

# Introduction to Cheminformatics: Accessing data through the CompTox Dashboard

**Antony John Williams**

**[williams.antony@epa.gov](mailto:williams.antony@epa.gov)**

*Center for Computational Toxicology and Exposure, US-EPA, RTP, NC*

UNC Chapel Hill: September 20<sup>th</sup> 2021

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

# Who am I?

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



- Computational chemist at the US-EPA – **scientist**
- Responsibility for cheminformatics projects, internal & external collaborations, “product marketing” – **cheminformatician**
- Work with a team of people developing software solutions – “**product & project manager**”
- Scientific publications, books, blogger – **author**; I am @ChemConnector – **social networker**



## Antony Williams

### ORCID iD

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

Print view

### Also known as

ChemConnector, ChemSpiderman, Tony Williams

### Country

United States

### Keywords

NMR, Computer-assisted Structure Elucidation, Chemistry, ChemSpider, Cheminformatics, Open PHACTS, PharmaSea, Open Science, Computational Toxicology

### Other IDs

Scopus Author ID: 55258539900  
US EPA VIVO: [williams.antony](#)  
ResearcherID: C-3089-2009

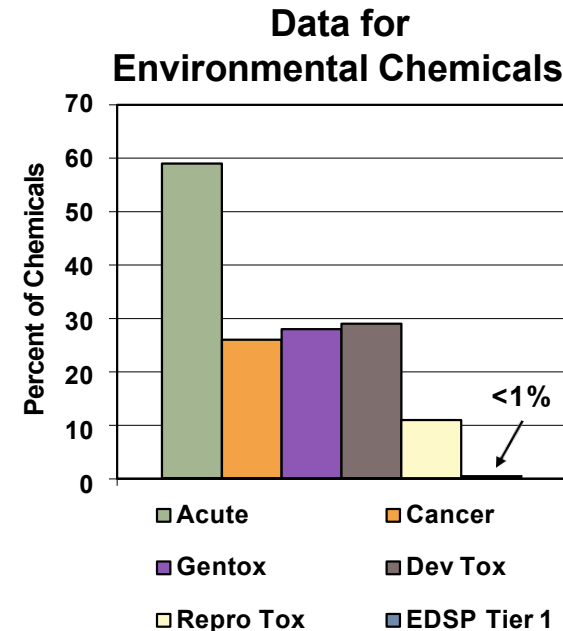
- A *very short* overview of cheminformatics focused on
  - Chemical identifiers and some associated challenges
  - Molecular fingerprints
  - Molecular similarity
  - Structure-based modeling (QSAR/QSPR/QSUR)
- An overview of the CompTox Chemicals Dashboard and how it can help to:
  - Search, source, visualize and download data for singleton or thousands of chemicals
  - Perform real-time prediction calculations and read-across
  - Navigate into dozens of other online resources that contain additional data

# Problem: Too Many Chemicals and Too Few Resources

- Fast characterization of human and ecological risk posed by existing and emerging chemicals is a critical challenge
- Chemistry never stops. But there is sparse and distributed data...



CAS REGISTRY® contains more than **171 million unique organic and inorganic chemical substances**, such as alloys, coordination compounds, minerals, mixtures, polymers and salts, and more than 68 million protein and DNA sequences



Modified from Judson *et al.*, EHP 2010

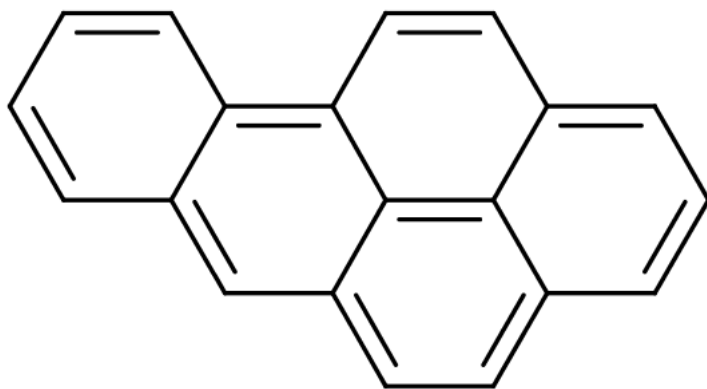
- Develop a “first-stop-shop” for environmental chemical data to support EPA and partner decision making:
  - **Centralized location** for relevant chemical data
  - Chemistry, exposure, hazard and dosimetry
  - Combination of existing data and predictive models
  - Publicly accessible, periodically updated, curated
- Easy access to data improves efficiency and ultimately accelerates chemical risk assessment

# Cheminformatics and the Dashboard

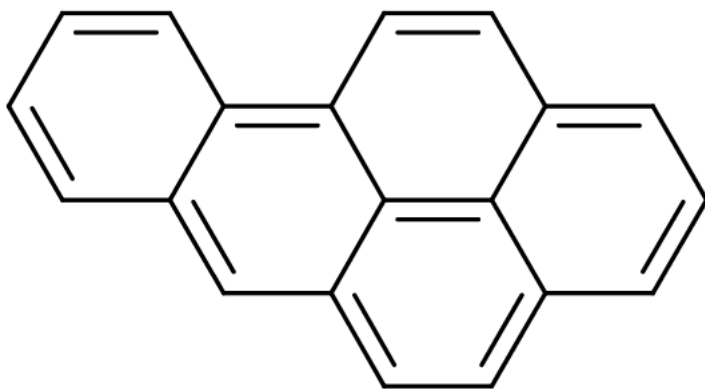


- Cheminformatics is the application of computer science and informatics-based approaches to:
  - Represent chemical structures, substances and reactions
  - Store chemistry-related data
  - Search for chemistry related data
  - Model data sets to provide predictive capabilities
  - Visualize and analyse chemistry related data
- The US-EPA uses cheminformatics (and bioinformatics) to manipulate, integrate, store, model and deliver access to our data. The CompTox Chemicals Dashboard is built on a solid cheminformatics foundation

# Types of Chemical Identifiers



- Structural Identifiers
- The visual depiction
- Multiple electronic formats
- InChI (Key): FMMWHPNWAFFZXNH-UHFFFAOYSA-N
- Common Name: Benzo(a)pyrene
- Systematic Name: Benzo[*pqr*]tetraphene
- CAS Registry Number(s) : 50-32-8
- Lots of other “common names and trade names”



## INTRINSIC PROPERTIES

- Formula :  $C_{20}H_{12}$
- Molecular weight: 252.316 g/mol
- Monoisotopic Mass: 252.093900 g/mol

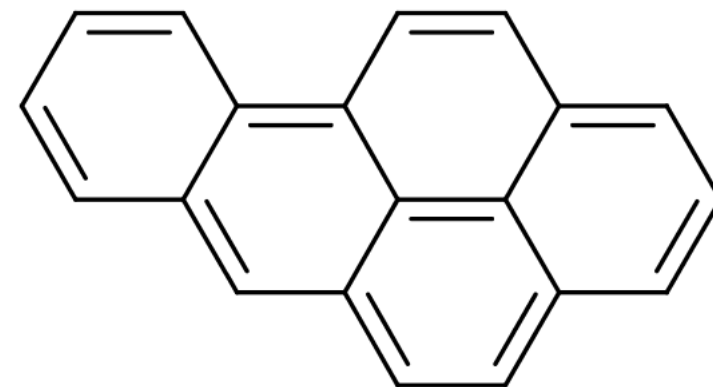
## MEASURED PROPERTIES

- LogKow 6.13
- Melting Pt 177°C
- Boiling Pt 485°C
- ....and many more



# How to Store a Chemical Structure

- Multiple approaches:
  - Names and identifiers
  - 2D or 3D structure “molfile”



```
Mrv1533009301517202D
  0  0  0      0  0      999 V3000
M  V30 BEGIN CTAB
M  V30 COUNTS 20 24 0 0 0
M  V30 BEGIN ATOM
M  V30 1 C 5.3801 0 0 0
M  V30 2 C 6.9201 0 0 0
M  V30 3 C 7.6901 -1.33 0 0
M  V30 4 C 9.2302 -1.33 0 0
M  V30 5 C 10.0003 -2.67 0 0
M  V30 6 C 9.2302 -4.0001 0 0
M  V30 7 C 7.6901 -4.0001 0 0
M  V30 8 C 6.9201 -5.3301 0 0
M  V30 9 C 5.3801 -5.3301 0 0
M  V30 10 C 4.6201 -4.0001 0 0
M  V30 11 C 5.3801 -2.67 0 0
M  V30 12 C 6.9201 -2.67 0 0
M  V30 13 C 4.6201 -1.33 0 0
M  V30 14 C 3.0801 -1.33 0 0
M  V30 15 C 2.31 -2.67 0 0
M  V30 16 C 3.0801 -4.0001 0 0
M  V30 17 C 0.77 -2.67 0 0
M  V30 18 C 0 -1.33 0 0
M  V30 19 C 0.77 0 0 0
M  V30 20 C 2.31 0 0 0
M  V30 END ATOM
M  V30 BEGIN BOND
M  V30 1 2 1 2
M  V30 2 1 1 13
M  V30 3 1 2 3
M  V30 4 2 3 4
M  V30 5 1 3 12
M  V30 6 1 4 11
```


- SMILES:
  - c1cc2c3ccc4cccc5ccc(cc2cc1)c3c45
  - C1=CC2=CC3=CC=C4C=CC=C5C=CC(=C2C=C1)C3=C45
  - and many other variants....
- InChI=1S/C20H12/c1-2-7-17-15(4-1)12-16-9-8-13-5-3-6-14-10-11-18(17)20(16)19(13)14/h1-12H
- InChIKey: FMMWHPNWAFFZXNH-UHFFFAOYSA-N


# If We Database Chemical Structures...


- ...then we can search the dataset by inherent **structural** properties
  - Formula
  - Mass
  - Substructure
  - Structural similarity
- ...we can **integrate** other info into the database for retrieval
- ...available data, both experimental and predicted, is a click away
- ...data can be downloaded, distributed and shared
- ...linking out to other resources enabled by adopting specific standards
- ...structure collections, with associated data, are available for modeling


# CompTox Chemicals Dashboard


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

 United States Environmental Protection Agency  
Home Advanced Search Batch Search Lists Predictions Downloads  
875 Thousand Chemicals  
Chemicals Product/Use Categories Assay/Gene  
SEARCH  
Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey  
Identifier subsetting search  
See what people are saying: read the dashboard comments!  
Cite the Dashboard Publication click here

 United States Environmental Protection Agency  
Home Advanced Search Batch Search Lists Predictions Downloads  
Batch Search  
Step Four: Select Data Output Format and Choose Data Fields to Download  
Please enter one identifier per line  
Select Input Type(s)  
Identifiers  
Chemical Name  
CASRN  
InChIKey  
DTXSID Substance  
DTXSID Compound ID  
InChIKey System  
MS-Ready Formulation  
Exact Formulation  
Monoterpene Mixture  
Display All Chemicals Download Chemical Data  
Select Output Format  
Excel Download  
Customize Results  
Select All  
Select All Lists  
Chemical Identifiers  
DTXSID  
Chemical Name  
To Not activate quickly, click Search in left sidebar  
Download Data for Excel

 UNITED STATES ENVIRONMENTAL PROTECTION AGENCY  
Chemicals Product/Use Categories Assay/Gene  
List of Terpenes added to Dashboard  
19th, 2019 at 12:28:28 AM  
Chemical list related to terpenes found in vape has been added to the dashboard and is available at: [comptox.epa.gov/dashboard/chemical\\_lists/VAPETERPENES](https://comptox.epa.gov/dashboard/chemical_lists/VAPETERPENES)


 United States Environmental Protection Agency  
Home Advanced Search Batch Search Lists Predictions Downloads  
Bisphenol A  
80-05-7 | DTXSID7020182  
Searched by Expert Validated Synonym.  
Hazard  
Data Type  
Toxicity Value  
Download Columns  
HAZARD  
ADME  
EXPOSURE  
BIODIVERSITY  
SIMILAR COMPOUNDS  
GENRA (BETA)  
RELATED SUBSTANCES  
SYNONYMS  
LITERATURE  
LINKS  
DETAILS  
EXECUTIVE SUMMARY  
PROPERTIES  
ENV. RATE/TRANSPORT  
HAZARD  
ADME  
EXPOSURE  
BIODIVERSITY  
SIMILAR COMPOUNDS  
GENRA (BETA)  
RELATED SUBSTANCES  
SYNONYMS  
LITERATURE  
LINKS  
TOXCAST DATA  
ASSAY DETAILS  
AC50 (AD50) 5.75  
Scaled log(5.00)  
Assay Endpoint Name: OT\_UR\_EstRn\_0480  
Assay Description: TSC  
Gene Symbol: ESR1  
Organism: Human  
Tissue: Kidney  
Assay Format Type: cell-based  
Biological Process Target: protein stabilization  
Detection Technology: Protein-fragment Complementation Analysis  
Analysis Direction: positive  
Intended Target Family: nuclear receptor  
Description: Data from the assay component OT\_UR\_EstRn\_0480 was analyzed into 1 assay endpoint. This assay endpoint, OT\_UR\_EstRn\_0480, was analyzed in the positive fitting direction relative to DMSO as the negative control and baseline of activity. Using a type of binding reporter, measures of receptor for gain-of-signal activity can be used to understand the binding at the pathway-level as they relate to the gene ESR1. Furthermore, this assay endpoint can be referred to as a primary readout, because the performed assay has only produced 1 assay endpoint. To generalize the intended target to other related targets, this assay endpoint is associated to the "nuclear receptor" intended target family, where the subfamily is "nuclear".

 United States Environmental Protection Agency  
Home Advanced Search Batch Search Lists Predictions Downloads  
Bisphenol A  
80-05-7 | DTXSID7020182  
Searched with a similarity threshold of 0.8  
378 of 390 chemicals visible  
Select all Download Send to Batch Search Similarity 0.8  
4-Compound  
CASRN:69-59-4  
DTXSID:DTXSID695858  
TOXCAT:100179  
4-(1,1-diphenyl-2-ethyl-1H-1,2,3-triazol-4-yl)phenol  
CASRN:939-97-2  
DTXSID:DTXSID939958  
TOXCAT:1  
4-E-C-Phenylphenylphenol  
CASRN:1571-75-1  
DTXSID:DTXSID157175-1  
TOXCAT:160179  
4-E-C-Ethene-1,1,1-triphenyl  
CASRN:2795-84-4  
DTXSID:DTXSID279584-4  
TOXCAT:140179  
4-(Diphenylmethyl)phenol  
CASRN:978-86-9  
DTXSID:DTXSID97886-9  
TOXCAT:1

# CompTox Chemicals Dashboard



## 883k Chemical Substances



United States  
Environmental Protection  
Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

Share▼



### CompTox Chemicals Dashboard

883 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](#)

#### 10th Release of the CompTox Chemicals Dashboard Now Live July 12th 2020

July 21st, 2020 at 9:32:02 PM

The 10th release of the Dashboard is now live with >7000 additional substances added to the dataset, updates to Bioactivity Data (ToxCast/Tox21), updates to the ToxVal data (under the Hazard tab), a new Safety Tab integrating the Globally Harmonized System of Classification and Labeling of Chemicals (via PubChem), over thirty new lists and a number of bug fixes. Our next release is scheduled for late Spring/Early Summer 2021. and is presently in development. It will be a full re-architecting of the entire application. Watch this space for updates. The release addresses a number of minor bugs and includes a short list of additional functionality as described in the [Release Notes here](#).

UNITED STATES


Discover

Connect


Ask

# BASIC Search


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




Benzo(a)pyrene  
DTXSID2020139




Benzo(a)pyrene diolepoxide 1  
DTXSID9036779




Benzo(a)pyrene- 7,8,9-triol,7,8,9,10-tetrahydro-, (7-alpha,8-beta,9-beta)-  
DTXSID00210066




Benzo(a)pyrene-1-methanol  
DTXSID40235374




Benzo(a)pyrene-1,6-dione, 7-methyl-  
DTXSID70229645




Benzo(a)pyrene-10-methanol  
DTXSID20235817



Benzo(a)pyrene-10-sulfonic acid, 7,8,9,10-tetrahydro-7,8,9-trihydroxy-, (7alpha,8beta,9beta)  
DTXSID80154378



Benzo(a)pyrene-11,12-diol  
DTXSID70215609



Benzo(a)pyrene-11,12-diol, 11,12-dihydro-, cis-  
DTXSID20214501

- Type ahead search using Names, synonyms and CASRNs
- Millions of identifiers
- Substring search

**Search Results**  
Searched with 'Synonym Substring': Benzo(A)Pyrene

**183 chemicals**

# Search for classes of chemicals

- Examples: “perfluoro”

**Chemicals**

Product/Use Categories

Assay/Gene

 perfluoro

☒ Identifier substring search

## Search Results

Searched with 'Synonym Substring': Perfluoro

Select all



Download

Send to Batch Search

Substring



DTXSID

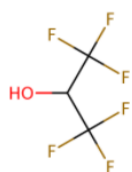
CASRN

TOXCAST

2098 chemicals

Hide chemicals that are:

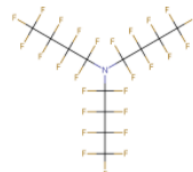
Filter by Name or CASRN



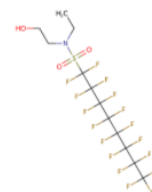
2H-Perfluoro-2-propanol  
DTXSID:DTXSID1022134  
CASRN:920-66-1  
TOXCAST:-



Perfluorooctanesulfonyl fluoride  
DTXSID:DTXSID5027140  
CASRN:307-35-7  
TOXCAST:-



Perfluorotributylamine  
DTXSID:DTXSID0027141  
CASRN:311-89-7  
TOXCAST:-



N-Ethyl-N-(2-hydroxyethyl)perfluorooctanesulfonamide  
DTXSID:DTXSID6027426  
CASRN:1691-99-2  
TOXCAST:-



N-Methyl-N-(2-hydroxyethyl)perfluorooctanesulfonamide  
DTXSID:DTXSID7027831  
CASRN:24448-09-7  
TOXCAST:-

0 related chemical  
structures with this  
substance

Perfluoro compounds, C5-18  
DTXSID:DTXSID5029059  
CASRN:86508-42-1  
TOXCAST:7/235



# Challenges with Nomenclature

- Be CAREFUL with names! There is a LOT of confusion in the public domain. CHOOSE sources wisely!
- There are MANY public databases but not many are curated
- All public databases have value but not many curate data
- Example: METHANE on PubChem  
<https://pubchem.ncbi.nlm.nih.gov/compound/297>

# CAS Registry Numbers on PubChem

## 2.3.1 CAS



74-82-8

- ▶ CAMEO Chemicals; CAS Common Chemistry; ChemIDplus; DrugBank; EPA Chemicals under the TSCA; EPA DSSTox; European Chemicals Agency (ECH...

8006-14-2

- ▶ CAMEO Chemicals; EPA Chemicals under the TSCA; EPA DSSTox; European Chemicals Agency (ECHA)

7440-44-0

- ▶ ChemIDplus

7782-40-3

- ▶ ChemIDplus

7782-42-5

- ▶ ChemIDplus

16291-96-6

- ▶ ChemIDplus

64365-11-3



# CASRN lookup on the dashboard

## Search Results

Searched using Batch Search

8 chemicals

Select all

Download

Send to Batch Search

Default



CASRN

DTXSID

Molecular Formula



Hide chemicals that are:

Filter by Name or CASRN



[Methane](#)

CASRN:74-82-8

DTXSID:DTXSID8025545

Molecular Formula:CH4



[Carbon](#)

CASRN:7440-44-0

DTXSID:DTXSID9027651

Molecular Formula:C

0 related chemical  
structures with this  
substance

[Natural gas](#)

CASRN:8006-14-2

DTXSID:DTXSID2027676

Molecular Formula:-

2 related chemical  
structures with this  
substance

[Graphite](#)

CASRN:7782-42-5

DTXSID:DTXSID2049634

Molecular Formula:-

1 related chemical  
structure with this  
substance

[Charcoal](#)

CASRN:16291-96-6

DTXSID:DTXSID2051217

Molecular Formula:-

1 related chemical  
structure with this  
substance

[Diamond](#)

CASRN:7782-40-3

DTXSID:DTXSID10905072

Molecular Formula:-

0 related chemical  
structures with this  
substance

[Activated charcoal](#)

CASRN:64365-11-3

DTXSID:DTXSID801019028

Molecular Formula:-

1 related chemical  
structure with this  
substance

[Carbon nanotubes](#)

CASRN:308068-56-6

DTXSID:DTXSID301020377

Molecular Formula:-

# Methane is Diamond and Nanotubes?

- These are all Depositor Names for Methane ☹️

## 2.4.2 Depositor-Supplied Synonyms



UN 1971 (Salt/Mix)	Fullerene soot, (as produced)	Carbon Nanotube sponges XFCN01	DTXSID9027651
UN 1972 (Salt/Mix)	MWNTs Butyl acetate suspension	Carbon Nanotube sponges XFCN07	Graphite electrode, rotrode disc
Activated carbon, pellets 3mm	QuadraPure C, 0.3-0.8mm	Carbon Nanotube sponges XFCN08	Carbon conductive cement adhesive
Graphene quantum dots(Powder)	6GRV67N0U2	Carbon, activated, -4+8 mesh	Conductive Flexible TPU Filament
Multiwall Nanotubes 5-15 nm	GO quantum dots yellow(Powder)	Carbon, activated, 2mm & down	GO quantum dots yellow(1mg/ml)
GO quantum dots(C: 1mg/ml)	Graphene electric aqueous slurry	CHEMBL2106049	Graphite powder, -20+84 mesh
;) MWNTs ethyl acetate suspension	Graphene powder Physical methods	Diamond Synthesized, 95% Nano	Carbon black, Super P Conductive
Reduced Graphene Oxide@ SnO2	Reduced Graphene Oxide@ Co3O4	Diethyl Cyanomethyl Phosphonate	DTXSID50179391
Carbon nanotubes aqueous slurry	Carbon Conductive Adhesive Tapes	DTXSID8025545	NanoIntegris metallic SWCNTs70%



# Detailed Chemical Pages

## One more identifier – the **DTXSID**

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

SAFETY

ADME

EXPOSURE

BIOACTIVITY

SIMILAR COMPOUNDS


GENRA (BETA)

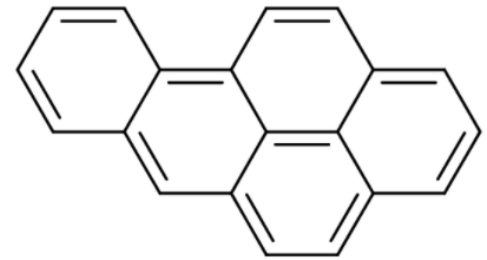
RELATED SUBSTANCES

SYNONYMS

LITERATURE

LINKS

 Benzo(a)pyrene  
50-32-8 | DTXSID2020139  
Searched by DSSTox Substance Id.



Wikipedia

Benzo[a]pyrene is a polycyclic aromatic hydrocarbon and the result of incomplete combustion of organic matter at temperatures between 300 °C (572 °F) and 600 °C (1,112 °F). The ubiquitous compound can be found in coal tar, tobacco smoke and many foods, especially grilled meats. The substance with the formula C<sub>20</sub>H<sub>12</sub> is one of the benzopyrenes, formed by a benzene ring fused to pyrene. Its diol epoxide metabolites (more commonly known as BPDE) react and bind to ...  
[Read more](#)

Quality Control Notes

Intrinsic Properties

Molecular Formula: C<sub>20</sub>H<sub>12</sub> [Mol File](#) [Find All Chemicals](#)

Average Mass: 252.316 g/mol [Isotope Mass Distribution](#)

Monoisotopic Mass: 252.0939 g/mol

Structural Identifiers


Linked Substances

Presence in Lists

- Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

# Detailed Chemical Pages

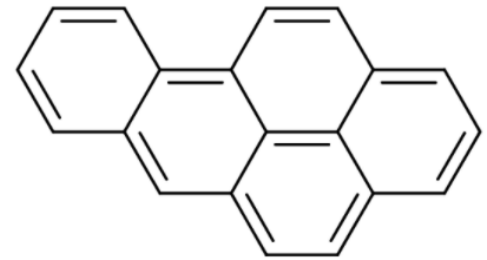
## Easy Navigation



### Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.



#### DETAILS




- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ▶ SAFETY
- ▶ ADME
- ▶ EXPOSURE
- ▶ BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- ▶ LITERATURE
- LINKS



#### Wikipedia


**Benzo[a]pyrene** is a polycyclic aromatic hydrocarbon and the result of incomplete combustion of organic matter at temperatures between 300 °C (572 °F) and 600 °C (1,112 °F). The ubiquitous compound can be found in coal tar, tobacco smoke and many foods, especially grilled meats. The substance with the formula C<sub>20</sub>H<sub>12</sub> is one of the benzopyrenes, formed by a benzene ring fused to pyrene. Its diol epoxide metabolites (more commonly known as BPDE) react and bind to ...  
[Read more](#)

#### Quality Control Notes

#### Intrinsic Properties

 **Molecular Formula:** C<sub>20</sub>H<sub>12</sub>  Mol File  Find All Chemicals

 **Average Mass:** 252.316 g/mol  Isotope Mass Distribution

 **Monoisotopic Mass:** 252.0939 g/mol

#### Structural Identifiers

#### Linked Substances

#### Presence in Lists

- Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances

# From the Chemical Details Page... all chemicals with same FORMULA

## Intrinsic Properties



Molecular Formula:  $C_{20}H_{12}$




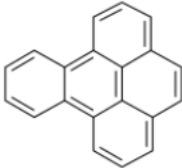
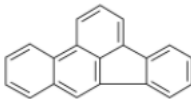
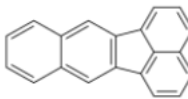

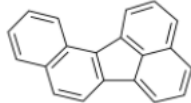
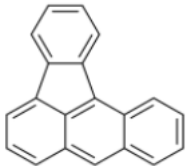
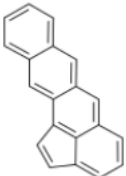
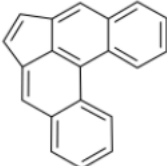
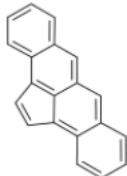
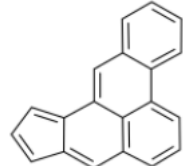
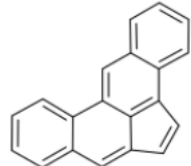
Mol File

 Find All Chemicals

## Search Results

Searched by Exact Molecular Formula:  $C_{20}H_{12}$ .

Download ▾ Send to Batch Search Default ▾ CASRN X DTXSID X ▾ 27 chemicals Hide chemicals that are: ▾ Filter by Name or C

 <p>Benzo(a)pyrene CASRN:50-32-8 DTXSID:DTXSID2020139</p>	 <p>Benzo(e)pyrene CASRN:192-97-2 DTXSID:DTXSID3023764</p>	 <p>Benzo(b)fluoranthene CASRN:205-99-2 DTXSID:DTXSID0023907</p>	 <p>Benzo(k)fluoranthene CASRN:207-08-9 DTXSID:DTXSID0023909</p>	 <p>Perylene CASRN:198-55-0 DTXSID:DTXSID4047753</p>	 <p>Benzo(j)fluoranthene CASRN:205-82-3 DTXSID:DTXSID8052691</p>
 <p>Benzo[a]fluoranthene CASRN:203-33-8 DTXSID:DTXSID4059756</p>	 <p>Cyclopenta(de)naphthalene CASRN:16683-64-0 DTXSID:DTXSID80168197</p>	 <p>Benz(a)acephenanthrylene CASRN:192-28-9 DTXSID:DTXSID70172748</p>	 <p>Cyclopenta(fg)naphthalene CASRN:19770-52-6 DTXSID:DTXSID40173469</p>	 <p>Benzo(de)cyclopent(a)anthracene CASRN:198-46-9 DTXSID:DTXSID60173507</p>	 <p>Benz(e)aceanthrylene CASRN:199-54-2 DTXSID:DTXSID30173675</p>

# How many chemicals are associated through LINKED SUBSTANCES?

- Atrazine, is a herbicide – in MANY commercial products
- The dashboard has salt forms, isotopically labelled forms, multicomponent forms
- How do we identify what they are???

## Linked Substances

**Same Connectivity:** [6 records](#) (based on first layer of InChI)


**Mixtures, Components and Isotopomers:** [DTXCID90112: 25 records;](#)


**Similar Compounds:** [73 records](#) (based on Tanimoto coefficient >0.8)


# A little more about the InChI


- An InChIKey is made up of two blocks...
  - Block 1 – “the connectivity” of atoms and bonds
  - Block 2 – isotopes, charge, stereo

### Structural Identifiers

 **IUPAC Name:** 6-Chloro-N~2~-ethyl-N~4~-(propan-2-yl)-1,3,5-triazine-2,4-diamine

 **SMILES:** CCNC1=NC(NC(C)C)=NC(Cl)=N1

 **InChI String:** 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)

 **InChIKey:** MXWJVTOOROXGIU-UHFFFAOYSA-N

Search Google for:

- The InChIKey is VERY USEFUL

- Demo an internet search using InChIs – Cholesterol has the InChIKey: HVYWMOMLDIMFJA-DPAQBDIFSA-N
- Demo Atrazine – Linked Substances – Skeleton
- More about Linked Substances....



# Linked Substances – more interesting

- We map chemicals together using cheminformatics approaches
- Use desalting, destereo, split multicomponents etc to map chemicals together

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

Journal of Cheminformatics

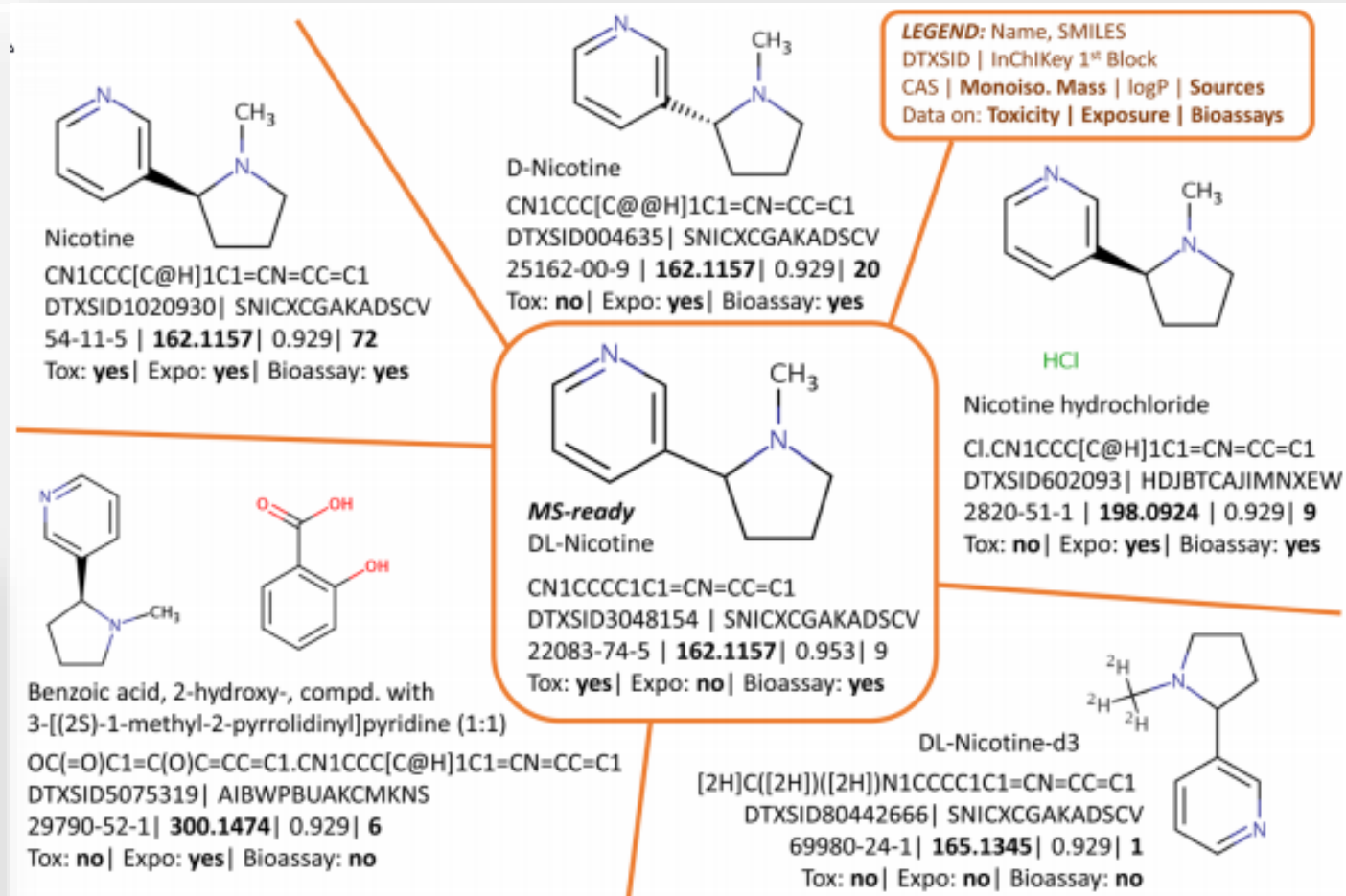
## METHODOLOGY

Open Access



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

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

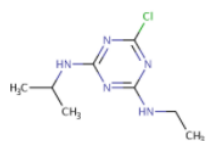


# Atrazine Linked Substances

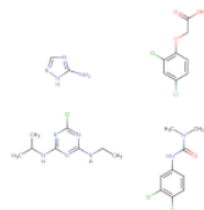
## MS-Ready Mappings of Atrazine (Isotopes pre-filtered)

20 of 25 chemicals visible

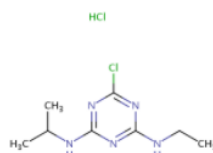
Select all Download Send to Batch Search Default DTXSID CASRN TOXCAST Isotopes Filter by Name or CASRN



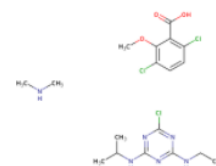
**Atrazine**  
DTXSID:DTXSID9020112  
CASRN:1912-24-9  
TOXCAST:62/1024



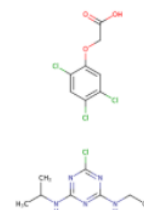
**Anox M**  
DTXSID:DTXSID50156021  
CASRN:128996-76-9  
TOXCAST:-



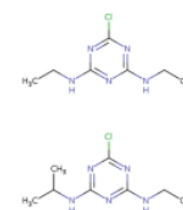
**s-Triazine, 2-chloro-4-(ethylamino)-6-(isopropylamino)-1,3,5-triazine hydrochloride**  
DTXSID:DTXSID30165459  
CASRN:15386-47-7  
TOXCAST:-



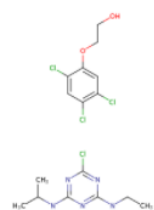
**Marksman**  
DTXSID:DTXSID80166936  
CASRN:160544-50-3  
TOXCAST:-



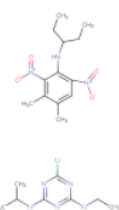
**Acetic acid, (2,4,5-trichlorophenoxy)-, methyl ester**  
DTXSID:DTXSID70192527  
CASRN:39283-62-0  
TOXCAST:-



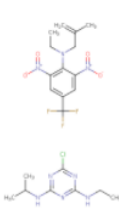
**1,3,5-Triazine-2,4-diamine, 6-chloro-N,N-dimethyl-**  
DTXSID:DTXSID60192556  
CASRN:39331-45-8  
TOXCAST:-



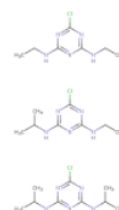
**Buvinol**  
DTXSID:DTXSID10199555  
CASRN:51602-05-2  
TOXCAST:-



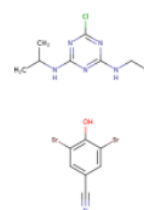
**Atrazine mixture with pendimethalin**  
DTXSID:DTXSID10209527  
CASRN:60704-01-0  
TOXCAST:-



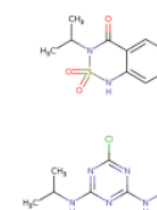
**Maizor**  
DTXSID:DTXSID20215154  
CASRN:64867-15-8  
TOXCAST:-



**Polytriazine**  
DTXSID:DTXSID00222508  
CASRN:72172-70-4  
TOXCAST:-



**Benzonitrile, 3,5-dibromo-4-hydroxy-, methyl ester**  
DTXSID:DTXSID20226063  
CASRN:75084-56-9  
TOXCAST:-



**Bentazon / atrazine**  
DTXSID:DTXSID80226064  
CASRN:75084-57-0  
TOXCAST:-

# A little more about our data quality

- Five full time curators register and curate data to elevate quality

## Record Information



**Citation:** U.S. Environmental Protection Agency. CompTox Chemicals Dashboard. <https://comptox.epa.gov/dashboard/DTXSID0020022> (accessed October 18, 2020), 5-(2-Chloro-4-(trifluoromethyl)phenoxy)-2-nitrobenzoic acid

### Data Quality:

Level 1: Expert curated, highest confidence in accuracy and consistency of unique chemical identifiers

Level 2: Expert curated, unique chemical identifiers using multiple sources

Level 3: Programmatically curated from high quality EPA source, unique chemical identifiers have no conflicts in ChemID and PubChem

Level 4: Programmatically curated from ChemID, unique chemical identifiers have no conflicts in PubChem

Level 5: Programmatically curated from ACToR or PubChem, unique chemical identifiers with low confidence, single public source

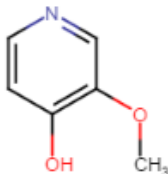
# Underneath the Dashboard

View/Edit a Single Record   Structure Search   Browse/Curate Records   Export DSSTox   Chemotypes   Manage Chemical Lists   Manage Property Data   Add Deleted Casms

Preferred Name matched <b>null</b>  
You are viewing the record associated with  
DTXSID80198757  
CASRN: 62885-41-0

4-Hydroxy-3-methoxy

Valid license cannot be found



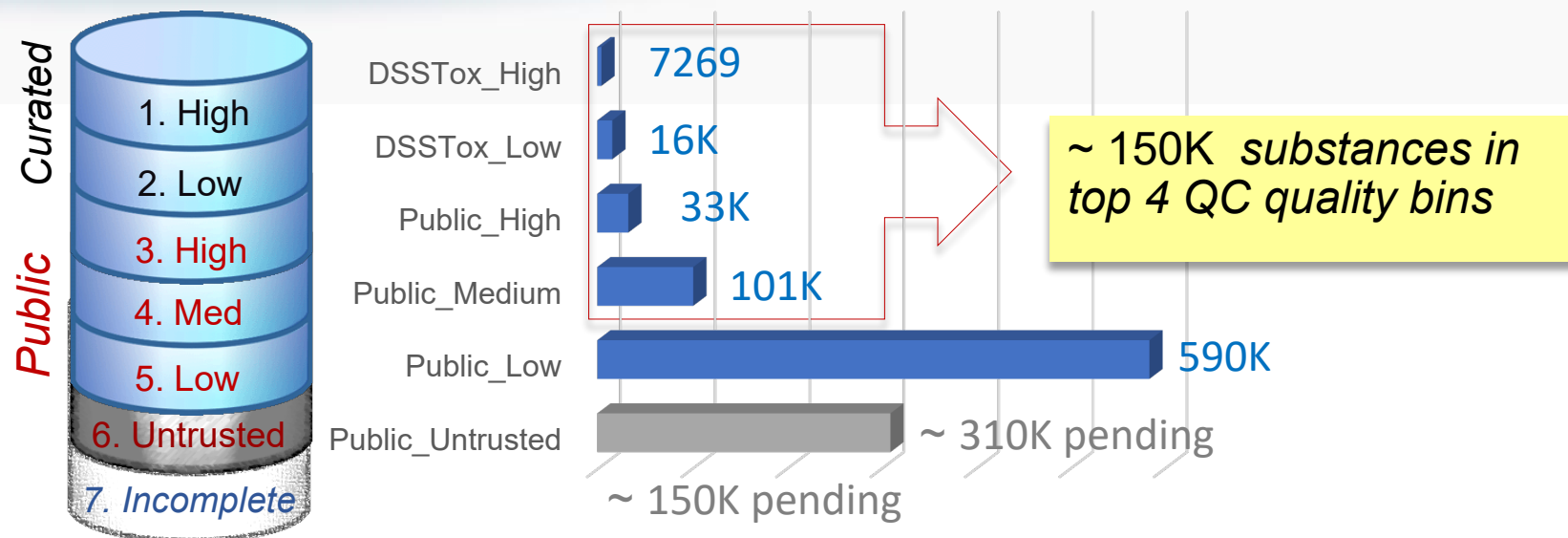
Chemical structure of 4-Hydroxy-3-methoxypyridine (DTXSID80198757) is displayed. The structure shows a pyridine ring with a hydroxyl group (-OH) at the 4-position and a methoxy group (-OCH<sub>3</sub>) at the 3-position.

Calculate from Structure

Substance_ID:	DTXSID80198757	Compound_ID:	DTXCID40121248
CAS:	62885-41-0	Chemical Shown:	Tested Chemical
Name:	4-Hydroxy-3-methoxypyridine	Private Notes:	
Substance Type:	Single Compound	Source of CAS-Compound:	STN(DSSTox)
QC Level:	DSSTox_High	Double Stereo:	None
Data Source:	STN(DSSTox)	Chiral Stereo:	None
QC Notes:	CAS [50700-60-2] assigned by DSSTox to pyridin-one tautomer form, which resolves to hydroxy form thru InChI	Chemical Form:	Organic
		Organic Form:	Parent

# Distribution of curated data

## Now at >1.2 MILLION substances



QC Levels	
DSSTox_High:	Hand curated and validated
DSSTox_Low:	Hand curated and confirmed using multiple public sources
Public_High:	Extracted from EPA SRS and confirmed to have no conflicts in ChemID and PubChem
Public_Medium:	Extracted from ChemID and confirmed to have no conflicts in PubChem
Public_Low:	Extracted from ACToR or PubChem
Public_Untrusted:	Postulated, but found to have conflicts in public sources

# A little more about our data quality

Computational Toxicology 12 (2019) 100096



ELSEVIER

Contents lists available at ScienceDirect

## Computational Toxicology

journal homepage: [www.elsevier.com/locate/comtox](http://www.elsevier.com/locate/comtox)

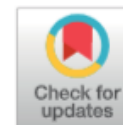


### EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research

Christopher M. Grulke<sup>a</sup>, Antony J. Williams<sup>a</sup>, Inthirany Thillanadarajah<sup>b</sup>, Ann M. Richard<sup>a,\*</sup>

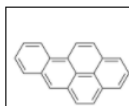
<sup>a</sup> National Center for Computational Toxicology, Office of Research & Development, US Environmental Protection Agency, Mail Drop D143-02, Research Triangle Park, NC 27711, USA

<sup>b</sup> Senior Environmental Employment Program, US Environmental Protection Agency, Research Triangle Park, NC 27711, USA





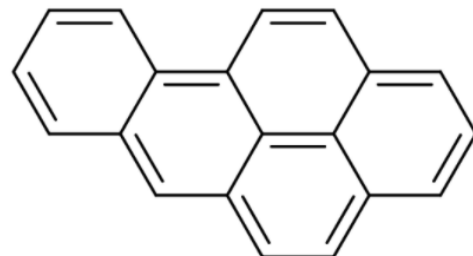
# Navigating data via the Left Hand Tabs



## Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.



### DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ SAFETY

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

LINKS

### Wikipedia

**Benzo[a]pyrene** is a polycyclic aromatic hydrocarbon and the result of incomplete combustion of organic matter at temperatures between 300 °C (572 °F) and 600 °C (1,112 °F). The ubiquitous compound can be found in coal tar, tobacco smoke and many foods, especially grilled meats. The substance with the formula  $C_{20}H_{12}$  is one of the benzopyrenes, formed by a benzene ring fused to pyrene. Its diol epoxide metabolites (more commonly known as BPDE) react and bind to

...  
[Read more](#)

### Quality Control Notes

### Intrinsic Properties



Molecular Formula:  $C_{20}H_{12}$



Mol File



Find All Chemicals



Average Mass: 252.316 g/mol



Isotope Mass Distribution




Monoisotopic Mass: 252.0939 g/mol

### Structural Identifiers

### Linked Substances

### Presence in Lists

# Experimental and Predicted Data



## Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

Property

Summary

Download Columns

Property	Experimental average	Predicted average
<a href="#">Water Solubility</a>	8.40e-9 (4)	1.75
<a href="#">LogKow: Octanol-Water</a>	6.13 (2)	6.24
<a href="#">Vapor Pressure</a>	5.49e-9 (1)	3.61e-9
<a href="#">Boiling Point</a>	495 (3)	480
<a href="#">Henry's Law</a>	4.57e-7 (1)	4.59e-7
<a href="#">Melting Point</a>	177 (8)	189
<a href="#">Surface Tension</a>	-	53.9
<a href="#">Flash Point</a>	-	234
<a href="#">Density</a>	-	1.28

- Physchem and Fate & Transport experimental and predicted data
- Data can be downloaded as Excel, TSV and CSV files
- Predictions: multiple algorithms
  - EPI Suite: Estimation Program Interface
  - ACD/Labs (commercial)
  - TEST: Toxicity Estimation Software Tool
  - OPERA: **O**PEn structure–activity/property **R**elationship **A**pp



## ToxVal Database

- >50k chemicals
- >770k tox. values
- >30 sources of data
- ~5k journals cited
- ~70k citations

Hazard

DataType  
Toxicity Value

Human Eco

Download Columns 10 Search query

More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	7	cancer slope factor	-	chronic	23.5	(mg/kg-day)-1	-	dermal	-	Alaska DEC	<a href="#">Alaska DEC</a>
	7	cancer unit risk	-	chronic	0.21	(mg/l)-1	-	inhalation	-	Alaska DEC	<a href="#">Alaska DEC</a>
	7	cancer slope factor	-	chronic	3.08	(mg/kg-day)-1	-	inhalation	-	Alaska DEC	<a href="#">Alaska DEC</a>
	7	cancer unit risk	-	chronic	0.88	(mg/m3)-1	-	inhalation	-	Alaska DEC	<a href="#">Alaska DEC</a>
	7	cancer slope factor	-	chronic	7.3	(mg/kg-day)-1	-	oral	-	Alaska DEC	<a href="#">Alaska DEC</a>
	7	MEG	Short-term Critical Air	short-term	80	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Short-term Marginal Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Soil Negligible Soil	chronic	12	mg/kg	-	Soil	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Long-Term, SL/d Negligible Water	chronic	0.0134	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>
	7	MEG	Short-term Negligible Air	short-term	0.6	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	<a href="#">DOD</a>

<< < 1 2 3 4 > >>

Showing 1 to 10 of 32 records

[anced\\_search/index](#)

## GHS Data

Print Page

PUBCHEM > BENZO[A]PYRENE > LABORATORY CHEMICAL SAFETY SUMMARY (LCSS) > GHS CLASSIFICATION




CID 2336

# Benzo[a]pyrene

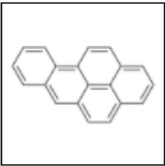
## GHS Classification



Showing 6 of 6

Pictogram(s)	<div></div> <div>Irritant    Health Hazard    Environmental Hazard</div>
Signal	<u>Danger</u>
GHS Hazard Statements	H317: May cause an allergic skin reaction [ <u>Warning</u> Sensitization, Skin] H340: May cause genetic defects [ <u>Danger</u> Germ cell mutagenicity] H350: May cause cancer [ <u>Danger</u> Carcinogenicity] H360FD: May damage fertility; May damage the unborn child [ <u>Danger</u> Reproductive toxicity] H400: Very toxic to aquatic life [ <u>Warning</u> Hazardous to the aquatic environment, acute hazard] H410: Very toxic to aquatic life with long lasting effects [ <u>Warning</u> Hazardous to the aquatic environment, long-term hazard]
Precautionary Statement Codes	P201, P202, P261, P272, P273, P280, P281, P302+P352, P308+P313, P321, P333+P313, P363, P391, P405, and P501 (The corresponding statement to each P-code can be found at the <a href="#">GHS Classification</a> page.)

# Identifiers Support Searches in other systems



Benzo(a)pyrene  
50-32-8 | DTXSID2020139  
Searched by DSSTox Substance Id.

## Synonyms

Download

25

Search query

Synonym	Quality
Benzo(a)pyrene	Valid
Benzo[pqr]tetraphene	Valid
Benzo[a]pyrene	Valid
50-32-8 <span>Active CAS-RN</span>	Valid
BaP	Valid
Benzo[a]pyrene	Good
3,4-Benz[a]pyrene	Good
3,4-Benzopyrene	Good
3,4-Benzpyrene	Good
6,7-Benzopyrene	Good
BENZ(A)PYREN	Good
Benz(a)pyrene	Good
Benz[a]pyrene	Good

- CASRNs are very useful, and still limited
- Not every chemical has a STRUCTURE...substances vs structures
- “Chemical Abstracts Service” – numbers don’t exist until they abstracted and indexed
- Not every chemical on the dashboard necessarily has a CASRN – how would you find those that didn’t??? Hint: Search NOCAS\_
- There are ~6000 chemicals without CASRN on the dashboard
- A chemical can also have many deleted CASRNs

## What chemicals are in hair care products?

Chemicals

**Product/Use Categories**

Assay/Gene

Q

hair

CPDat PRODUCT category: hair coloring  
*general hair coloring products which can not be classified into a more refined category*

CPDat PRODUCT category: hair coloring hair bleach  
*products for lightening or removing color from hair on the head*

CPDat PRODUCT category: hair coloring hair color - permanent  
*hair colors and dyes characterized as permanent*

CPDat PRODUCT category: hair coloring hair color - professional  
*hair colors and dyes characterized as for professional use*

CPDat PRODUCT category: hair coloring hair color activator  
*chemical activators for hair coloring products*

CPDat PRODUCT category: hair coloring hair color developer  
*chemical developers for hair coloring products*

CPDat PRODUCT category: hair styling and care dry shampoo  
*products for removing oil and dirt from hair*

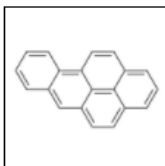
CPDat PRODUCT category: hair styling and care  
*general hair styling or hair care products which do not fit into a more refined category*

# Let's Talk Exposure



- Types of Exposure Data on the Dashboard
  - Consumer product categories and uses
  - Products containing the chemical
  - Predicted exposure levels from modeling (more in next session)

# Sources of Exposure to Chemicals



Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

## Chemical Weight Fractions

 Download ▾

Columns ▾

10 ▾

Search query

Product Name	Product Use Category	Minimum Weight Fraction	Maximum Weight Fraction	Data Type	Source
m-525-1-5x pah mixtures 0.5 mg/ml for method 525	Not Yet Categorized:			MSDS	<a href="#">SIRI</a>
mm6125 surface conditioner	Not Yet Categorized:			Health Product Declaration	<a href="#">Health Product Declaration Collaborative</a>
monolithic membrane 6125 (mm6125) / monolithic membrane	Not Yet Categorized:			Health Product Declaration	<a href="#">Health Product Declaration Collaborative</a>
organic potablewatr pw 32_component h:reg semi-volatile 690	Not Yet Categorized:	0.00	1.00e-3	MSDS	<a href="#">SIRI</a>
polynuclear aromatic hydrocarbon mixture_ep84627	Not Yet Categorized:			MSDS	<a href="#">SIRI</a>
prestone(r) power steering fluid	engine maintenance: auto fluids and additives			MSDS	<a href="#">CPCPdb</a>
r-12 shield tite wet surface coating	Not Yet Categorized:	0.00	0.500	MSDS	<a href="#">SIRI</a>
sea tar 1010_0028	Not Yet Categorized:			MSDS	<a href="#">SIRI</a>
supelprime-hc kit pah mix_48909	Not Yet Categorized:			MSDS	<a href="#">SIRI</a>
supelprime-hc pah mix 1mL_48905	Not Yet Categorized:			MSDS	<a href="#">SIRI</a>

- What do you trust more? Experimental or predicted data?
- Do you trust individual models or consensus models
- What if there are no experimental data, how good are predictions?



# Data Curation Pipelines plus Manual Curation Processes

> SAR QSAR Environ Res. 2016 Nov;27(11):939-965. doi: 10.1080/1062936X.2016.1253611.

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

K Mansouri <sup>1 2</sup>, C M Grulke <sup>2</sup>, A M Richard <sup>2</sup>, R S Judson <sup>2</sup>, A J Williams <sup>2</sup>

Research article | [Open Access](#) | Published: 08 March 2018

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

[Kamel Mansouri](#) , [Chris M. Grulke](#), [Richard S. Judson](#) & [Antony J. Williams](#)

[Journal of Cheminformatics](#) **10**, Article number: 10 (2018) | [Cite this article](#)

**9195** Accesses | **90** Citations | **25** Altmetric | [Metrics](#)

# Property and Fate and Transport Data ~25 MILLION pre-predicted values

- We have built QSPR models based on tens of thousands of property data points curated over the past decade
- We push our “QSAR-Ready” chemical structures through predictions to produce property predictions

Methodology | [Open Access](#) | Published: 30 August 2018

## **"MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies**

[Andrew D. McEachran](#) , [Kamel Mansouri](#), [Chris Grulke](#), [Emma L. Schymanski](#), [Christoph Ruttkies](#) & [Antony J. Williams](#) 

[Journal of Cheminformatics](#) **10**, Article number: 45 (2018) | [Cite this article](#)

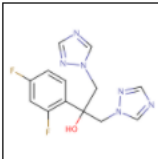
**4640** Accesses | **34** Citations | **13** Altmetric | [Metrics](#)

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT



Fluconazole

86386-73-4 | DTXSID3020627


Searched by DSSTox Substance Id.

Property

LogKow: Octanol-Water

Download Summary

## Predicted

 Download Predicted Data

Source	Result	Calculation Details	QMRf
EPISUITE	0.250	Not Available	Not Available
ACD/Labs Consensus	0.698	Not Available	Not Available
ACD/Labs	0.500	Not Available	Not Available
OPERA	0.501	<a href="#">OPERA Model Report [Inside AD]</a>	Available

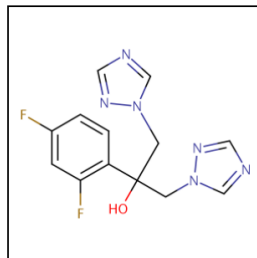
# OPERA Reports

OPERA Models: LogKow: Octanol-Water

Fluconazole

86386-73-4 | DTXSID3020627

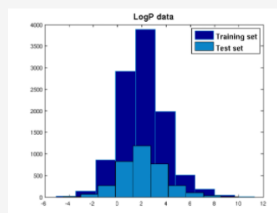
Print PDF



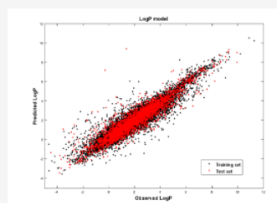
## Model Results

Predicted value: 0.501  
Global applicability domain: Inside  
Local applicability domain index: 0.998  
Confidence level: 0.732

## Model Performance



QMRP

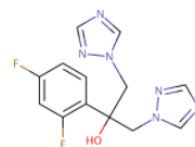


Weighted KNN model

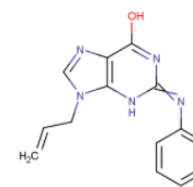
Weighted KNN model

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

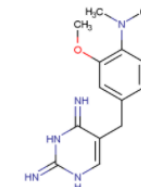
## Nearest Neighbors from the Training Set



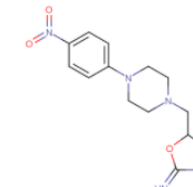
Fluconazole  
Measured: 0.5  
Predicted: 0.50



GUANINEN2PHENYL9ALLYL  
Measured: 1.75  
Predicted: 1.75



2,4-PYRIMIDINEDIAMINE, 5-[(4-(DIMETHYLAMINO)-3-M  
Measured: 1.87  
Predicted: 1.87



5-(1-P-NITROPHENYL-4-PIPERAZINYL)METHYL-2-AMINO  
Measured: 1.23  
Predicted: 1.23

# Similar reports for TEST predictions

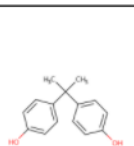
## Predicted Normal boiling point for 80-05-7 from Consensus method

Prediction results

Endpoint	Experimental value	Predicted value
Normal boiling point Å°C	N/A	359.93

Individual Predictions

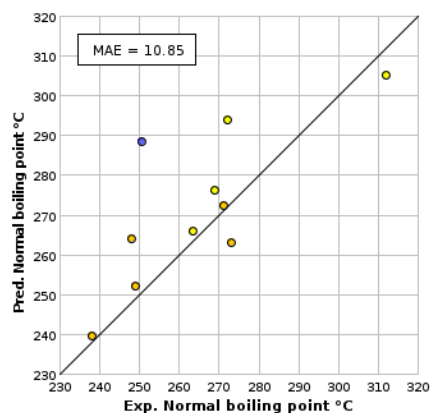
Method	Predicted value Å°C
Hierarchical clustering	372.06
Group contribution	377.41
Nearest neighbor	330.33



## 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

Prediction results (colors defined in table below)




Chemicals	MAE*
Entire set	11.46
Similarity coefficient $\geq 0.5$	10.85

\*Mean absolute error in Å°C

CAS	Structure	Similarity Coefficient	Experimental value Å°C	Predicted value Å°C
80-05-7 (test chemical)			N/A	359.93
14938-35-3		0.81	250.50	288.54
28994-41-4		0.75	312.00	305.28
96-76-4		0.71	263.50	265.93
4130-42-1		0.70	272.00	293.96
616-55-7		0.70	269.00	276.42
2052-14-4		0.69	271.00	272.30

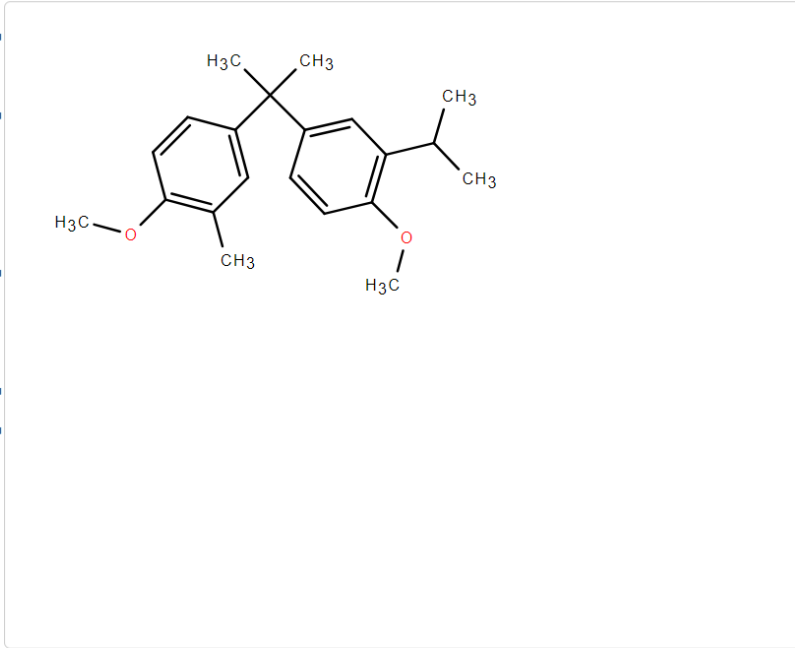
# Real-Time Predictions



United States  
Environmental Protection  
Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data



Chemical structure diagram showing a biphenyl system. The left ring has a methoxy group (H<sub>3</sub>C-O-) at the para position and a methyl group (-CH<sub>3</sub>) at the ortho position. The right ring has a methyl group (-CH<sub>3</sub>) at the ortho position, a propoxy group (-O-CH<sub>2</sub>-CH<sub>2</sub>-CH<sub>3</sub>) at the para position, and a methyl group (-CH<sub>3</sub>) at the other ortho position. The two rings are connected by a central carbon atom bonded to two methyl groups (H<sub>3</sub>C- and -CH<sub>3</sub>).

**Select properties to predict**

**H** **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**

- ✓ Toxicological properties
  - ✓ 96 hour fathead minnow LC50
  - ✓ 48 hour D. magna LC50
  - ✓ 48 hour T. pyriformis IGC50
  - ✓ Oral rat LD50
  - ✓ Bioconcentration 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

# 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		6.051 -Log10(mol/L) 0.278 mg/L	5.678 -Log10(mol/L) 0.656 mg/L	5.572 -Log10(mol/L) 0.836 mg/L	5.908 -Log10(mol/L) 0.386 mg/L	7.047 -Log10(mol/L) 0.028 mg/L
48 hour D. magna LC50		5.591 -Log10(mol/L) 0.802 mg/L	5.548 -Log10(mol/L) 0.884 mg/L	6.169 -Log10(mol/L) 0.212 mg/L	5.518 -Log10(mol/L) 0.948 mg/L	5.128 -Log10(mol/L) 2.329 mg/L
48 hour T. pyriformis IGC50		5.590 -Log10(mol/L) 0.804 mg/L	6.390 -Log10(mol/L) 0.127 mg/L		5.588 -Log10(mol/L) 0.806 mg/L	4.790 -Log10(mol/L) 5.068 mg/L
Oral rat LD50		2.400 -Log10(mol/kg) 1243.951 mg/kg	2.232 -Log10(mol/kg) 1829.942 mg/kg			2.568 -Log10(mol/kg) 845.609 mg/kg
Bioaccumulation factor		3.066 Log10 1164.438	3.090 Log10 1230.849	2.717 Log10 521.420	3.257 Log10 1806.262	3.200 Log10 1585.959
Developmental toxicity		true	true	true		true
Ames mutagenicity		false	false			false
Estrogen Receptor RBA		-0.710 Log10 0.195	-1.692 Log10 0.020	-1.515 Log10 0.031		1.077 Log10 11.931
Estrogen Receptor Binding		false	false	false		true
Normal boiling point		345.2 °C	306.6 °C		408.2 °C	320.7 °C
Melting point		74.3 °C	63.8 °C		41.0 °C	118.2 °C
Flash point		161.7 °C	143.5 °C		152.7 °C	188.9 °C
Vapor pressure		-5.955 Log10(mmHg) 1.109*10 <sup>-6</sup> mmHg	-5.534 Log10(mmHg) 2.925*10 <sup>-6</sup> mmHg		-5.903 Log10(mmHg) 1.249*10 <sup>-6</sup> mmHg	-6.428 Log10(mmHg) 3.735*10 <sup>-7</sup> mmHg
Density		0.959 g/cm <sup>3</sup>	0.977 g/cm <sup>3</sup>		0.843 g/cm <sup>3</sup>	1.057 g/cm <sup>3</sup>

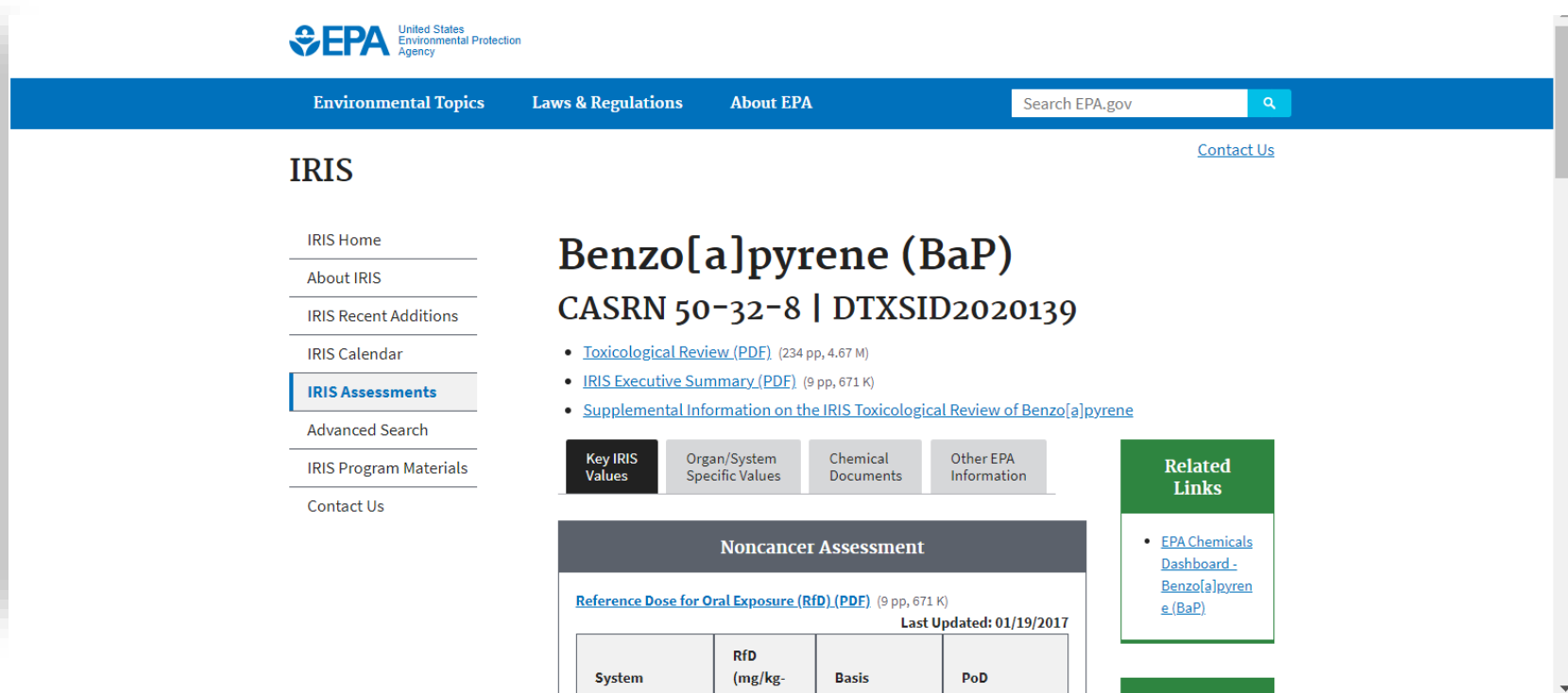
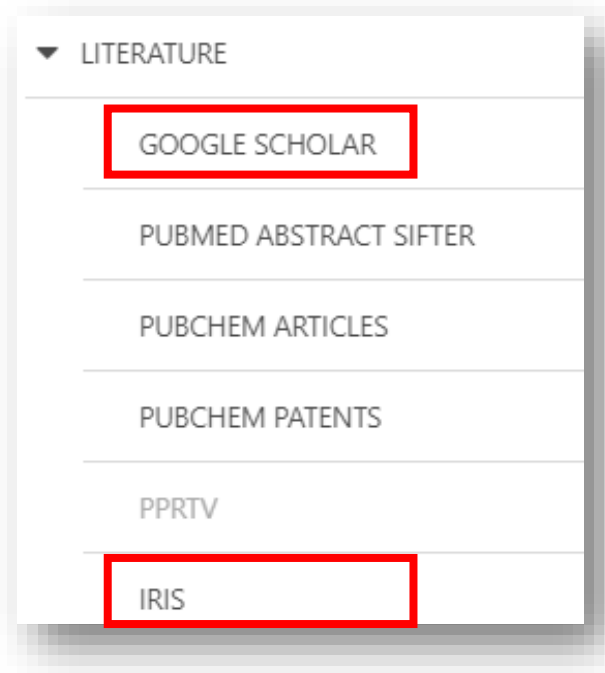


# What's the best way to search the internet for chemical data?

- We know how complex chemicals identifiers are...
  - CASRN(s)
  - Hundreds of names (maybe)
  - SMILES
  - InChIs
  - EINECS, EC numbers
- What can WE do to help you navigate the internet?

# Identifiers are used in the app

- Identifiers are used to feed and link into “Literature”



EPA United States Environmental Protection Agency

Environmental Topics Laws & Regulations About EPA Search EPA.gov

Contact Us

## IRIS

- IRIS Home
- About IRIS
- IRIS Recent Additions
- IRIS Calendar
- IRIS Assessments**
- Advanced Search
- IRIS Program Materials
- Contact Us

### Benzo[a]pyrene (BaP)

CASRN 50-32-8 | DTXSID2020139

- [Toxicological Review \(PDF\)](#) (234 pp, 4.67 M)
- [IRIS Executive Summary \(PDF\)](#) (9 pp, 671 K)
- [Supplemental Information on the IRIS Toxicological Review of Benzo\[a\]pyrene](#)

Key IRIS Values Organ/System Specific Values Chemical Documents Other EPA Information

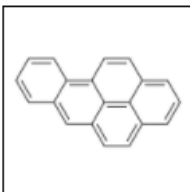
#### Noncancer Assessment

[Reference Dose for Oral Exposure \(RfD\) \(PDF\)](#) (9 pp, 671 K) Last Updated: 01/19/2017

System	RfD (mg/kg-day)	Basis	PoD

#### Related Links

- [EPA Chemicals Dashboard - Benzo\[a\]pyrene \(BaP\)](#)




Benzo(a)pyrene

50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.

## Abstract Sifter

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

Hazard



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.



("50-32-8" OR "Benzo(a)pyrene") AND (NOAEL OR NOEL OR LOEL OR Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

- Real-time retrieval of data from PubMed (~30 million abstracts and growing)
- Choose from set of pre-defined queries
- Adjust and fine tune queries based on interests

# Literature Searching

- “Sifting” of results using multiple terms
- Frequency counting terms
- Color highlighting of terms
- Download list to Excel
- Send list to PubMed for downloading ref. file
- Direct link via PubMed ID

To find articles quickly, enter terms to sift abstracts. 

dermal cancer pyrene      

<input type="checkbox"/>	dermal	cancer ↓	pyrene	Total	PMID	Year	Title	Authors	Journal	Rev
<input type="checkbox"/>	0	7	1	8	23922326	2013	Using immunotoxicity information to improve cancer risk a...	Zaccaria, McClure	International journal of toxicology	✓
<input type="checkbox"/>	8	7	2	17	16632147	2006	Development of a dermal cancer slope factor for benzo[a]...	Knafila, Philipps, Brecher, Petrovic, Richardson	Regulatory toxicology and pharmacology : RTP	✓
<input type="checkbox"/>	4	6	2	12	33359623	2020	Testing the validity of a proposed dermal cancer slope fac...	Magee; Forsberg	Regulatory toxicology and pharmacology : RTP	✓
<input type="checkbox"/>	0	5	1	6	28477805	2017	Pollution characteristics, sources and lung cancer risk of ...	Wang; Xia; Wu; Zhang; Sun; Yin; Zhou; Yang	Journal of environmental sciences (China)	
<input type="checkbox"/>	4	4	2	10	20888881	2010	Development and application of a skin cancer slope factor...	Knafila; Petrovic; Richardson; Campbell; Rowat	Regulatory toxicology and pharmacology : RTP	
<input type="checkbox"/>	4	4	1	9	16307791	2005	Health risk assessment on human exposed to environme...	Chen; Liao	The Science of the total environment	
<input type="checkbox"/>	2	4	1	7	11807932	2002	Cancer risk assessment for oral exposure to PAH mixtures.	Schneider; Roller; Kalberlah; Schuhmacher-Wolz	Journal of applied toxicology : JAT	
<input type="checkbox"/>	2	3	1	6	32460055	2020	PAHs in Chinese atmosphere Part II: Health risk assessm...	Ma; Zhu; Liu; Jia; Yang; Li	Ecotoxicology and environmental safety	
<input type="checkbox"/>	0	3	1	4	23379661	2013	Parent and halogenated polycyclic aromatic hydrocarbon...	Ni; Guo	Journal of agricultural and food chemistry	
<input type="checkbox"/>	0	3	1	4	20800879	2010	Health risk assessment on dietary exposure to polycyclic ...	Xia; Duan; Qiu; Liu; Wang; Tao; Jiang; Lu; Song; Hu	The Science of the total environment	
<input type="checkbox"/>	2	3	1	6	16293284	2005	Probabilistic risk assessment for personal exposure to car...	Liao; Chiang	Chemosphere	
<input type="checkbox"/>	0	2	1	3	17544483	2007	Health risk assessment for traffic policemen exposed to p...	Hu; Bai; Zhang; Wang; Zhang; Yu; Zhu	The Science of the total environment	
<input type="checkbox"/>	0	1	1	2	28795279	2017	Human health risk assessment and PAHs in a stretch of ri...	Srivastava; Sreekrishnan; Nema	Environmental monitoring and assessment	
<input type="checkbox"/>	0	1	1	2	12634119	2003	Deviation from additivity in mixture toxicity: relevance of n...	Lutz; Vamvakas; Kopp-Schneider; Schlatter; Stopper	Environmental health perspectives	
<input type="checkbox"/>	0	1	2	3	3709501	1986	The adsorption of polyaromatic hydrocarbons on natural a...	Menard; Noel; Khorami; Jouve; Dunnigan	Environmental research	
<input type="checkbox"/>	0	0	1	1	33136306	2020	Effects on Animal Outcomes of Regulatory Relevance of F...	Crumo; Boulanger; Farhat; Williams; Basu; Hecker	Environmental toxicology and chemistry	

**Development of a dermal cancer slope factor for benzo[a]pyrene.**  
Polycyclic aromatic hydrocarbons (PAHs) are commonly found at environmentally impacted sites in both Canada and the United States, and also occur naturally. Typically, benzo[a]pyrene (B[a]P) is selected as a standard to which the cancer potencies of other carcinogenic PAHs are compared. Cancer potency estimates for B[a]P have been published for the oral and inhalation routes of exposure, however, no such estimate has been established by a regulatory agency for dermal exposure. The main objectives of the current investigation were to: evaluate approaches used to examine the relative carcinogenicity of PAHs; to conduct a review of mammalian dermal carcinogenicity studies for B[a]P; and derive a cancer slope factor for dermal exposure to PAHs using B[a]P as a surrogate for other PAHs. The toxicological database of dermal B[a]P studies was examined for relevant animal bioassays. Seven relevant studies were identified. A cancer slope factor for B[a]P was developed using the benchmark dose approach and the linearized multistage model. The upper 95th CI at the 5% effect level above background incidence was used as the point of departure for low-dose linear extrapolation. An average slope factor of 0.55 (microg/animal day)<sup>-1</sup> was calculated for mice, which was converted to a dose-equivalent slope factor of 25 (mg/kg day)<sup>-1</sup>. This latter slope factor is proposed for application to human health risk assessment with no scaling adjustment. Dermal potency equivalency factor values were identified which may be used with other carcinogenic PAH in the calculation of total B[a]P equivalent dermal cancer risk estimates. An identified area for further investigation is the consideration of scaling in extrapolating the calculated dermal cancer slope factor from mice to humans.

# External Links – Also use Identifiers Names, CASRN, PubChem IDs, InChIs...























Benzo(a)pyrene




















50-32-8 | DTXSID2020139

Searched by DSSTox Substance Id.




















## General

-  EPA Substance Registry Service
-  PubChem
-  ChempSpider
-  CPCat
-  DrugBank
-  Wikipedia
-  MSDS Lookup
-  ChEMBL
-  ToxPlanet
-  ACS Reagent Chemicals
-  Wolfram Alpha
-  ECHA Infocard
-  ChemAgora
-  Consumer Product Information Database
-  ChEBI
-  NIST Chemistry Webbook
-  WEBWISER
-  PubChem Safety Sheet
-  Consumer Product Information Database
-  PubChem: Chemical Vendors













## Toxicology

-  ACToR
-  DrugPortal
-  CCRIS
-  ChemView
-  CTD
-  eChemPortal
-  Gene-Tox
-  HSDB
-  ACToR PDF Report
-  CREST
-  National Air Toxics Assessment
-  ECOTOX
-  ChemView
-  Chemical Checker
-  BindingDB
-  CalEPA OEHHHA
-  NIOSH IDLH Values
-  LactMed
-  ECOTOX






## Publications

-  Toxline
-  PPRTVWEB
-  PubMed
-  IRIS Assessments
-  EPA HERO
-  NIOSH Skin Notation Profiles
-  NIOSH Pocket Guide
-  RSC Publications
-  BioCaddie DataMed
-  Springer Materials
-  Bielefeld Academic Search Engine
-  CORE Literature Search
-  Google Books (Text Search)
-  Google Patents (Text search)
-  Google Scholar (Text search)
-  Google Patents (Structure search)
-  Google Books (Structure Search)
-  Google Scholar (Structure search)
-  Federal Register

## Analytical

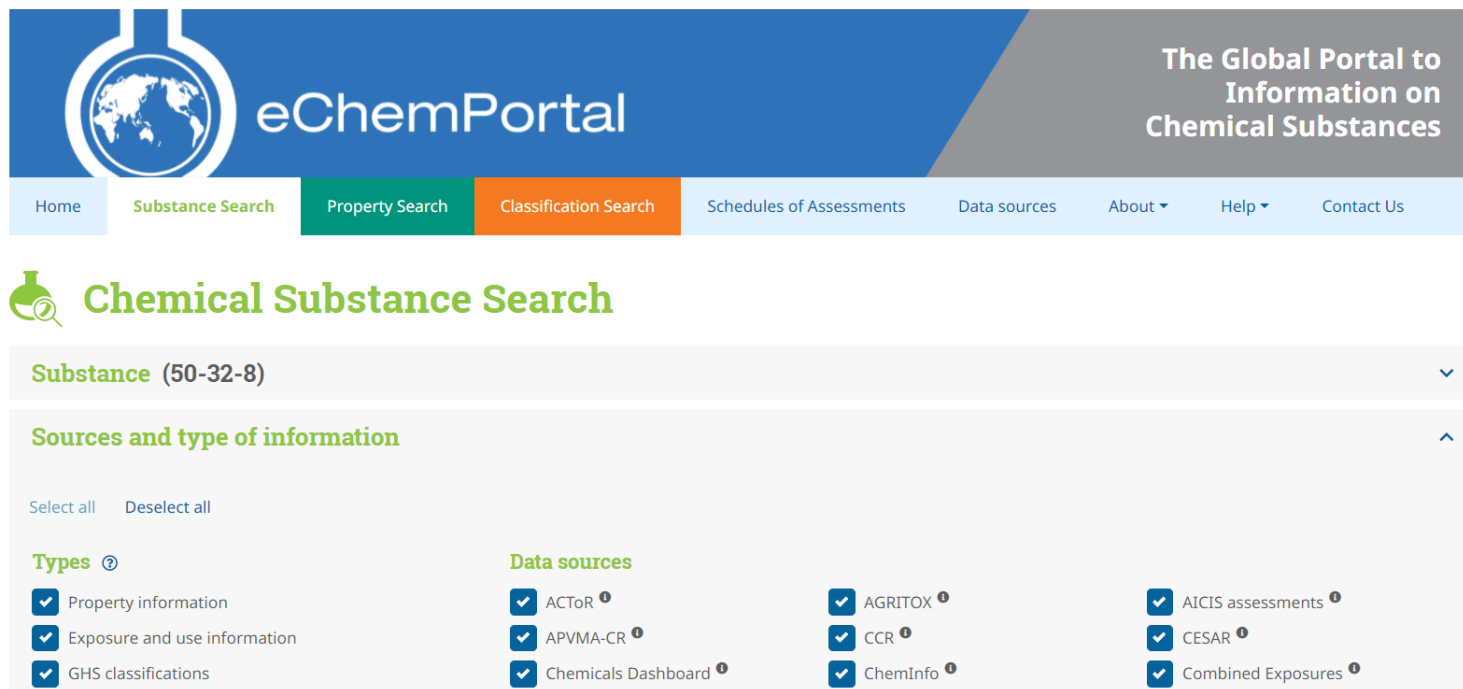
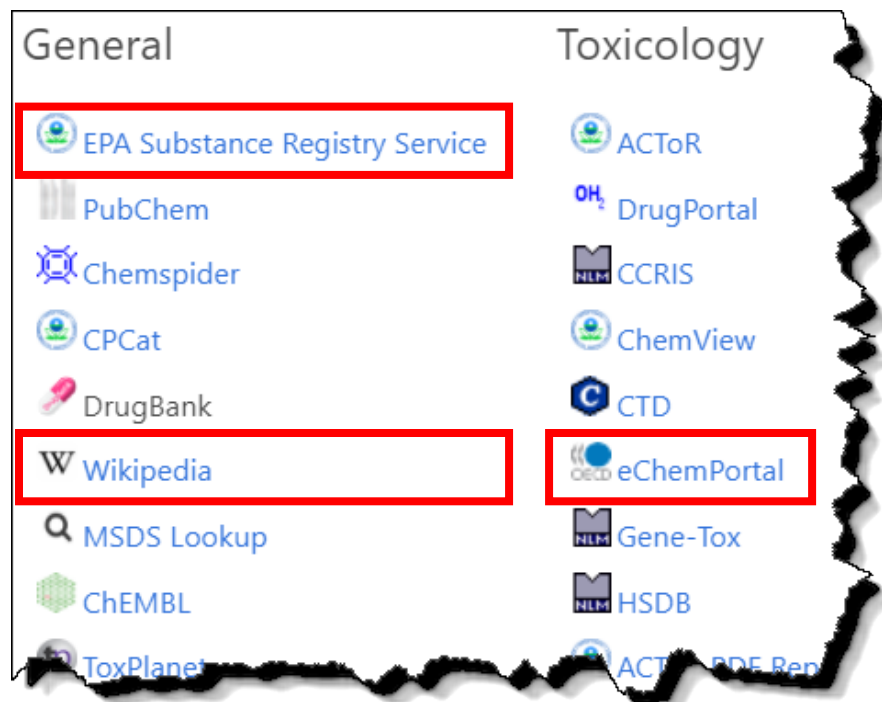
-  RSC Analytical Abstracts
-  Tox21 Analytical Data
-  MONA: MassBank North America
-  mzCloud
-  NIST IR Spectrum
-  NIST MS Spectrum
-  MassBank
-  NIST Antoine Constants
-  IR Spectra on PubChem
-  NIST Kovats Index values
-  Protein DataBank
-  National Environmental Methods Index

## Prediction

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

# External Links

- Links to ~90 websites providing access to additional data on the chemical of interest




The screenshot shows the eChemPortal website. The header features the OECD logo and the text 'The Global Portal to Information on Chemical Substances'. Below the header is a navigation bar with links: Home, Substance Search (highlighted), Property Search, Classification Search, Schedules of Assessments, Data sources, About, Help, and Contact Us. The main content area is titled 'Chemical Substance Search' and shows a search result for 'Substance (50-32-8)'. Under the heading 'Sources and type of information', there are two sections: 'Types' and 'Data sources'. The 'Types' section has checkboxes for 'Property information', 'Exposure and use information', and 'GHS classifications', all of which are checked. The 'Data sources' section has a grid of checkboxes for various sources, all of which are checked: ACToR, APVMA-CR, Chemicals Dashboard, AGRITOX, CCR, ChemInfo, AICIS assessments, CESAR, and Combined Exposures.

# Chemical Lists and Categories

# PFAS lists of Chemicals

## 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)



# Curated List of Pesticides

- Find list of interest

- Select list and send to batch

**EPA** United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

## Select List

### PESTICIDES|EPA: Pesticide Chemical Search Database

Search EPAPCS Chemicals

☐ Identifier substring search

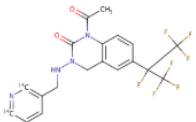
#### List Details

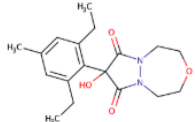
**Description:** The entries in this list have been classified in the U.S. as pesticidal "active ingredients" (conventional, antimicrobial, or biopesticidal agents), and were sourced from the Pesticide Chemical Search database (<https://iaspub.epa.gov/apex/pesticides/f?p=chemicalsearch:1>) created by EPA's Office of Pesticide Programs. Chemical Search provides a single point of reference for easy access to information previously published in a variety of locations, including various EPA web pages and Regulations.gov. Chemical search contains the following: 1) More than 20,000 regulatory documents; 2) Links to over 800 dockets in Regulations.gov 3) Links to pesticide tolerance (or maximum residue levels) information; 4) A variety of web services providing easy access to other scientific and regulatory information on particular chemicals from other EPA programs and federal government sources.

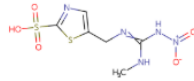
**Number of Chemicals:** 4012

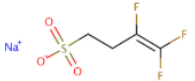
Select all Download **Send to Batch Search** CASRN DTXSID Mono.Mass 3988 chemicals

Hide chemicals that are: Filter by Name or CASRN

  
1-Acetyl-6-(1,1,1,2,3,3,3-heptafluoro-2-pyridyl)-4-methyl-2-pyridone  
CASRN: NOCAS\_920532  
DTXSID: DTXSID00920532  
Mono.Mass: 468.114807

  
8-(2,6-Diethyl-4-methylphenyl)-8-hydroxy-2,3-dihydro-1,4-benzodioxin-5-one  
CASRN: NOCAS\_920508  
DTXSID: DTXSID10920508  
Mono.Mass: 332.173607

  
5-(((Methylamino)(nitroamino)methylene)amino)-2-methylthiophene-3-sulfonic acid  
CASRN: NOCAS\_912338  
DTXSID: DTXSID20912338  
Mono.Mass: 295.004511

  
Sodium 3,4,4-trifluoro-3-buten-1-sulfonate  
CASRN: NOCAS\_912336  
DTXSID: DTXSID00912336  
Mono.Mass: 211.973094

- Singleton searches are great but...
- ...we generally want data on LOTS of chemicals!
- Typical questions
  - What are the structures for a set of chemical names? Set of CASRN's?
  - Can I get chemical lists in Excel files? As a list of SMILES strings?  
Can I get an SDF file?
  - Can I include predicted properties in the download file? OPERA?  
TEST?
  - Are “these chemicals” screened in Toxcast?
  - I’m a mass spectrometrists and need masses and formulae for a list of chemicals

# Access data *en masse* for thousands of chemicals....

## 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 ⓘ
- ☐ DSSTox Compound ID ⓘ
- ☐ InChIKey Skeleton ⓘ
- ☐ MS-Ready Formula(e) ⓘ
- ☐ Exact Formula(e) ⓘ
- ☐ Monoisotopic Mass ⓘ

#### Enter Identifiers to Search (One per line. Searches should be limited to <5000 identifiers.)

Sodium azide  
O,O-Dimethyl S-[(4-oxo-1,2,3-benzotriazin-3(4H)-yl)methyl] phosphorodithioate  
(Trichloromethyl)benzene  
100-44-7  
(Chloromethyl)benzene  
98-07-7  
111-44-4  
1-Chloro-2-(2-chloroethoxy)ethane  
Chloro(chloromethoxy)methane  
2-(Butan-2-yl)-4,6-dinitrophenol

👁 Display All Chemicals

⋮ Download Chemical Data

# Select Output Format and Content

## Step Four: Select Data Output Format and Choose Data Fields to Download

### Select Output Format:







 Excel 

 Download






### Customize Results

- ☐ Select All
- ☐ Select All in Lists






### 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 

### Metadata

- ☐ Curation Level Details 
- ☐ NHANES/Predicted Exposure 
- ☐ Data Sources 
- ☐ Include ToxVal Data Availability 
- ☐ Assay Hit Count 
- ☐ Number of PubMed Articles 
- ☐ PubChem Data Sources 
- ☐ CPDat Product Occurrence Count 
- ☐ IRIS 
- ☐ PPRTV 
- ☐ Wikipedia Article
- ☐ QC Notes 
- ☐ Include links to ACToR reports - SLOW! (BETA) 

### Presence in Lists:

- ☐ 40CFR116.4 Designation of Hazardous Substances (Above Ground Storage Tanks) 
- ☐ 40CFR355 Extremely Hazardous Substance List and Threshold Planning Quantities 
- ☐ AEGLs: Acute Exposure Guideline Levels 
- ☐ ANDROGEN: Androgen Receptor Chemicals 
- ☐ ARTICLE; Bench-Mark Dose Human Health Assessment List (Wignall et al., 2014) 
- ☐ ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (CERAPP) 
- ☐ ARTICLE: Collaborative Estrogen Receptor Activity Prediction Project (COMPARA) 
- ☐ ATSDR Toxicological Profiles 
- ☐ ATSDR: Minimal Risk Levels (MRLs) for Hazardous Substances 
- ☐ ATSDR: Toxic Substances Portal Chemical List 
- ☐ California Office of Environmental Health Hazard Assessment 
- ☐ Canadian Domestic Substances List 2019 
- ☐ CATEGORY: Amino acids 
- ☐ CATEGORY: Color Index dyes 
- ☐ CATEGORY: Flame Retardants 

# Batch Search CASRN

## Batch Search

Step 1

Step 2

Step 3







Step 4

Step 5






### Step Four: Select Data Output Format and Choose Data Fields to Download

Please enter one identifier per line






#### 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 

#### Metadata

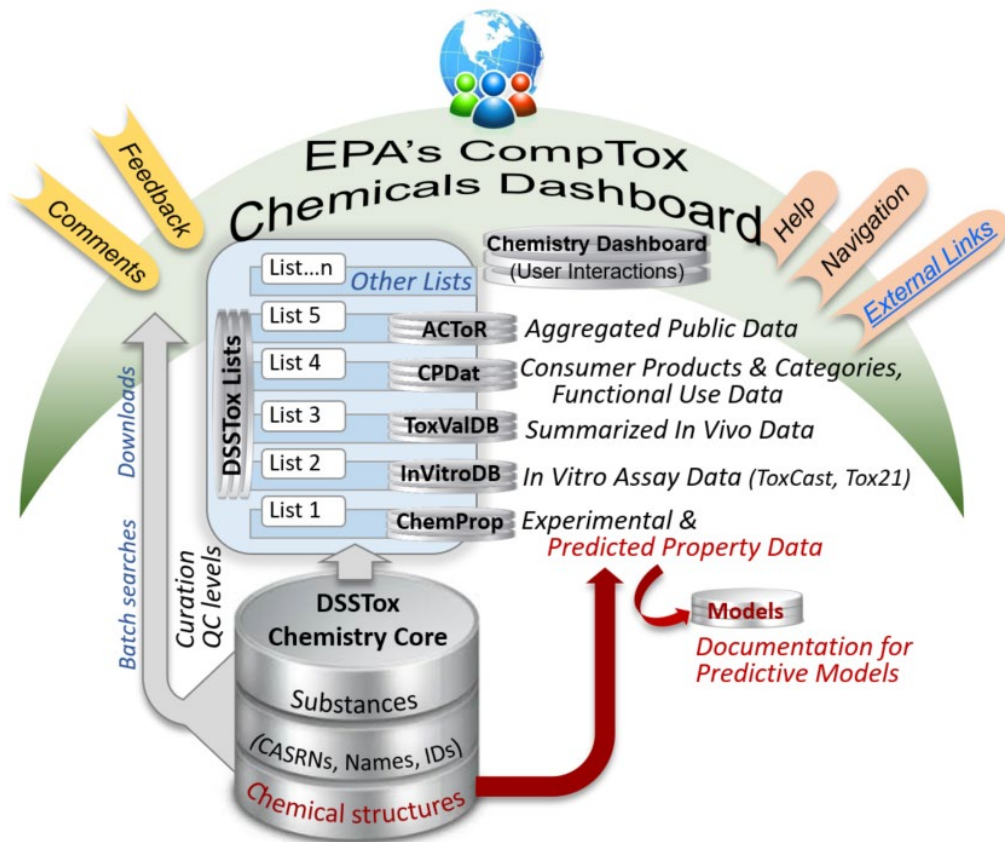
- ☐ Curation Level Details 
- ☐ NHANES/Predicted Exposure 
- ☐ Data Sources 
- ☒ Include ToxVal Data Availability 
- ☒ Assay Hit Count 
- ☒ Number of PubMed Articles 

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	INPUT	FOUND_BY	DTXSID	PREFERRED	CASRN	INCHIKEY	IUPAC NAME	MOLECULAR	TOXVAL_D	TOXCAST	TOXCAST	MS_READY	NUMBER_CIRIS	LINK	PPRTV_LIN	ATMOSPHE	BIOCONCE
2	534-52-1	CAS-RN	DTXSID10	2-Methyl-4,6	534-52-1	ZXVONLUN	2-Methyl-4,6	C7H6N2O5	Y	27.56	261/947	CC1=C(O)C	179	-	Y	1.571E-12	2.88959
3	115-21-9	CAS-RN	DTXSID70	Ethyl silicon	115-21-9	ZOYFEXPF	Trichloro(ethyl)	C2H5Cl3Si	Y	-	-	-	-	-	-	8.289E-12	13.9658
4	111-44-4	CAS-RN	DTXSID90	Bis(2-chloro	111-44-4	ZNSMNVML	1-Chloro-2-(	C4H8Cl2O	Y	1.12	10/891	CICCOCCC	12	Y	-	2.647E-12	9.99608
5	2763-96-4	CAS-RN	DTXSID50	Muscimol	2763-96-4	ZJQHPWUV	5-(Aminome	C4H6N2O2	Y	-	-	NCC1=CC(C	4308	-	-	1.179E-10	5.05695
6	1464-53-5	CAS-RN	DTXSID00	2,2'-Bioxiran	1464-53-5	ZFIVKAOQ	2,2'-Bioxiran	C4H6O2	Y	-	-	C1OC1C1C	363	-	-	4.383E-12	1.27107
7	22224-92-6	CAS-RN	DTXSID30	Fenamiphos	22224-92-6	ZCJPOPBZ	Ethyl 3-methyl	C13H22NO3	Y	10.8	105/972	CCOP(=O)(C	58	Y	-	1.66E-11	2.3394
8	359-06-8	CAS-RN	DTXSID40	Fluoroacetyl	359-06-8	ZBHDYQJ	Fluoroacetyl	C2H2ClFO	Y	-	-	FCC(Cl)=O	-	-	-	3.513E-13	4.49379
9	5344-82-1	CAS-RN	DTXSID40	1-(o-Chloro	5344-82-1	YZUKKTCD	N-(2-Chloro	C7H7ClN2S	Y	-	-	NC(=S)NC1	-	-	-	2.482E-11	9.95206
10	7446-18-6	CAS-RN	DTXSID10	Thallium (I)	7446-18-6	YTQVHRVI	Dithallium(1-	O4STI2	Y	-	-	-	34	Y	Y	-	-
11	62207-76-5	CAS-RN	DTXSID40	Bis(3-fluoro	62207-76-5	YRZXYIHD	-	C16H14CoF	Y	-	-	-	-	-	-	-	-
12	66-81-9	CAS-RN	DTXSID60	Cycloheximi	66-81-9	YPHMISFO	4-[(2R)-2-[(	C15H23NO4	Y	32.27	294/911	CC1CC(C)C	18709	-	-	1.756E-11	2.81761
13	106-96-7	CAS-RN	DTXSID30	Propargyl bi	106-96-7	YORCIIVHU	3-Bromopro	C3H3Br	Y	-	-	BrCC#C	-	-	-	1.069E-11	10.4968
14	315-18-4	CAS-RN	DTXSID70	Mexacarbat	315-18-4	YNEVBPNZ	4-(Dimethyl	C12H18N2C	Y	5.11	12/235	CNC(=O)OC	27	-	-	1.447E-11	26.2914
15	110-00-9	CAS-RN	DTXSID60	Furan	110-00-9	YLQBMQCI	Furan	C4H4O	Y	0.0	0/235	O1C=CC=C	919	Y	-	4.019E-11	5.01648
16	3037-72-7	CAS-RN	DTXSID20	4-(diethoxyn	3037-72-7	YHFFINXFN	4-[Diethoxy(	C9H23NO2	Y	-	-	-	-	-	-	5.95E-12	6.71292
17	75-44-5	CAS-RN	DTXSID00	Phosgene	75-44-5	YGYAWVD	Carbonyl dic	CCl2O	Y	-	-	ClC(Cl)=O	489	Y	-	9.994E-16	13.0711
18	2032-65-7	CAS-RN	DTXSID30	Methiocarb	2032-65-7	YFBPRJGD	3,5-Dimethy	C11H15NO2	Y	18.14	88/485	CNC(=O)OC	65	-	-	1.446E-11	34.1692
19	2778-04-3	CAS-RN	DTXSID20	Endothion	2778-04-3	YCAGGFXS	[(5-Metho	C9H13O6P	Y	-	-	COC1=COCC	-	-	-	4.355E-11	1.18341
20	12108-13-3	CAS-RN	DTXSID90	(Methylcyclo	12108-13-3	YASXMYPV	Tricarbonyl(	C9H7MnO3	Y	2.56	11/430	-	68	-	-	-	-
21	7803-51-2	CAS-RN	DTXSID20	Phosphine	7803-51-2	XYFCBTPG	Phosphane	H3P	Y	-	-	-	928	Y	-	-	-
22	107-18-6	CAS-RN	DTXSID80	Allyl alcohol	107-18-6	XXROGKLT	Prop-2-en-1	C3H6O	Y	3.99	17/426	OCC=C	627	Y	Y	2.592E-11	4.03901
23	108-05-4	CAS-RN	DTXSID30	Vinyl acetate	108-05-4	XTXRWKR	Ethethyl acet	C4H6O2	Y	1.7	4/235	CC(=O)OC	206	Y	-	2.5E-11	5.52157
24	19624-22-7	CAS-RN	DTXSID10	Pentaborane	19624-22-7	XPIBKKWN	-	B5H9	Y	-	-	-	-	-	-	-	-
25	75-74-1	CAS-RN	DTXSID00	Tetramethyl	75-74-1	XOOGZRUE	Tetramethyl	C4H12Pb	Y	-	-	C[Pb](C)(C)	24	-	-	-	-

Worksheet1

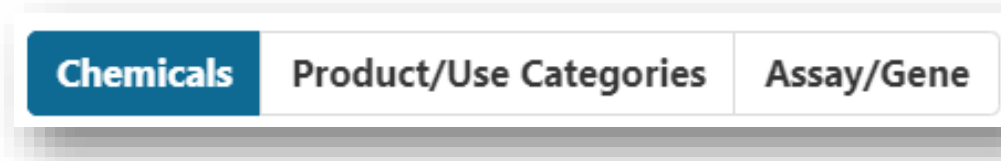


# Summary and Conclusion



- CompTox Chemicals Dashboard - a central hub for environmental data

- ~900k chemical substances
- Integrating property data, hazard data, exposure data, *in vitro* bioactivity data
- Interrogation of bioactivity data -
- Multiple types of searches

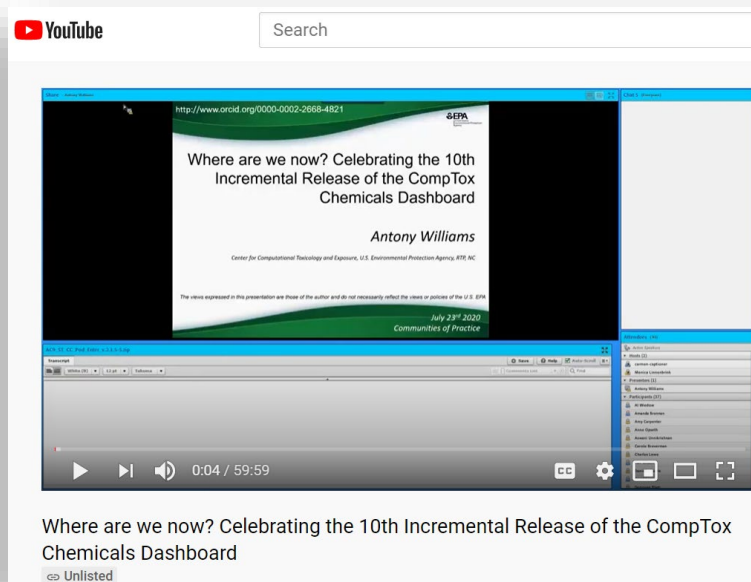
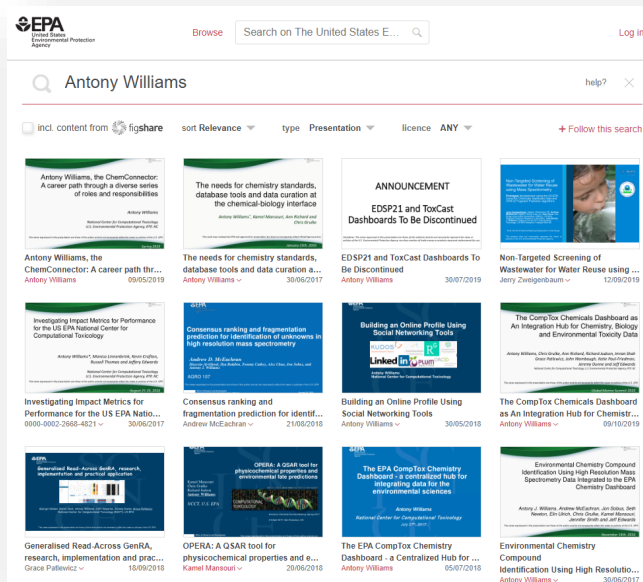


- Batch search for thousands of chemicals
- Real-time property and toxicity predictions
- Downloadable files – CSV, TSV and Excel

- The CompTox Chemistry Dashboard: a community data resource for environmental chemistry, *J. Cheminformatics*, **9**, 61 (2017)
- EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research, *Comp. Tox.* **12**, 100096 (2019)
- OPERA models for predicting physicochemical properties and environmental fate endpoints, *J. Cheminformatics*, **10**, 10 (2018)
- Screening Chemicals for Estrogen Receptor Bioactivity Using a Computational Model, *Environ. Sci. Technol.* **49**, 8804-8814 (2015)
- ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology, *Chem. Res. Toxicol.* **29**, 1225-51 (2016)
- Development and Validation of a Computational Model for Androgen Receptor Activity, *Chem. Res. Toxicol.* **30**, 946-964 (2017)
- CERAPP: Collaborative Estrogen Receptor Activity Prediction Project, *Environ. Health Perspect.* **124**, 1023 (2016)
- Abstract Sifter: a comprehensive front-end system to PubMed, *F1000*, **6**, 2164 (2017)

# You want to know more...

- Lots of resources available
  - Presentations: <https://tinyurl.com/w5hq55>
  - Communities of Practice Videos: <https://rb.gy/qsbn01>
  - Manual: <https://rb.gy/4fgydc>
  - Latest News: [https://comptox.epa.gov/dashboard/news\\_info](https://comptox.epa.gov/dashboard/news_info)

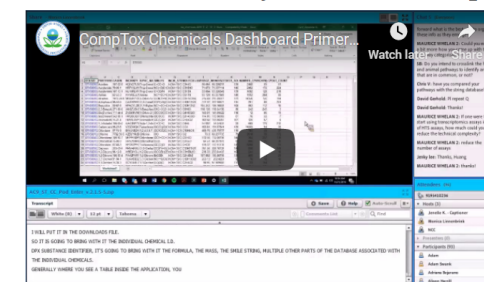


## CompTox Chemicals Dashboard primer videos

The CompTox Chemicals Dashboard is a one-stop-shop for chemistry, toxicity and exposure information for over 875,000 chemicals. Data and models within the Dashboard also help with efforts to identify chemicals of most need of further testing and reducing the use of animals in chemical testing.

Explore the wealth of data and features available in the CompTox Chemicals Dashboard with these instructional videos narrated by EPA scientists.

## General Chemistry and Search Capabilities





# Acknowledgments

- Contact: [Williams.Antony@epa.gov](mailto:Williams.Antony@epa.gov)
- Feedback and follow-up is welcomed! Your questions help
- The dashboard is based on the efforts of many more team members than us. Many collaborators provide data also.



EPA's Center for Computational Toxicology and Exposure