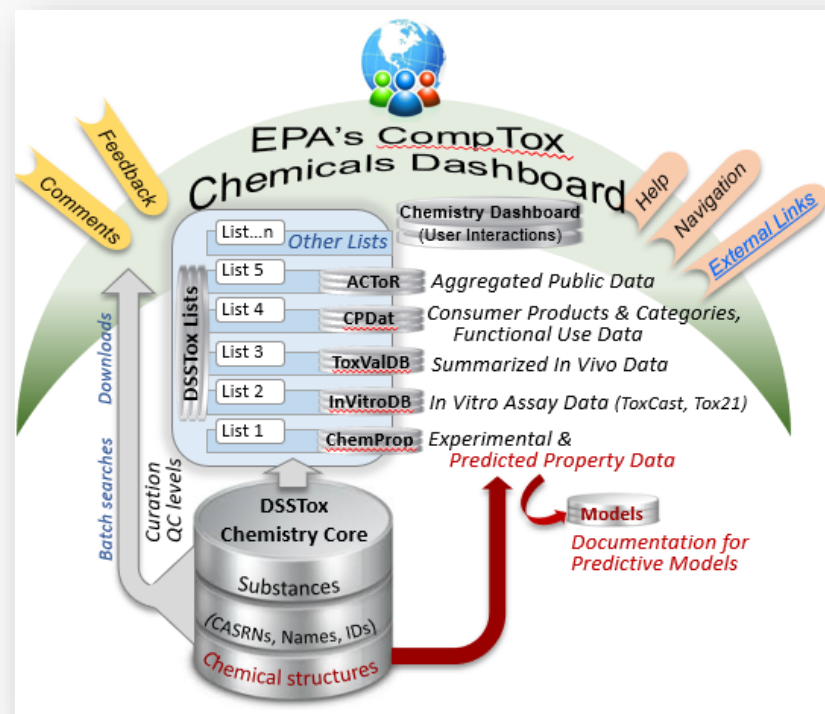
BibTeX

Ref. manager

Endnote

Conclusion

- Dashboard access to data for ~875,000 chemicals
- Dashboard **CONSUMES** a lot of open source libraries and open data
- We **PRODUCE** open models and data to the community in exchange
- We are committed to an open API to provide more complete data access and real time predictions



Acknowledgements



- *NCCT IT development team*
- *Tommy Cathey, ACTOR Web Services*
- *Nancy Baker, Abstract Sifter*
- *Todd Martin & Valery Tkachenko, WebTEST*
- *Kathie Dionisio & Kristin Isaacs, CPDat*
- *Thanks to Emma Schymanski, University of Luxembourg, for coordinating all efforts with the NORMAN Network for curation of lists on the Suspect Exchange*

Antony Williams

NCCT, US EPA Office of Research and Development,

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^{1*} , Christopher M. Grulke¹, Jeff Edwards¹, Andrew D. McEachran², Kamel Mansouri^{1,2,4}, Nancy C. Baker³, Grace Patlewicz¹, Imran Shah¹, John F. Wambaugh¹, Richard S. Judson¹ and Ann M. Richard¹

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