

# The CompTox Dashboard: Data and Tools to Support Chemical and Environmental Risk Assessment and the ENTACT project

*Antony Williams<sup>1</sup>, Chris Grulke<sup>1</sup>, Andrew McEachran<sup>2</sup>,  
Emma Schymanski<sup>3</sup>, Jon Sobus<sup>4</sup>, Elin Ulrich<sup>4</sup>, Ann Richard<sup>1</sup>,  
Jeremy Dunne<sup>1</sup> and Jeff Edwards<sup>1</sup>*

- 1) *National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC*
- 2) *Oak Ridge Institute of Science and Education (ORISE) Research Participant, RTP, NC*
- 3) *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg*
- 4) *National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, 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*

*August 2018  
ACS Fall Meeting, Boston*

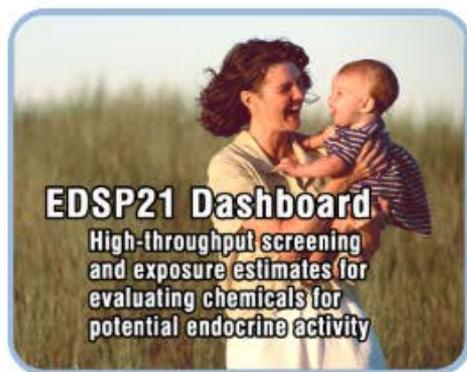
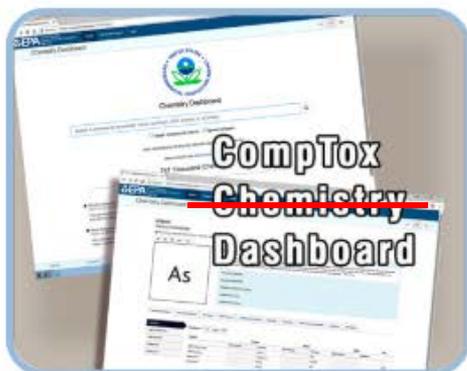
# Coming soon to a screen near you: The CompTox Portal

Environmental Topics

Laws & Regulations

About EPA

Search EPA.gov





- DSSTox was a concept in 2002
- Today it underpins the Dashboard

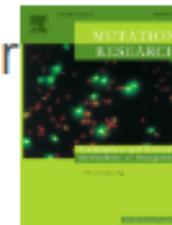


3



## Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis

Volume 499, Issue 1, 29 January 2002, Pages 27-52



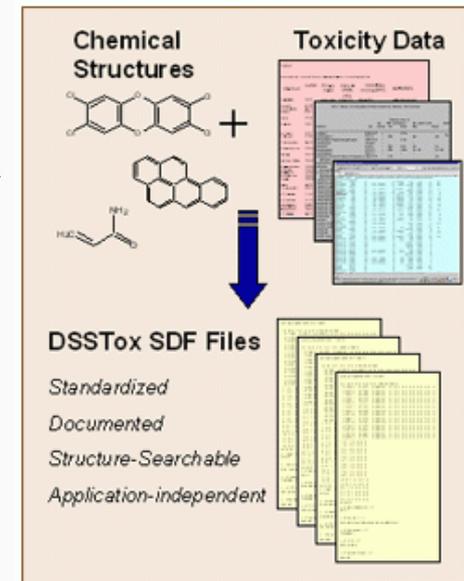
Mutation Research Frontiers

### Distributed structure-searchable toxicity (DSSTox) public database network: a proposal

Ann M. Richard <sup>a</sup>  , ClarLynda R. Williams <sup>a, b</sup>

## GOAL: Link chemical structures to data for SAR

- First release of data files in 2004
- Focused on high impact sets of data
  - Carcinogenic Potency Database
  - Drinking water disinfection by-products
  - EPA's Integrated Risk Information System
  - FDA's Maximum Daily Dose dataset
  - EPA's Fat Head Minnow Toxicity dataset
  - etc...



- Managed chemical registration for ToxCast/Tox21 chemicals (our HTS screening research)

# And then came Chris Grulke... ChemReg registration for DSSTox



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

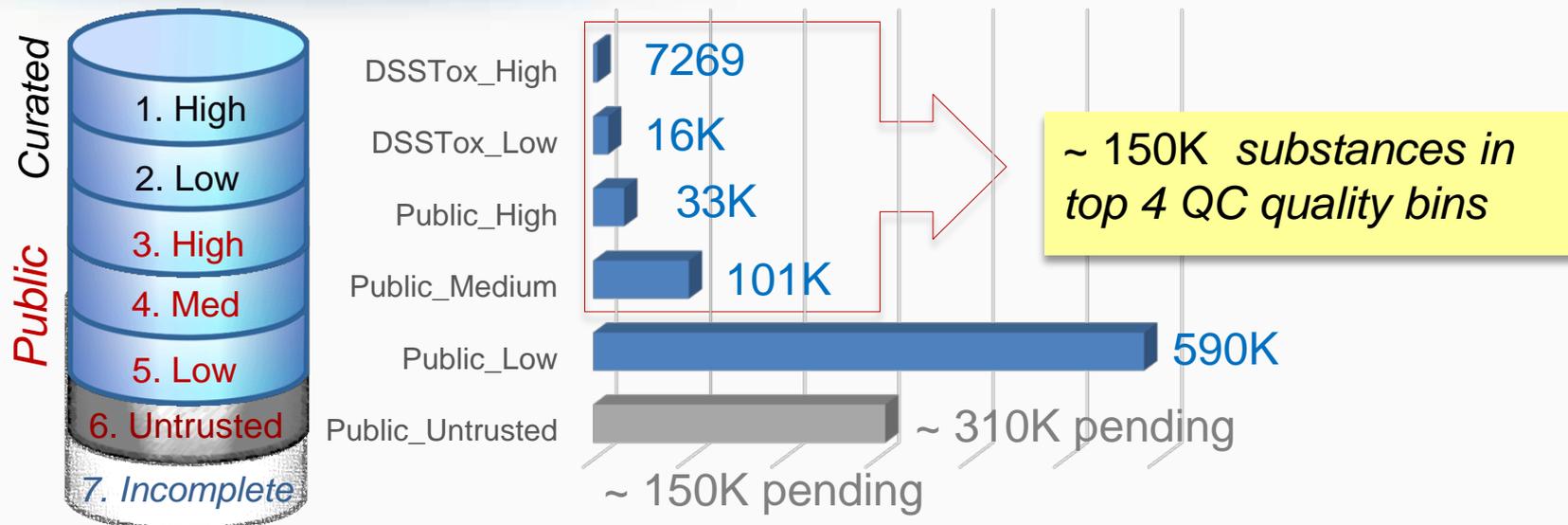
Valid license cannot be found

Calculate from Structure

Substance\_ID: DTXSID80198757  
CAS: 62885-41-0  
Name: 4-Hydroxy-3-methoxypyridine  
Substance Type: Single Compound  
QC Level: DSSTox\_High  
Data Source: STN(DSSTox)  
QC Notes: CAS [50700-60-2] assigned by DSSTox to pyridin-one tautomer form, which resolves to hydroxy form thru InChI

Compound\_ID: DTXCID40121248  
Chemical Shown: Tested Chemical  
Private Notes:  
Source of CAS-Compound: STN(DSSTox)  
Double Stereo: None  
Chiral Stereo: None  
Chemical Form: Organic  
Organic Form: Parent

# Distribution of curated data



## 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

# Internal List Conflicts: ChemReg List Curation

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

Welcome cgnulke

Editing Listname: ALANWOOD

Internal Check Results	
Description	Records
All	1718
Duplicate CASRN	2
Duplicate STRUCTURE, Duplicate STRUCTURE_INCHIKEY	2
Invalid casrn	2
NONE	1712

List: ALANWOOD

(1 of 69)   1 2 3 4 5 6 7 8 9 10   25

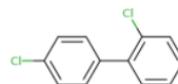
Record ID	External ID	1st Identifier	Warning
DTXRID303936283	(3-ethoxypropyl)mercury bromide	6012-84-6	NONE
Identifier			
(3-ethoxypropyl)mercury bromide			
6012-84-6			
InChI=1S/C5H11O.BrH.Hg/c1-3-5-6-4-2;;;/h1,3-5H2,2H3;1H;/q;+1/p-1			
BWUIOGHGUVLNSX-UHFFFAOYSA-M			
BWUIOGHGUVLNSX-UHFFFAOYSA-M			
DTXRID003936284	1,2-dichloropropane	78-87-5	NONE
DTXRID703936285	1,3-dichloropropene	542-75-6	NONE
DTXRID403936286	1-methylcyclopropene	3100-04-7	NONE
DTXRID103936287	1-naphthol	90-15-3	NONE
DTXRID803936288	2-(octylthio)ethanol	3547-33-9	NONE
DTXRID503936289	2,3,5-tri-iodobenzoic acid	88-82-4	NONE
DTXRID803936290	2,3,6-TBA	50-31-7	NONE
DTXRID503936291	2,4,5-T	93-76-5	NONE
DTXRID203936292	2,4,5-TB	93-80-1	NONE
DTXRID903936293	2,4-D	94-75-7	NONE

# Chemical Collections (e.g. PCBs)

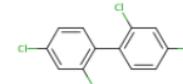
## Successor Substances (209)

CAS-RN	Relationship	Source
32774-16-6	is a Representative Isomer of this	
2051-60-7	is a Representative Isomer of this	
2051-61-8	is a Representative Isomer of this	
2051-62-9	is a Representative Isomer of this	
13029-08-8	is a Representative Isomer of this	
16605-91-7	is a Representative Isomer of this	
25569-80-6	is a Representative Isomer of this	
33284-50-3	is a Representative Isomer of this	
34883-43-7	is a Representative Isomer of this	
34883-39-1	is a Representative Isomer of this	
33146-45-1	is a Representative Isomer of this	
2050-67-1	is a Representative Isomer of this	
2974-92-7	is a Representative Isomer of this	
2974-90-5	is a Representative Isomer of this	
34883-41-5	is a Representative Isomer of this	Public
2050-68-2	is a Representative Isomer of this	Public

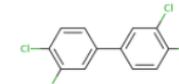
Download as:



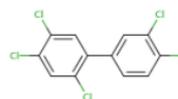
2,4-Dichlorobiphenyl  
34883-43-7



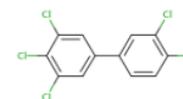
2,2',4,4'-Tetrachlorobiphenyl  
2437-79-8



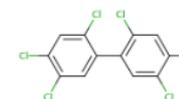
3,3',4,4'-Tetrachlorobiphenyl  
32598-13-3



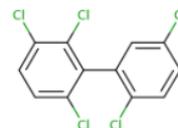
2,3,4,4',5-Pentachlorobiphenyl  
31508-00-8



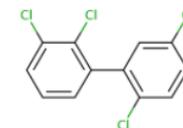
3,3',4,4',5-Pentachlorobiphenyl  
57465-28-8



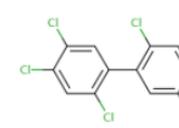
2,2',4,4',5,5'-Hexachlorobiphenyl  
35055-27-1



2,2',3,5',6-Pentachlorobiphenyl  
38379-99-8



2,2',3,5'-tetrachlorobiphenyl  
41464-39-5



2,2',4,5,5'-Pentachlorobiphenyl  
37880-73-2

### Related Chemicals

Found 209 chemicals

# The CompTox Dashboard

## General Overview and New Features 3.0

- Version 3.0 release, August 2018:
  - New user interface design
  - ~762,000 chemicals with related property data
  - New searches – product and use categories, assays and genes
  - More data – physchem, hazard data, bioactivity curves
  - Expanded number of chemical lists to review
  - Enhanced batch search capabilities
  - New chemical lists added (including PFAS chemical lists)
  - Handling of UVCB chemicals – many Markush



Chemicals Product/Use Categories Assay/Gene

762 Thousand Chemicals

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey

Identifier substring search

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

## Latest News

[Read more news](#)

### YouTube video regarding using the Dashboard for Non-Targeted Analysis

March 7th, 2018 at 9:43:36 AM

Ar  
Mar

A YouTube video discussing the application of the CompTox Chemistry Dashboard to support non-targeted analysis by mass spectrometry is available. This short video summarizes the advantages of the dashboard in terms of data quality and focused data set for environmental non-targeted analysis. [View it here on Youtube.](#)



### Discover.

[About/Disclaimer](#)  
[Accessibility](#)  
[Privacy](#)

### Connect.

[ACToR](#)  
[DSSTox](#)  
[Downloads](#)

### Ask.

[Contact](#)  
[Help](#)

# CompTox Dashboard Chemicals

762 Thousand Chemicals



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

Q Bisphenol A

-  Bisphenol A  
DTXSID7020182
-  Bisphenol A bis(2-hydroxyethyl ether) diacrylate  
DTXSID6066991
-  Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate  
DTXSID1066992
-  Bisphenol A bis(2-hydroxypropyl) ether  
DTXSID8051592
-  Bisphenol A carbonate polymer  
DTXSID6027840
-  Bisphenol A diglycidyl ether  
DTXSID6024624
-  Bisphenol A glycidyl methacrylate  
DTXSID7044841
-  Bisphenol A propoxylate diglycidyl ether  
DTXSID10399098
-  Bisphenol A propoxylate glycerolate diacrylate  
DTXSID40400126

# CompTox Dashboard Products and Use Categories



762 Thousand Chemicals



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

hair color

CPDat PRODUCT category: personal care hair color  
*hair colors and dyes characterized as permanent*

CPDat PRODUCT category: personal care hair color  
*hair colors and dyes characterized as for professional use*

CPDat PRODUCT category: personal care hair color  
*hair colors and dyes characterized as temporary*

CPDat PRODUCT category: personal care hair color  
*hair coloring products not otherwise categorized*

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

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

CPDat PRODUCT category: personal care hair color toner  
*chemical toners for hair coloring products*



## Discover.

[About/Disclaimer](#)  
[Accessibility](#)  
[Privacy](#)

## Connect.

[ACToR](#)  
[DSSTox](#)  
[Downloads](#)

## Ask.

[Contact](#)  
[Help](#)

# CompTox Dashboard Assays and Genes



762 Thousand Chemicals

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

Q estrogen

GENE: ESR1  
*estrogen receptor 1*

GENE: ESR2  
*estrogen receptor 2 (ER beta)*

GENE: ESRR A  
*estrogen-related receptor alpha*

GENE: ESRR B  
*estrogen-related receptor beta*

GENE: ESRR G  
*estrogen-related receptor gamma*

and curating data, major updates to the batch searching functionality and access to real time predictions for both physiochemical and toxicity endpoints. A [list of release notes](#) is available for your review. We look forward to your feedback.



## Discover.

[About/Disclaimer](#)  
[Accessibility](#)  
[Privacy](#)

## Connect.

[ACToR](#)  
[DSSTox](#)  
[Downloads](#)

## Ask.

[Contact](#)  
[Help](#)

# Detailed Chemical Pages Redesign

## DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▶ BIOACTIVITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

SYNONYMS

▶ LITERATURE

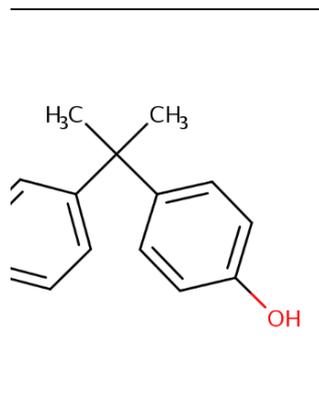
LINKS

COMMENTS

Bisphenol A

DTXSID7020182

US Tox Substance Id.



Batch Search Lists Predictions Downloads

Copy

Share

Submit Comment

Search all data

### Wikipedia

**Bisphenol A (BPA)** is an organic synthetic compound with the chemical formula  $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4\text{OH})_2$  belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water. It has been in commercial use since 1957.

BPA is a starting material for the synthesis of plastics, primarily

...  
[Read more](#)

### Intrinsic Properties

### Structural Identifiers

### Linked Substances

### Presence in Lists

### Record Information

### Quality Control Notes

# Physicochemical properties

Property

Summary

## Summary

Download

Columns

Search query

Property	Experimental average	Predicted average	Experimental median	Predicted median	Experimental range	Predicted range	Unit
LogP: Octanol-Water	3.32 (1)	3.29		3.43	3.32	2.40 to 3.64	
Melting Point	155 (7)	139	156	138	153 to 156	125 to 157	°C
Boiling Point	200 (1)	363		360	200	343 to 401	°C
Water Solubility	5.26e-4 (1)	9.64e-4		1.00e-3	5.26e-4	5.44e-4 to 1.31e-3	mol/L
Vapor Pressure	-	8.37e-7		3.43e-7	-	6.83e-8 to 2.59e-6	mmHg
Flash Point	-	190		190	-	188 to 192	°C
Surface Tension	-	46.0			-	46.0	dyn/cm
Index of Refraction	-	1.60			-	1.60	
Molar Refractivity	-	68.2			-	68.2	cm <sup>3</sup>
Polarizability	-	27.0			-	27.0	Å <sup>3</sup>
Density	-	1.17		1.17	-	1.14 to 1.20	g/cm <sup>3</sup>
Molar Volume	-	200			-	200	cm <sup>3</sup>
Thermal Conductivity	-	150			-	150	mW/(m <sup>2</sup> K)

# Detailed OPERA Prediction Reports

[Source](#)
[Result](#)
[Calculation Details](#)

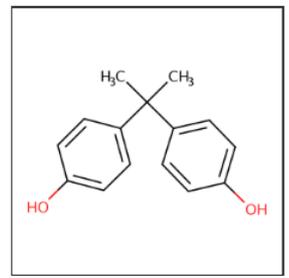
Experimental Values

- [PhysPropNCCT](#)
- Predicted Values
- [EPISUITE](#)
- [NICEATM](#)
- [ACD/Labs Conse](#)
- [ACD/Labs](#)
- [OPERA](#)

OPERA Models: LogP: Octanol-Water

**Bisphenol A**  
80-05-7 | DTXSID7020182

[Save PDF](#)



**Model Results**

Predicted value: 3.35

Global applicability domain: inside

Local applicability domain index: 0.88

Confidence level: 0.75

**Model Performance**

Weighted KNN model

[QMR](#)

5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.85	0.89	0.86	0.87	0.88	0.78

Nearest Neighbors from the Training Set

**Bisphenol A**  
Measured: 3.32  
Predicted: 3.35

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

**Flurbiprofen**  
Measured: 4.18  
Predicted: 3.83

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

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

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

# GenRA (Generalised Read-Across)

GenRA

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

Neighbors by: Chem: Morgan Fgrpts

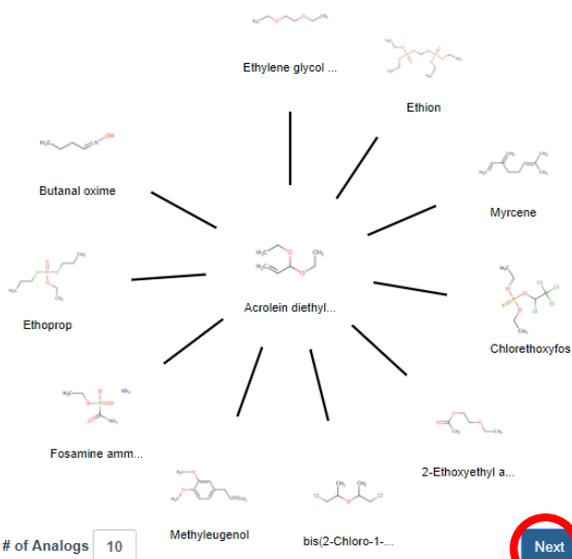
Filter by: invivo data

Summary Data Gap Analysis

Group: ToxRef

By: Tox Fingerprint

Generate Data Matrix



	bio_hx21	bio_hxct	chl_m_ct	tox_brf
Fluconazole	3	714	15	0
Hexaconazole	43	819	18	345
Flusilazole	28	819	9	345
Cyproconazole	14	819	16	408
Pyrasulfotole metabolite ...	0	0	18	234
Myclobutanil	15	818	15	345
Fenbuconazole	34	819	17	345
Tetraconazole	35	819	20	345
Metconazole	35	215	15	82
Ipconazole	46	232	16	180
Bromuconazole	24	277	13	345

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

Data gap analysis

# Access to Chemical Hazard Data

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT

## HAZARD

- ▶ ADME
- ▶ EXPOSURE
- ▶ BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- ▶ LITERATURE
- LINKS
- COMMENTS

### Data Type

Point of Departure

Download

Human Eco

Columns 10

Search query

More	Priority	Toxval type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Subsource	Source
	5	BMDL-10	-	chronic	0.609	mg/kg-day	human	-	mouse	EFSA CEF	EFSA
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECHA
	6	NOAEL	-	reproductive	3.75	mg/kg-day	reproductive	oral	rat	-	HPVIS
	5	NOEL	Systemic	repeat dose	3.75	mg/kg-day	repeat dose toxicity : oral	oral	rat	-	ECHA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECHA
	5	NOEL	Systemic	repeat dose	4.5	mg/kg-day	repeat dose toxicity : oral	oral	mouse	-	ECHA
	7	LEL	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB
	7	nel	-	chronic	5	mg/kg-day	reproductive multigeneration	oral	rat	open_lit	ToxRefDB
	5	NOAEL	-	chronic	5	mg/kg-day	human	-	mouse	EFSA AFC	EFSA
	7	nel	-	subchronic	5	mg/kg-day	subchronic	oral	rat	unpublished_submission	ToxRefDB

- ToxVal Database contains following data:
  - 30,050 chemicals
  - 772,721 toxicity values
  - 29 sources of data
  - 21,507 sub-sources
  - 4585 journals cited
  - 69,833 literature citations

# In Vitro Bioassay Screening

## ToxCast and Tox21

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

BIOACTIVITY

TOXCAST SUMMARY

PUBCHEM

TOXCAST DATA

TOXCAST MODELS

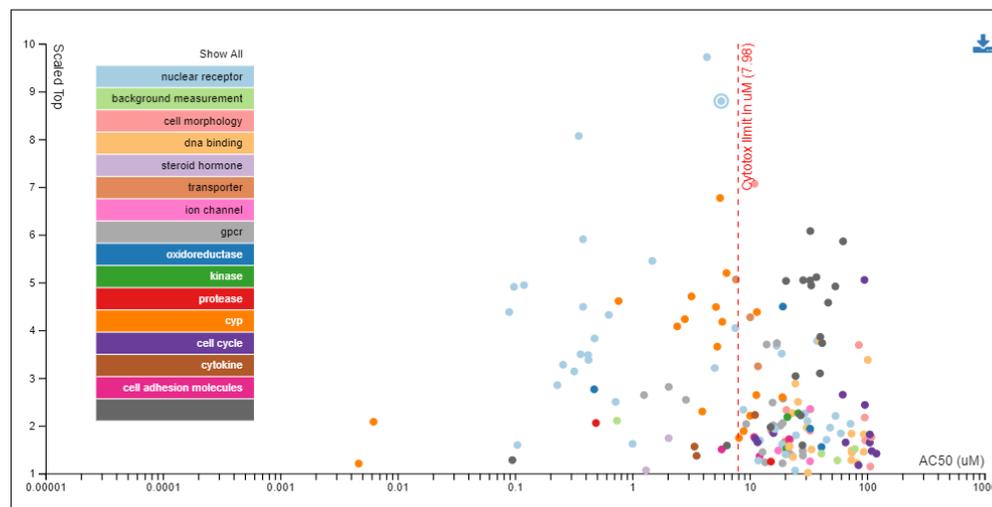
SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

### Chemical Activity Summary i

TOXCAST DATA



ASSAY DETAILS

AC50 (uM): 5.73  
Scaled top: 8.80  
Assay Endpoint Name: OT\_ER\_ERaEa\_0480  
Assay Description: 742  
Gene Symbol: ESR1  
Organism: human  
Tissue: kidney  
Assay Format Type: cell-based  
Biological Process Target: protein stabilization  
Detection Technology: Protein-fragment Complementation  
Analysis Direction: positive  
Intended Target Family: nuclear receptor  
Description: Data from the assay component OT\_ER\_ERaEa\_0480 was analyzed into 1 assay endpoint. This assay endpoint, OT\_ER\_ERaEa\_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 relateable targets, this assay endpoint is annotated to the 'nuclear receptor' intended target family, where the subfamily is 'steroidal'.

# In Vitro Bioassay Screening

## ToxCast and Tox21

### Bisphenol A

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

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

▶ ADME

▶ EXPOSURE

▼ BIOACTIVITY

TOXCAST: SUMMARY

PUBCHEM

TOXCAST: DATA

TOXCAST: MODELS

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

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

Assay Selection 1 Selected

A Single Assay Can Have Multiple Charts

Number of Charts: 6

Active  Inactive  All

Filter

Filter assays

Assay Set: ER (1 of 18 Selected)

ACEA\_T47D\_80hr\_Positive

ATG\_ERE\_CIS\_up

ATG\_ERa\_TRANS\_up

NVS\_NR\_bER

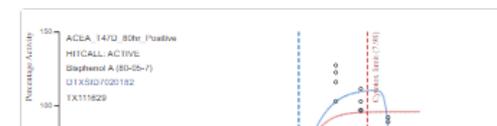
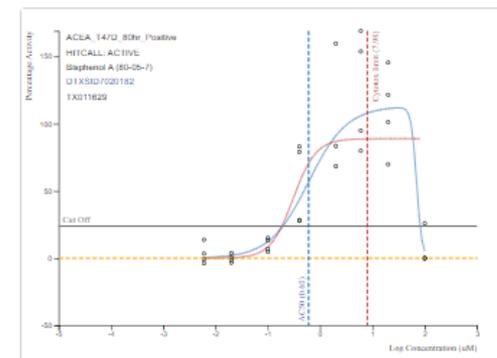
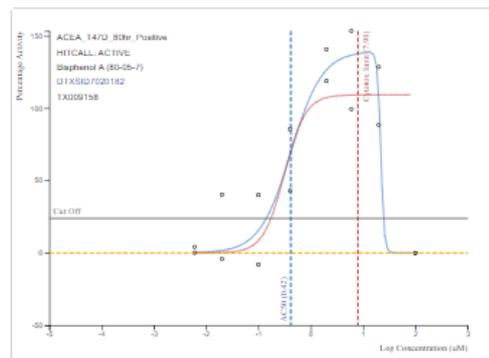
NVS\_NR\_hER

NVS\_NR\_mERa

OT\_ER\_ERaERa\_0480

OT\_ER\_ERaERa\_1440

OT\_ER\_ERaERb\_0480



# Sources of Exposure to Chemicals

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME

Download

Columns 10

### Product and Use Categories (PUCs) i

Search query

Product or Use Categorization	Categorization type	Number of Unique Products
manufacturing, metals	CPCat Cassette	17
adhesive	CPCat Cassette	17
	CPCat Cassette	16
	CPCat Cassette	12
	CPCat Cassette	11
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	8
	CPCat Cassette	7
	CPCat Cassette	6

### EXPOSURE

- PRODUCT & USE CATEGORIES**
- CHEMICAL WEIGHT FRACTION
- CHEMICAL FUNCTIONAL USE
- TOXICS RELEASE INVENTORY
- MONITORING DATA
- EXPOSURE PREDICTIONS
- PRODUCTION VOLUME

# Sources of Exposure to Chemicals

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

HAZARD

ADME

EXPOSURE

PRODUCT & USE CATEGORIES

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

**TOXICS RELEASE INVENTORY**

MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

BIOACTIVITY

## Toxics Release Inventory

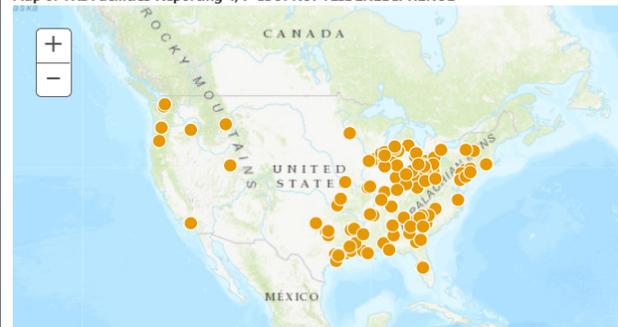
Print Page

### 2015 TRI Factsheet: Chemical - 4,4'-ISOPROPYLIDENEDIPHENOL, 000080057

Data Source: 2016 Dataset (released March 2018)

The [Toxics Release Inventory \(TRI\)](#) tracks the management of certain toxic chemicals that may pose a threat to human health and the environment. Certain industrial facilities in the U.S. must report annually how much of each chemical is recycled, combusted for energy recovery, treated for destruction, and disposed of or otherwise released on- and off-site. This information is collectively referred to as production-related waste managed.

#### Map of TRI Facilities Reporting 4,4'-ISOPROPYLIDENEDIPHENOL



#### Quick Facts for 2015

	Chemical	United States
<b>Number of TRI Facilities:</b>	120	22,130
<b>Total Production-Related Waste Managed:</b>	15.8 million lbs	27.1 billion lbs
<b>Total On-site and Off-site Disposal or Other Releases:</b>	2.5 million lbs	3.4 billion lbs
<b>Total On-site:</b>	39.4 thousand lbs	2.9 billion lbs
• Air:	28.7 thousand lbs	686.4 million lbs
• Water:	4.4 thousand lbs	198.2 million lbs
• Land:	6.2 thousand lbs	2.0 billion lbs

# Identifiers to Support Searches

## 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

25

Search query

**Synonym**

**Quality**

**Bisphenol A**

Valid

4,4'-(Propane-2,2-diyldiphenol

Valid

Phenol, 4,4'-(1-methylethylidene)bis-

Valid

80-05-7 Active CAS-RM

Valid

**BPA**

Valid

4,4'-Propane-2,2-diyldiphenol

Valid

Phenol, 4,4'-(1-methylethylidene)bis-

Valid

4-06-00-06717 Bellstein Registry Number

Bellstein

(4,4'-Dihydroxydiphenyl)dimethylmethane

Good

2,2-Bis(4'-hydroxyphenyl)propane

Good

2,2'-Bis(4-hydroxyphenyl)propane

Good

2,2-BIS-(4-HYDROXY-PHENYL)-PROPANE

Good

2,2-Bis(4-hydroxyphenyl)propane

Good

2,2-Bis(p-hydroxyphenyl)propane

Good

2,2-Di(4-Hydroxyphenyl) Propane

Good

# Literature Searches and Links

## Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. FATE/TRANSPORT

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

Optionally, edit the query before retrieving.

"80-05-7" OR "Bisphenol A"

### LITERATURE

GOOGLE SCHOLAR

PUBMED ABSTRACT SIFTER

PUBCHEM ARTICLES

PUBCHEM PATENTS

PPRTV

IRIS

# Literature Searches and Links

Chemical Properties | Env. Fate/Transport | Hazard | ADME (Beta) | Exposure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

Google Scholar  
**PubMed Abstract Sifter**  
 PubChem Articles  
 PubChem Patents  
 PPRTV  
 IRIS

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

Hazard  Retrieve Articles **i**

13 of 13 articles loaded...

Optionally, edit the query before retrieving.  
 ("57-55-6" OR "1,2-Propylene glycol" OR "Propylene Glycol") AND (NOAEL or NOEL OR LOEL or Rfd OR "reference dose" OR "reference concentration" OR "adverse effect level"[tiab] OR "cancer slope factor"[tiab])

To find articles quickly, enter terms to sift abstracts. **i**

Download / Send to... **i** Download Sifter for Excel **i**

<input type="checkbox"/>	PMID	Year	Title	Authors	Journal	Rev
<input type="checkbox"/>	27101543	2016	Electronic cigarettes: a systematic review of available studies on hea...	Zulkifli; Abidin; Abidin; Amer Nordin; Praveena; Sye...	Reviews on environmental health	
<input type="checkbox"/>	26787428	2016	Toxicological assessment of a prototype e-cigaret device and three fl...	Werley; Kirkpatrick; Oldham; Jerome; Langston; Lill...	Inhalation toxicology	
<input type="checkbox"/>	26475513	2015	Deriving Biomonitoring Equivalents for selected E- and P-series glyco...	Poet; Ball; Hays	International journal of hygiene and environmental h...	
<input type="checkbox"/>	26120296	2015	Potential harmful health effects of inhaling nicotine-free shisha-pen v...	Kienhuis; Soeteman-Hernandez; Bos; Cremers; Kle...	Tobacco induced diseases	
<input type="checkbox"/>	25527861	2014	Efinaconazole: Developmental and reproductive toxicity potential of ...	Glynn; Jo; Minowa; Sanada; Nejishima; Matsuuchi; ...	Reproductive toxicology (Elmsford, N.Y.)	
<input type="checkbox"/>	25038564	2014	Nonclinical safety assessment of Efinaconazole Solution (10%) for o...	Jo; Glynn; Nejishima; Sanada; Minowa; Calvarese; ...	Regulatory toxicology and pharmacology : RTP	
<input type="checkbox"/>	24138296	2013	Solvent-based formulations for intravenous mouse pharmacokinetic ...	Thackaberry; Wang; Schweiger; Messick; Valle; De...	Xenobiotics: the fate of foreign compounds in biolog...	
<input type="checkbox"/>	21683116	2011	Non-clinical safety and pharmacokinetic evaluations of propylene gly...	Werley; McDonald; Lilly; Kirkpatrick; Wallery; Byron;...	Toxicology	
<input type="checkbox"/>	18830862	2008	Final report on the safety assessment of methoxyisopropanol and m...		International journal of toxicology	
<input type="checkbox"/>	15876203	2005	Using physiologically-based pharmacokinetic modeling to address n...	Kiman; Sweeney; Corley; Gargas	Risk analysis : an official publication of the Society f...	
<input type="checkbox"/>	12583407	2003	Significance of 2-methoxypropionic acid formed from beta-propylene...	Carney; Pottenger; Johnson; Liberacki; Tomesi; Dry...	Toxicological sciences : an official journal of the Soc...	

ptox.zn.epa.gov/dashboard/dsstoxdb/results

# How can we curate our data?

- Crowdsourcing is well proven nowadays
- Comments can be added at a record level



- Submitted comments are reviewed by administrators and responded to

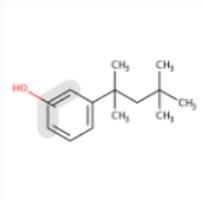
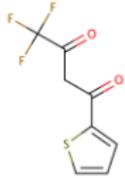
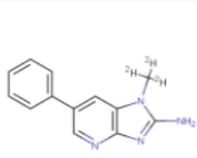
# Public Crowdsourced Comments

[https://comptox.epa.gov/dashboard/comments/public\\_index](https://comptox.epa.gov/dashboard/comments/public_index)

## Crowdsourced Comments

Show 10 entries

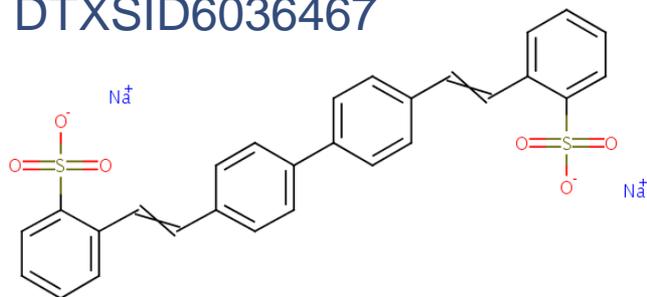
Search:

Chemical	Structure	Date	Comment	Status
<a href="#">(1,1,3,3-Tetramethylbutyl)phenol</a>		2017-07-15	Octylphenol redirects here, yet the name and related chemicals are 1,1,3,3-tetramethylbutylphenol - which is only a subset of all octylphenol isomers? Is this CAS only for these alkyl isomers?	★
<a href="#">1,3-Butanedione, 4,4,4-trifluoro-1-(2-thienyl)-</a>		2017-03-30	Synonym: TTFA (Any way to bank these reCAPTCHAs so I don't have to do it everytime?)	★
<a href="#">1-(2H&lt;sup&gt;2&lt;/sup&gt;H&lt;sub&gt;3&lt;/sub&gt;)-Methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine</a>		2017-05-06	1-(2H <sup>2</sup> H <sub>3</sub> )Methyl-6-phenyl-1H-imidazo[4,5-b]pyridin-2-amine 210049-13-1   DTXSID70670097 contains an error in the empirical formula due to an error in the deuterium representation and subsequent counting	★

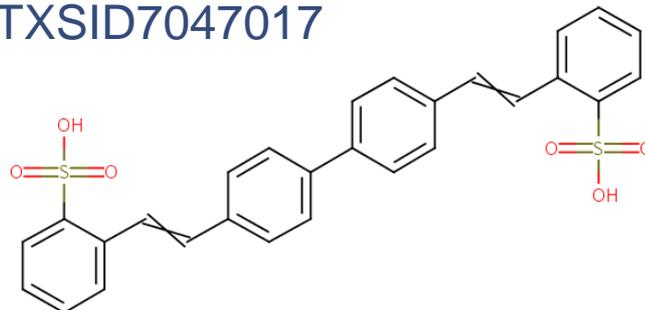
# MassBank/CompTox Curation of External Data

- A “nice” example: 4-4'-Bis(2sulfostyryl)biphenyl

Purchased: CAS: 27344-41-8  
DTXSID6036467



Registered: CAS: 38775-22-3 (UFZ)  
DTXSID7047017



Comment from structure source: to my knowledge the stilbene-derived fluorescent whitening agents are all trans (E) isomers, as the cis (Z) isomers are not fluorescent (although they might undergo photo-isomerisation to the cis isomers under UV light, and clothing gets yellowish again then...). Thus I would consider the E,E form the correct one, although I don't know whether it is synthesized in a way that x % of the technical product are actually the inactive E,Z (dunno if this is still active?) or Z,Z forms and thus the undefined stereo would be correct. In the environment E,E is partly transformed to E,Z (maybe also Z,Z) and thus both isomers occur. see: <http://pubs.acs.org/doi/pdf/10.1021/es960748a> CAS number on record was 27344-41-8 which is DTXSID6036467.

Add A Comment

Related to comment just added to DTXSID7047017 - it would be useful for us to know if this CAS number is undefined stereochemistry (a technical mix) or defined ... we would update our records accordingly.

User comment posted 4 months ago

ed to the mixed E/Z-form of the chemical. The E/E form is 6036467 is ALSO the mixed EZ form

I have a stereo-defined version of this on record as Fluorescent brightener 351 c1cc(cc1)S(=O)(=O)C=C/c2ccc(cc2)c3ccc(cc3)C=C/c4c(cccc4)S(=O)(=O)O SQAQVQFOMMLRPR-IWGRKNQJSA-N

User comment posted 4 months ago

The CAS Number is indeed for the undefined stereochemistry form.

Admin reply posted 4 months ago

There is NO indication of the E-defined stereochemistry in the registry and it appears to be a mix of isomers.

Admin reply posted 4 months ago



- The ENTACT project has driven specific functionality Mass Spectrometry needs
  - Specific searches within the dashboard
  - Mappings between chemicals in the database
  - Making use of specific representations for UVCBs
  - Addition of specific lists of chemicals
  - Collaborative cross-linking between sites
  - Research into best approaches for candidate ranking

# Mass and Formula Searches Supporting Mass Spectrometry

## Advanced Search

### Mass Search

**Select Adduct:**  ▼

Da

### Molecular Formula Search

MS Ready Formula 

Exact Formula 

### Generate Molecular Formula(e)

Da

Default Options: C[1-50] H[0-100] O[0-20] N[0-20] P[0-20] S[0-10]

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

# The Dashboard to Support MS-Analysis

## MS-Ready Structures Underpin Analysis

**Mass Search** ⓘ

± Min/Max M

Mass Da ± Error Da ppm

**Molecular Formula Search** ⓘ

Molecular Formula  MS Ready Formula ⓘ  Exact Formula ⓘ

**Generate Molecular Formula(e)** ⓘ

± Min/Max

Mass Da

Default Options: C[1-50] H[0-100] O[0-20] N[0-20]  
Include Halogens:  F[0-20]  Cl[0-20]  Br[0-20]

Options ▾

**Select Input Type(s)**

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

Step One Step Two Step Three Step Four Step Five Step Six

**Step Five: Choose Data Fields to Download**

Please enter one identifier per line

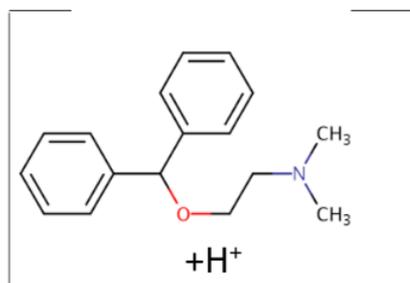
Enter Identifiers to Search (searches should be limited to <1000 identifiers)

C14H22N2O3  
C10H12N2O  
C14H18N4O3  
C12H11N7  
C8H9NO2

Display All Chemicals Download Chemical Data

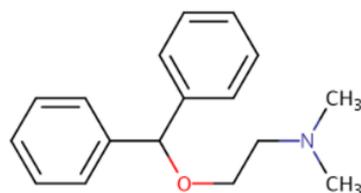
# Specific Data-Mappings “MS-Ready Structures”

A) Molecular Ion



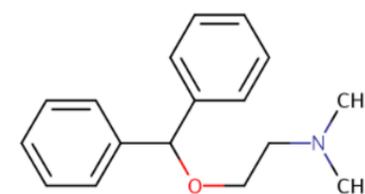
$m/z \approx 256.1702$

B) MS-Ready Form

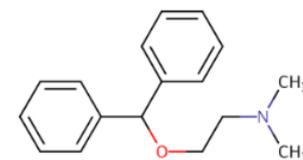


monoisotopic mass= 255.1623  
 $C_{17}H_{21}NO$   
DTXCID802949

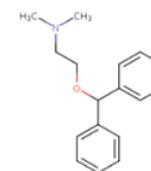
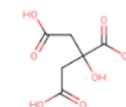
C) Mappings from MS-Ready



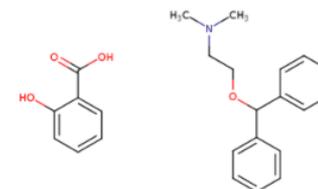
Diphenhydramine  
 $C_{17}H_{21}NO$  | 255.1623  
DTXSID4022949



Diphenhydramine  
hydrochloride  
 $C_{17}H_{22}ClNO$  | 291.1390  
DTXSID4020537



Diphenhydramine citrate  
 $C_{23}H_{29}NO_8$  | 447.1893  
DTXSID80237211



Diphenhydramine salicylate  
 $C_{24}H_{27}NO_4$  | 393.1940  
DTXSID10225883

- Already through peer review

**Title: “MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies**

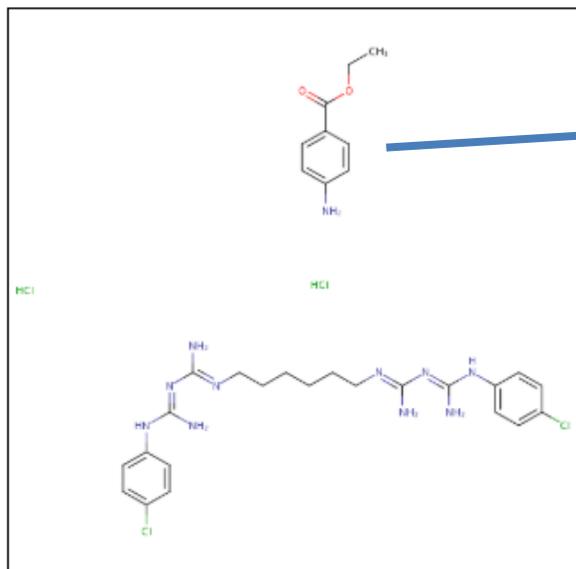
Andrew D. McEachran<sup>1,2\*</sup>, Kamel Mansouri<sup>1,2,3</sup>, Chris Grulke<sup>2</sup>, Emma L. Schymanski<sup>4</sup>, Christoph Ruttkies<sup>5</sup> and Antony J. Williams<sup>2\*</sup>

# MS-Ready Mappings

## Progaron

108532-15-6 | DTXSID20148579

Searched by DSSTox Substance Id.



Intrinsic Properties

Structural Identifiers

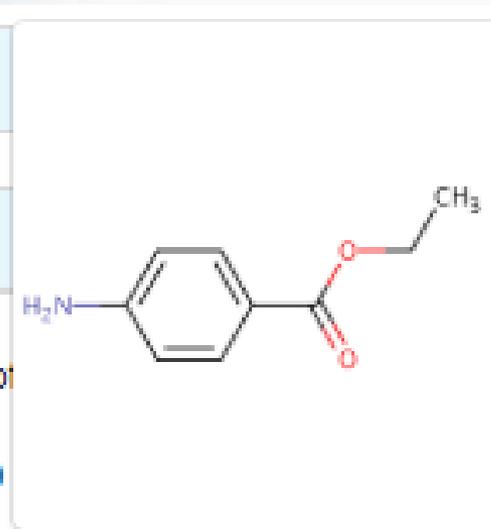
Linked Substances

Same Connectivity: 1 record (based on)

Mixtures, Components and Neutralizer:

MS-Ready Mappings: DTXCID0013314;

Similar Compounds: 0 records



[DTXCID301804:11 records;](#)

Presence in Lists

Record Information

Quality Control Notes

# MS-Ready Mappings Set

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

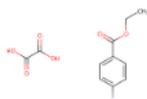
9 of 11 chemicals visible

Download / Send

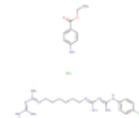
Show info: DTXSID CASRN Select all

Sort by: DTXSID

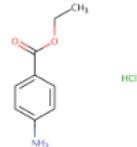
Filter by: Name or CASRN Isotopes



Anesthesine oxalate  
DTXSID: DTXSID20148337  
CASRN: 107948-47-0



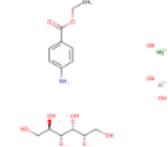
Progaron  
DTXSID: DTXSID20148879  
CASRN: 108532-15-8



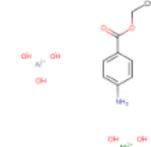
Benzocaine hydrochloride  
DTXSID: DTXSID50177812  
CASRN: 23239-88-5



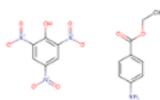
Anesthesine succinate  
DTXSID: DTXSID60148336  
CASRN: 107948-46-9



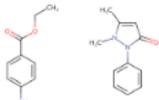
Almagel A-neo  
DTXSID: DTXSID80227559  
CASRN: 76741-92-9



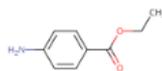
Almagel  
DTXSID: DTXSID70227560  
CASRN: 76741-95-2



Ethyl 4-aminobenzoate-2,4,6-trinitrophenyl  
DTXSID: DTXSID70787033  
CASRN: 5982-70-7



Antipyrine mixture with benzocaine  
DTXSID: DTXSID80212886  
CASRN: 63448-01-1



Benzocaine  
DTXSID: DTXSID8021804  
CASRN: 94-09-7

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

## Molecular Formula Search

 MS Ready Formula  Exact Formula 

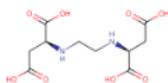
3 of 3 chemicals visible

Show info: "/> "/> 

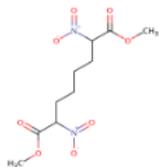
Sort by: "/> 

Filter by:

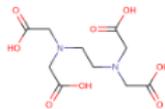




N,N'-Ethylenedi-L-aspartic acid  
DTXSID: DTXSID1051852  
CASRN: 20846-91-7



Dimethyl 2,7-dinitrooctanedioate  
DTXSID: DTXSID20468864  
CASRN: 67404-09-5



Ethylenediaminetetraacetic acid  
DTXSID: DTXSID6022977  
CASRN: 60-00-4

# MS-Ready Mappings

- **Same Input Formula: C<sub>10</sub>H<sub>16</sub>N<sub>2</sub>O<sub>8</sub>**
- **MS Ready Formula Search: 93 Chemicals**

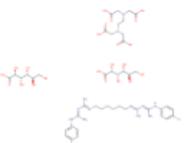
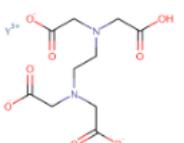
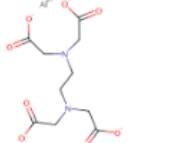
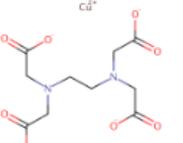
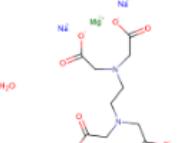
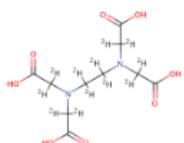
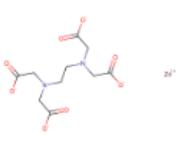
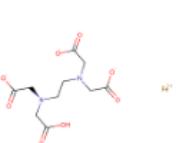
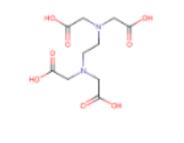
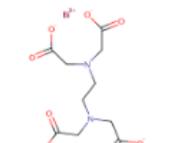
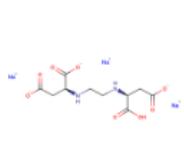
93 chemicals

Download / Send

Sort by: DTXSID

Show info:

Filter by:

 <p>Trisidine DTXSID: DTXSID00153984 CASRN: 123354-94-9</p>	 <p>Acetic acid, (ethylenedinitrilo)tetra-, yttrium salt DTXSID: DTXSID00154799 CASRN: 12558-71-3</p>	 <p>Acetic acid, (ethylenedinitrilo)tetra-, aluminum salt DTXSID: DTXSID00183708 CASRN: 29507-62-8</p>	 <p>EDTA, copper salt DTXSID: DTXSID0034564 CASRN: 12276-01-6</p>	 <p>Magnesium sodium 2,2',2'',2'''-(ethane-1,2-diyldinitrilo)tetraacetate DTXSID: DTXSID00583348 CASRN: 29932-54-5</p>	 <p>2,2',2'',2'''-[(2,2,2-trifluoroethylideneamino)ethane-1,2-diyldinitrilo]tetraacetate DTXSID: DTXSID00583949 CASRN: 203806-08-0</p>
 <p>Zincate(2-), [[N,N'-1,2-ethanediy]bis(N-carboxymethyl)ammonium]zinc salt DTXSID: DTXSID0065696 CASRN: 14025-21-9</p>	 <p>Ferrate(1-), [[N,N'-1,2-ethanediy]bis(N-carboxymethyl)ammonium]ferrate salt DTXSID: DTXSID0066163 CASRN: 17099-81-9</p>	 <p>PUBCHEM_54611985 DTXSID: DTXSID00715445 CASRN: 22239-30-1</p>	 <p>Glycine, N,N'-1,2-ethanediybis(N-carboxymethyl)ammonium salt DTXSID: DTXSID10236595 CASRN: 87731-78-0</p>	 <p>Bismuth Sodium Ethylenediaminetetraacetate DTXSID: DTXSID10437000 CASRN: 12558-49-5</p>	 <p>N,N'-Ethylenedi-(L-aspartic acid) trisodium salt DTXSID: DTXSID1051806 CASRN: 178949-82-1</p>

- 93 chemicals returned in total
  - Only 7 of the 93 are single component chemicals
  - 3 are neutral compounds and 1 is charged

# Advanced Searches

## Mass Search: Tile View

### Mass Search

Da

### Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

298 of 298 chemicals visible

Download / Send 

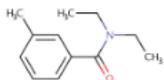
Show info:     

Select all 

Sort by:   

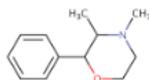
Filter by:





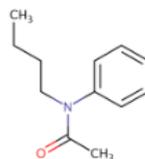
DEET

DTXSID: DTXSID2021995  
CASRN: 134-62-3  
TOXCAST: 14/663  
Mass Diff: 0.000014



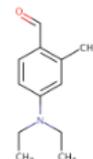
Phendimetrazine

DTXSID: DTXSID1023447  
CASRN: 634-03-7  
TOXCAST: 0  
Mass Diff: 0.000014



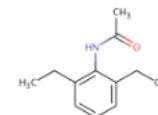
N-Butylacetanilide

DTXSID: DTXSID2042197  
CASRN: 91-49-6  
TOXCAST: 0  
Mass Diff: 0.000014



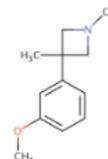
Benzaldehyde, 4-(diethylamino)-2-methyl-

DTXSID: DTXSID4059041  
CASRN: 92-14-8  
TOXCAST: 0  
Mass Diff: 0.000014



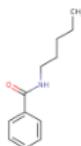
Acetanilide, 2',6'-diethyl-

DTXSID: DTXSID90168148  
CASRN: 16665-89-7  
TOXCAST: 0  
Mass Diff: 0.000014



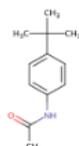
Azetidine, 1,3-dimethyl-3-(m-methoxyphenyl)-

DTXSID: DTXSID40173560  
CASRN: 19832-26-9  
TOXCAST: 0  
Mass Diff: 0.000014



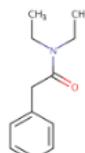
Benzamide, N-pentyl-

DTXSID: DTXSID20174198  
CASRN: 20308-43-4



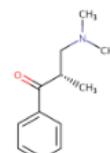
p-t-Butylacetanilide

DTXSID: DTXSID80174238  
CASRN: 20330-45-4



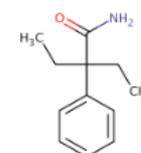
N,N-Diethylphenylacetamide

DTXSID: DTXSID00179048  
CASRN: 2431-98-1



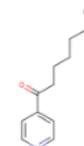
3-(Dimethylamino)-2-methylpropiofenone

DTXSID: DTXSID80180796  
CASRN: 26171-60-6



Butyramide, 2-ethyl-2-phenyl-

DTXSID: DTXSID60184653  
CASRN: 30568-30-0



1-Heptanone, 1-(4-pyridyl)-

DTXSID: DTXSID40188594  
CASRN: 32941-30-3

# Advanced Searches

## Mass Search: Table View

### Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

298 of 298 chemicals visible

Download / Send

Select all

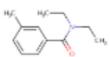
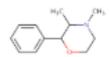
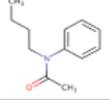
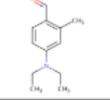
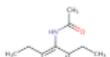


Sort by: Mass Difference



Filter by: Name or CASRN

Multicomponent Chemicals

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass	Mass Difference
	DTXSID2021995 <a href="#">ToxCast™</a>	DEET	134-82-3	Level 1	111	111	155	753	191.131014	0.000014
	DTXSID1023447	Phendimetrazine	634-03-7	Level 2	12	28	35	50	191.131014	0.000014
	DTXSID2042197	N-Butylacetanilide	91-49-6	Level 2	1	26	50	1	191.131014	0.000014
	DTXSID4059041	Benzaldehyde, 4-(diethylamino)-2-methyl-	92-14-8	Level 3	0	7	51	0	191.131014	0.000014
	DTXSID90168148	Acetanilide, 2,6'-diethyl-	16665-89-7	Level 4	0	4	33	0	191.131014	0.000014

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

## Batch Search



### Step Three: Select Download Data or Display Chemicals

Please enter one identifier per line 

#### Select Input Type(s)

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

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

```
126-98-7
78-85-3
598-31-2
78-95-5
108-05-4
79-21-0
3282-30-2
141-59-3
75-86-5
78-82-0
```

Chemical Data

# Batch Searching

Select Output Format:

Excel

Download

### Customize Results

- Select All
- Select All in Lists

### Chemical Identifiers

- DTXSID *i*
- Chemical Name *i*
- CAS-RN *i*
- InChIKey *i*
- IUPAC Name *i*

### Structures

- Mol File *i*
- SMILES *i*
- InChI String *i*
- MS-Ready SMILES *i*
- QSAR-Ready SMILES *i*

### Intrinsic And Predicted Properties

- Molecular Formula *i*
- Average Mass *i*
- Monoisotopic Mass *i*
- TEST Model Predictions *i*
- QSAR Model Predictions *i*

- Presence in Lists:**
- Acute exposure guideline levels
  - Algal Toxins
  - Androgen Receptor Chemicals
  - ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances
  - ATSDR Toxic Substances Portal Chemical List
  - Bisphenol Compounds
  - California Office of Environmental Health Hazard Assessment
  - Drinking Water Suspects, KWR Water, Netherlands
  - Endocrine Disruption Screening Program (EDSP) Universe of Chemicals
  - EPA Chemicals associated with hydraulic fracturing
  - EPA Consumer Products Suspect Screening Results
  - EPA Integrated Risk Information System (IRIS)
  - EPAHFR - EPA Chemicals associated with hydraulic fracturing
  - EU Cosmetic Ingredients Inventory (Combined 2000/2006)
  - French Monitoring List
  - HERO: Health and Environmental Research Online
  - ICCVAM in vitro cytotoxicity test methods
  - ICCVAM local lymph node assay (LLNA) from NIEHS
  - ICCVAM Skin Corrosion 2004 collection from NIEHS
  - ICCVAM test method evaluation report: in vitro ocular toxicity test methods
  - ITN ANTIBIOTIC LIST

70 Lists and growing

# Excel Output

INPUT	FOUND_BY	DTXCID_IN	DATA_SQ	TOXVAL_D	TOXCAST	TOXCAST	NUMBER_C	PUBCHEM	STO
C6H12O3	MS Ready	<a href="#">DTXCID701:</a>	51	Y	0.36	2/562	24	83	Y
C6H12O3	MS Ready	<a href="#">DTXCID003:</a>	67	Y	0.36	1/276	376	80	Y
C6H12O3	MS Ready	<a href="#">DTXCID106:</a>	65	Y	4.42	5/113	6	77	Y
C6H12O3	MS Ready	<a href="#">DTXCID105:</a>	45	Y	0.0	0/163	3	94	-
C6H12O3	MS Ready	<a href="#">DTXCID901:</a>	38	Y	-	-	14	110	Y
C6H12O3	MS Ready	<a href="#">DTXCID402:</a>	34	Y	0.0	0/113	-	53	Y
C6H12O3	MS Ready	<a href="#">DTXCID202:</a>	31	Y	-	-	-	36	Y
C6H12O3	MS Ready	<a href="#">DTXCID202:</a>	30	-	2.54	7/276	-	54	-
C6H12O3	MS Ready	<a href="#">DTXCID109:</a>	26	Y	-	-	-	46	-
C6H12O3	MS Ready	<a href="#">DTXCID202:</a>	24	Y	0.0	0/113	-	47	-
C6H12O3	MS Ready	<a href="#">DTXCID303:</a>	22	Y	-	-	-	89	-
C6H12O3	MS Ready	<a href="#">DTXCID302:</a>	20	Y	-	-	2	25	Y
C6H12O3	MS Ready	<a href="#">DTXCID407:</a>	19	Y	-	-	12	62	-
C6H12O3	MS Ready	<a href="#">DTXCID704:</a>	17	Y	-	-	-	64	-
C6H12O3	MS Ready	<a href="#">DTXCID704:</a>	16	Y	-	-	3	49	-

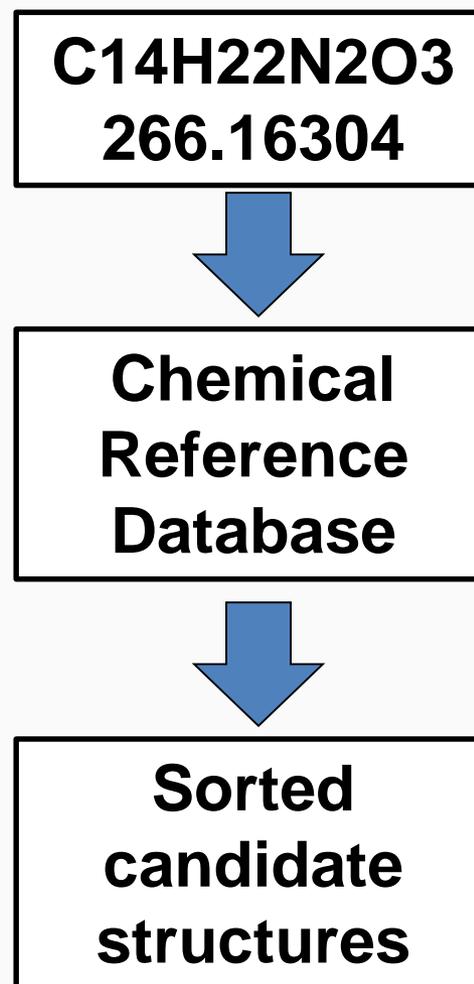
# Searching batches

## Formula (or mass) searching

	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3	338.116398	19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23ClN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S	406.213757997	35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35ClN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35ClN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O	176.094963014	40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13ClN2O	212.0716407	11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxyimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026	13140-86-8	N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3	290.137890456	68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3	290.137890456	51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19ClN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2-	C14H18N4O3	290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quin	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid-6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6025567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	50

# Data Source Ranking for Identification in SSA/NTA

- Mass and/or formula unknown to a researcher, contained within a reference database
- Most likely candidate chemicals have the most references/sources



# Comparing Dashboard with ChemSpider

Anal Bioanal Chem  
DOI 10.1007/s00216-016-0139-z



RAPID COMMUNICATION

## Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard

Andrew D. McEachran<sup>1</sup> · Jon R. Sobus<sup>2</sup> · Antony J. Williams<sup>3</sup>

- On same 162 chemicals, Dashboard outperforms ChemSpider

	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2 <sup>a</sup>	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

<sup>a</sup> Average rank in ChemSpider shown here does not include an outlier where the rank was 201, when added the average rank position is 3.5

- US EPA CompTox Dashboard Data Sources (DS)
-  Data Source Count
-  Reference Count
- Presence in  Database
- Predicted Environmental Media Occurrence
- CPDat Product Occurrence Count
- OPERA PhysChem Properties
- NORMAN Network Priority List

# All available via Batch Search

Excel

Download

### Customize Results

- Select All
- Select All in Lists

### Chemical Identifiers

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

### Structures

- Mol File
- SMILES
- InChI String
- MS-Ready SMILES

### 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
- Include links to ACToR reports - SLOW! (BETA)

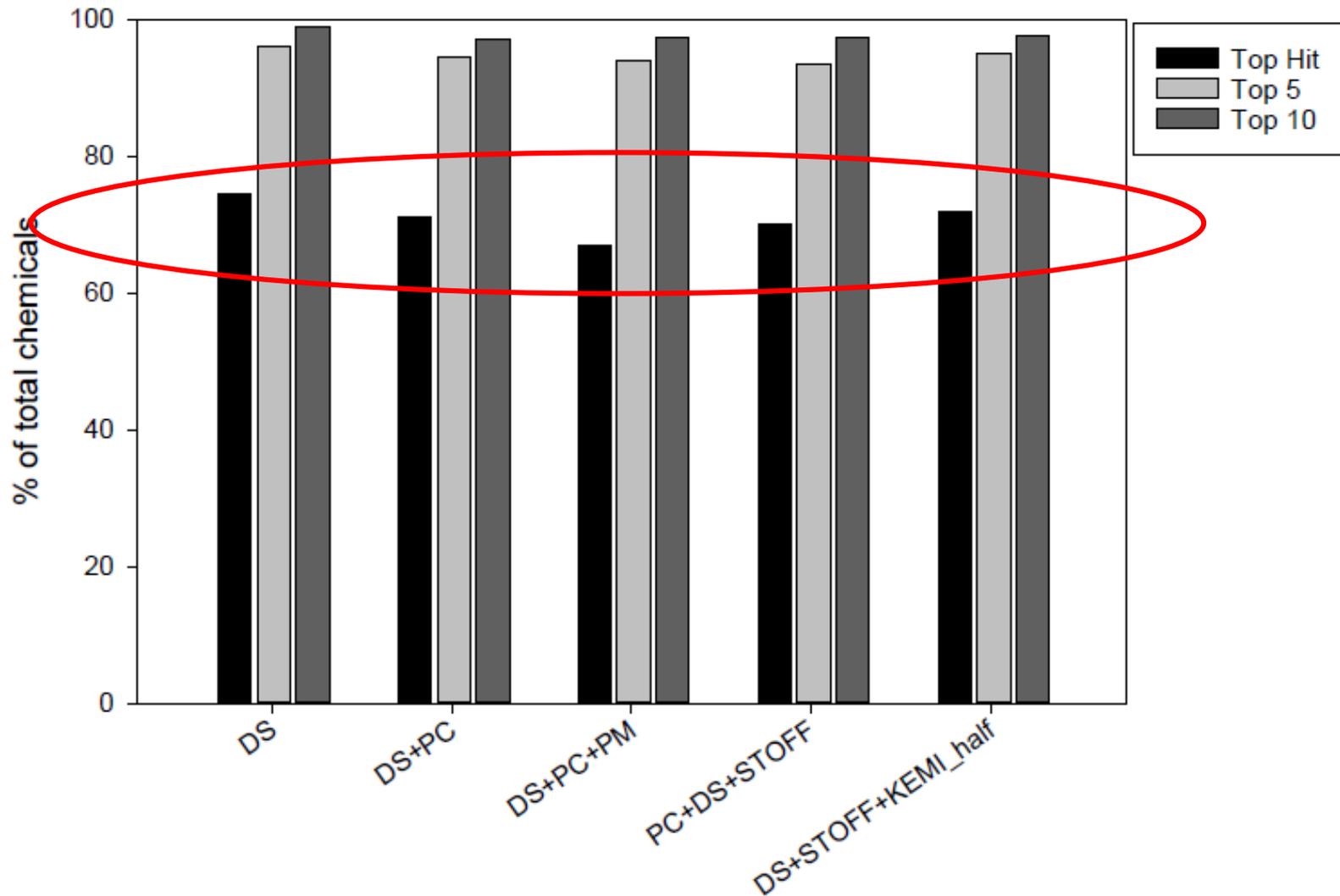
### Presence in Lists:

- ICCVAM test method evaluation report: in vitro ocular toxicity test methods
- 40CFR355
- A list of all PBDEs (Polybrominated diphenyl ethers)
- A list of all PCBs (Polychlorinated biphenyls)
- A list of polycyclic aromatic hydrocarbons
- Acute exposure guideline levels
- Algal Toxins
- Androgen Receptor Chemicals
- APCRA Chemicals for Prospective Analysis
- APCRA Chemicals for Retrospective Analysis
- APCRA Chemicals for Retrospective Analysis\_App\_List\_448\_Chemicals
- ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances
- ATSDR Toxic Substances Portal Chemical List
- Bisphenol Compounds
- California Office of Environmental Health Hazard Assessment
- Chemicals with interesting names
- CMAP
- DNT Screening Library
- Drinking Water Suspects, KWR Water, Netherlands
- EDSP Universe
- EPA Chemicals associated with hydraulic fracturing
  - Safer Choice Chemical List
  - Standard (no list)
  - Stockholm Convention on Organic Pollutants
  - STOFF-IDENT Database of Water-Relevant Substances
  - Superfund Chemical Data Matrix
  - Superfund Chemicals

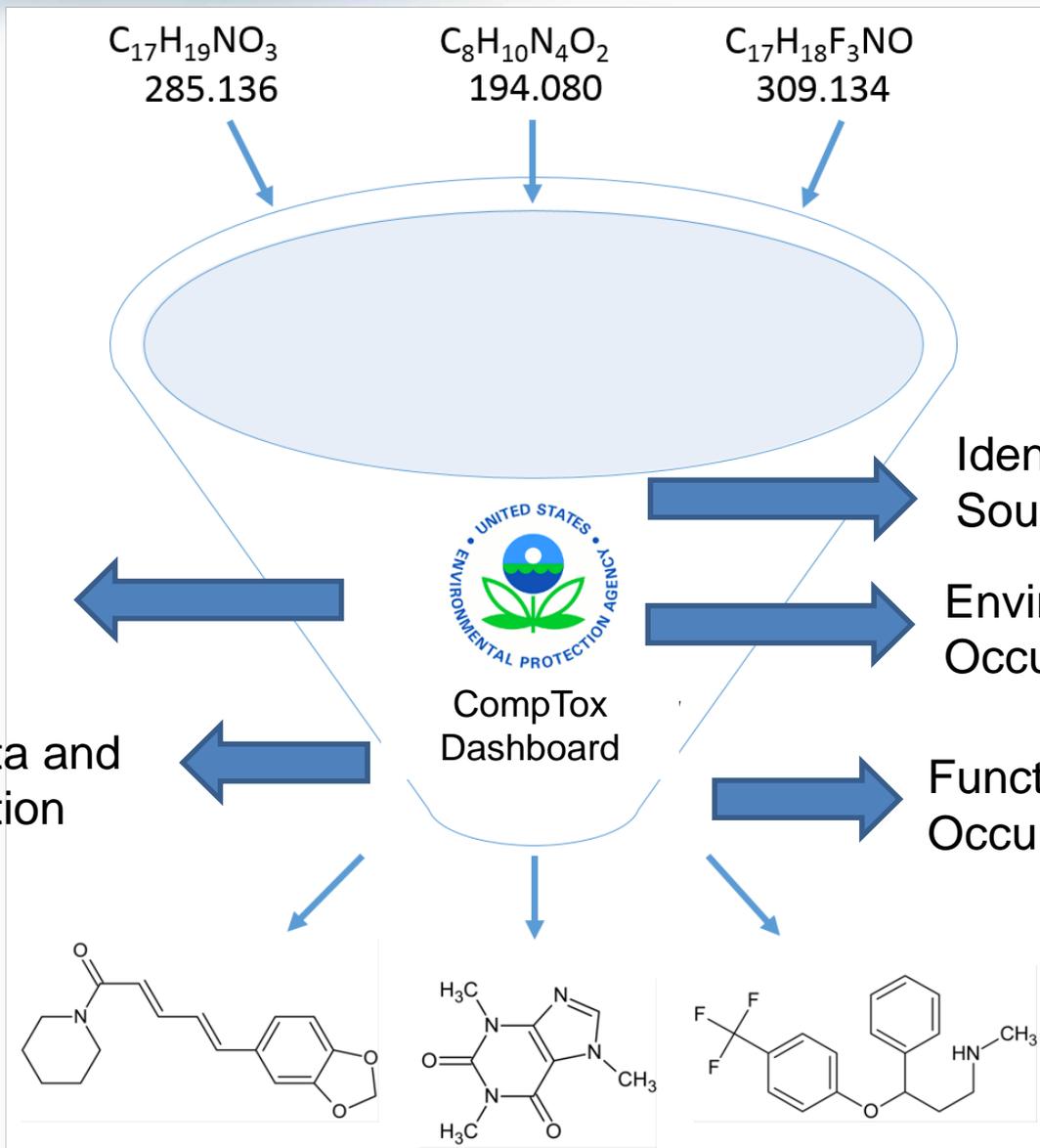


# Identification ranking: 1783 chemicals with multiple data streams

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$$



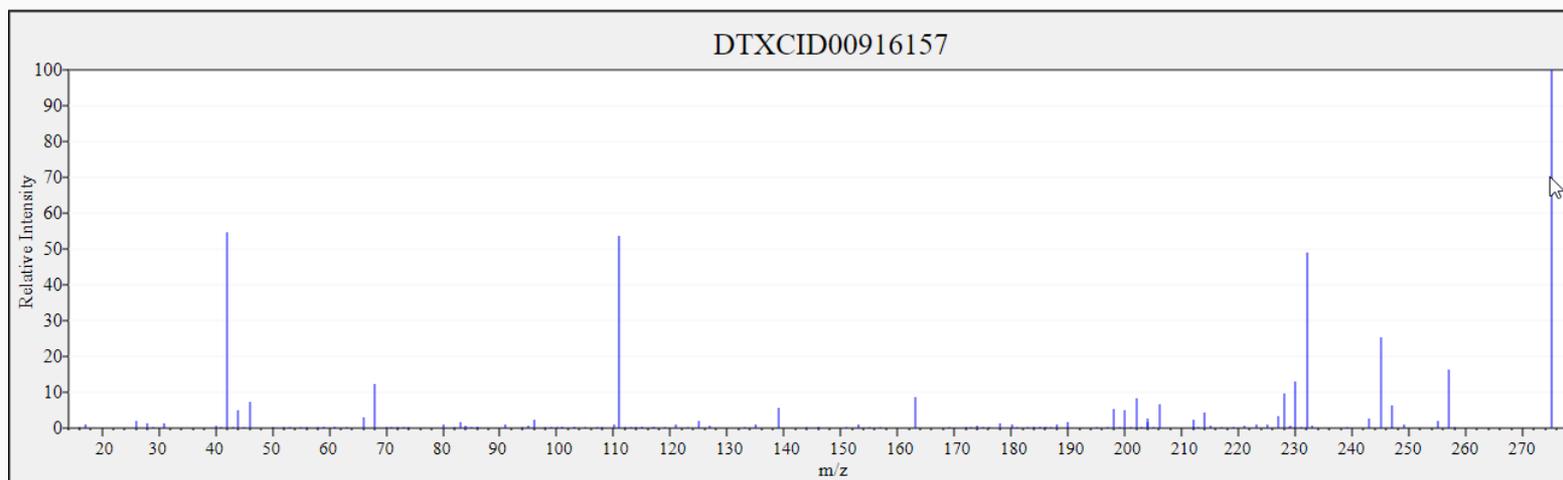
# Dashboard in NTA Workflows



# Work in Progress: Incorporating MS/MS data via the Dashboard

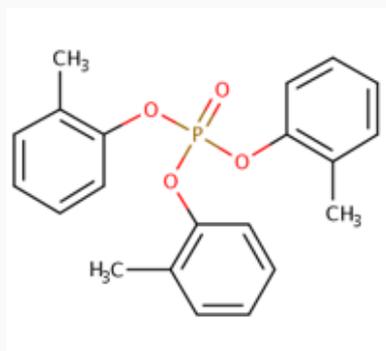


- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >700,000 structures, to be accessible via Dashboard

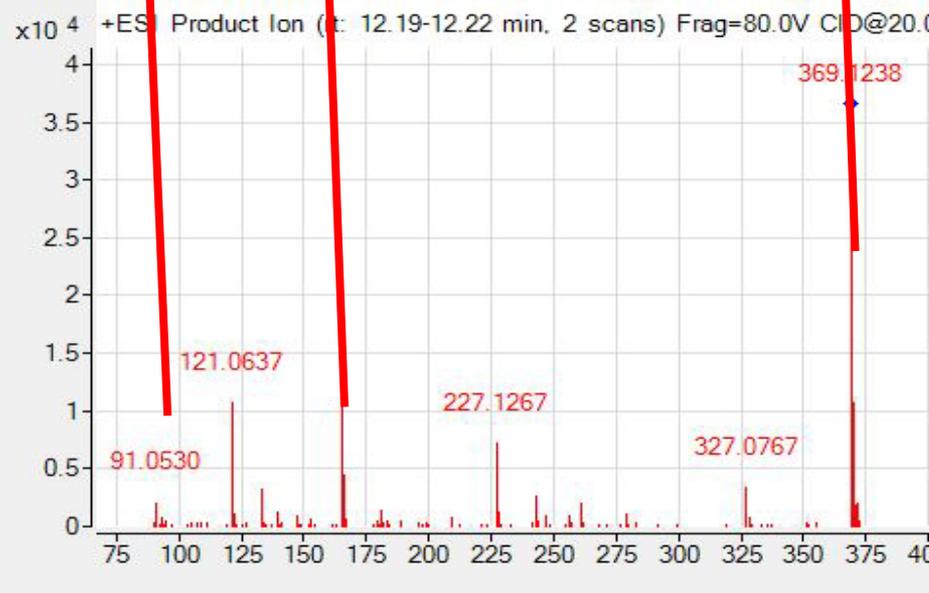
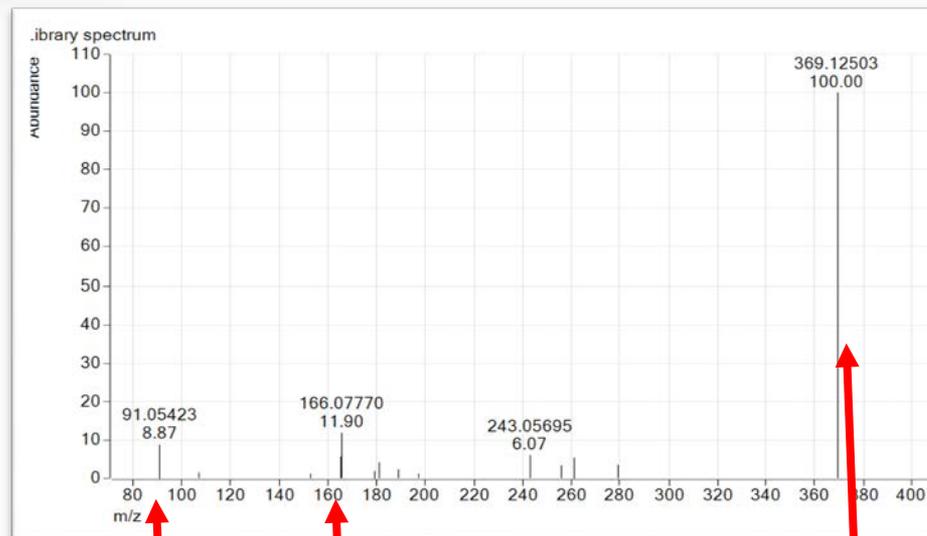


# Predicted Mass Spectra

Library Fragmentation  
Spectra (20eV)



Observed Fragmentation  
Spectra (20eV)



Match  
Score



- Critical Assessment of Small Molecule Identification
  - Training data= 312 peak lists (from 285 substances)
    - 234 MS/MS in positive mode
    - 58 in negative mode
  - Challenge Data= 208 peak lists (from 188 substances)
    - 127 in positive mode
    - 81 in negative mode
- Precursor ion search window= 15 ppm
- Fragment ion match threshold= 0.02 Da
- Candidates limited to Dashboard results within precursor ion search window



## CASMI 2016 Contest Challenge Set (n=208)

### CFM-ID only

	# Identified	% of Total
#1 Hits	89	43%
Top 5	154	74%
Top 10	174	84%
Top 20	190	91%

### CFM-ID +DSSTox Data Sources

	# Identified	% of Total
#1 Hits	154	74%
Top 5	195	94%
Top 10	198	95%
Top 20	202	97%

# Search Expt. vs. Predicted Spectra

## Non Target Analysis Prototype

### Mass Search

Da

±

Da

### Molecular Formula Search

*Mass or Formula must be entered before searching spectrum*

### Ionization Type

▼

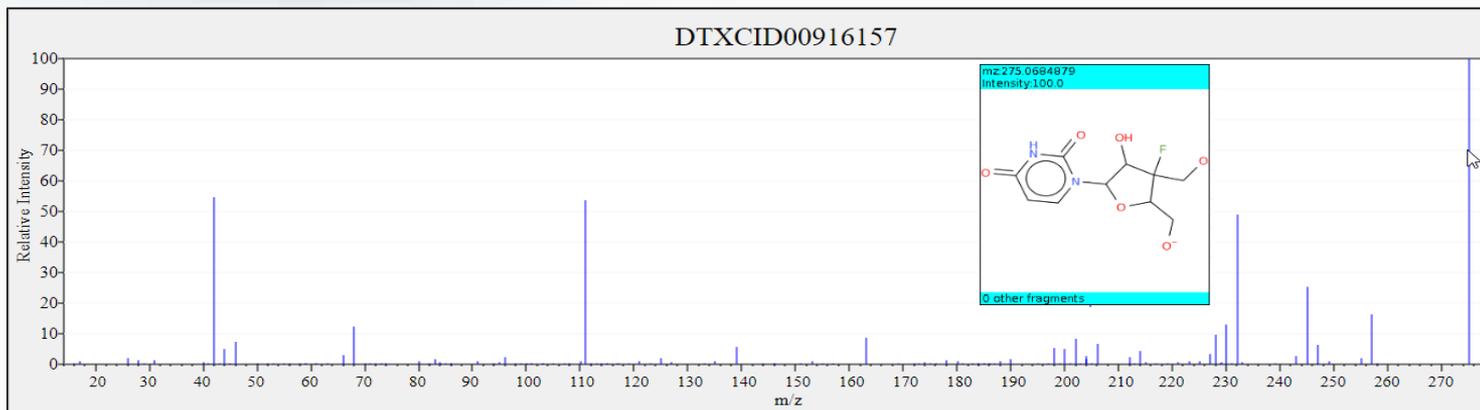
### Spectra Input

304.1332052 11.8199475  
198.0913404 7.306439699  
123.0440559 6.538348292  
196.0759904 5.269463115  
216.1019051 4.700461978  
200.1060005 4.80044284

Peak Match Window:

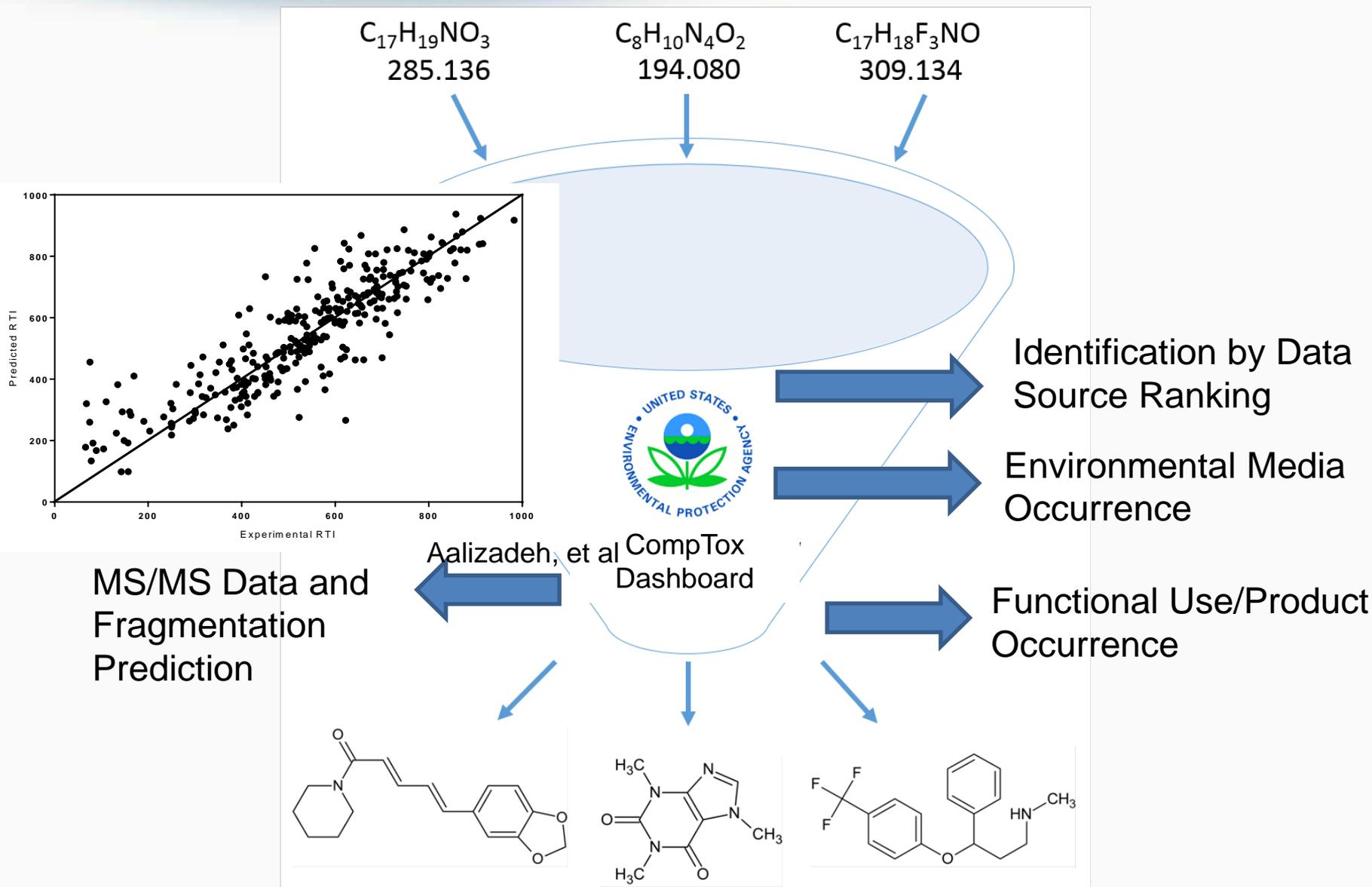
Da

# Future Access via the Dashboard



- Manuscript in Preparation for Nature Scientific Data
- Data will be available for download with publication
- (NO timeline...)

# Dashboard in NTA Workflows



# Future directions

## Intuitive Visualization of Results

$$SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \dots$$

 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

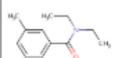
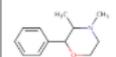
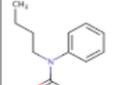
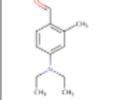
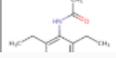
### Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

298 of 298 chemicals visible

Download / Send Select all

Sort by: Mass Difference Filter by: Name or CASRN Multicomponent Chemicals

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass	Mass Difference
	<a href="#">DTXSID2021995</a> <a href="#">ToxCast™</a>	DEET	134-62-3	Level 1	111	111	155	753	191.131014	0.000014
	<a href="#">DTXSID1023447</a>	Phendimetrazine	634-03-7	Level 2	12	28	35	50	191.131014	0.000014
	<a href="#">DTXSID2042197</a>	N-Butylacetanilide	91-49-6	Level 2	1	26	50	1	191.131014	0.000014
	<a href="#">DTXSID4059041</a>	Benzaldehyde, 4-(diethylamino)-2-methyl-	92-14-8	Level 3	0	7	51	0	191.131014	0.000014
	<a href="#">DTXSID90168148</a>	Acetanilide, 2,6'-diethyl-	16665-89-7	Level 4	0	4	33	0	191.131014	0.000014

# MetFrag Integration (more from Emma)

## Batch Search?



### Step Six: CI

Please enter one identifier per line

#### Select Input Type(s)

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

#### Metadata

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

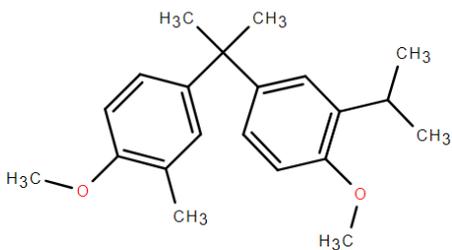
#### Enhanced Data Sheets

- MetFrag Input File (Beta) *i*
- ToxPrint single fingerprints *i*
- Abstract Sifter Input File (Beta) *i*
- Synonyms and Identifiers
- ToxPrint fingerprints (ChemoTyper format - CSV/TSV only)
- Associated ToxCast Assays

Step 6

- For every chemical registered into DSSTox we need to:
  - Generate QSAR-ready and MS-Ready forms
  - Pass the QSAR-ready form through predictors
    - OPERA PhysChem and Env. Fate and Transport
    - TEST PhysChem and Toxicity Models
    - Bioactivity models
  - More prediction algorithms are in development
- All predictions should be service-based
- We will deliver these services to users (eventually)

# Real-Time Predictions



Chiral



## Select properties to predict

H

TEST.

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

# Real-Time Predictions

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>

# API in development

## Prototype services available

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

JSON Raw Data Headers

Save Copy

```
uuid: "55547f4f-f966-48e8-b831-a0d217998064"
predictionTime: 1520539090089
software: "T.E.S.T (Toxicity Estimation Software Tool)"
softwareVersion: "5.01"
condition: "25°C"
endpoint: "Water solubility at 25°C"
method: "Hierarchical clustering"
▼ predictions:
  ▼ 0:
    id: "C_1520539090089"
    smiles: "OCC"
    expValMolarLog: "-1.337"
    expValMass: "1001180.703"
    predValMolarLog: "-1.338"
    predValMass: "1002625.241"
    molarLogUnits: "-Log10(mol/L)"
    massUnits: "mg/L"
```

- Prototype asynchronous pipelining of data already developed – to be integrated
- First prototype of QSAR/MS-ready processing about to go into testing
- Future plans (no deadline)
  - Batch processing of files to produce QSAR/MS-Ready (we encourage standard processes for QSAR modeling)
  - Web service access to QSAR/MS-Ready file generation

# A List of Lists of Chemicals

[https://comptox.epa.gov/dashboard/chemical\\_lists](https://comptox.epa.gov/dashboard/chemical_lists)



## Select List

Show  entries

Search:

Download

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
AEGLVALUES	<a href="#">Acute exposure guideline levels</a>	2018-04-20	174	Acute exposure guideline levels (AEGLs) describe the human health effects from once-in-a-lifetime, or rare, exposure to airborne chemicals.
ALGALTOX	<a href="#">Algal Toxins</a>	2017-11-21	54	A set of algal toxins of interest
ARCHEMICALS	<a href="#">Androgen Receptor Chemicals</a>	2018-05-01	110	The list of chemicals used to identify references with in vitro AR binding . From Kleinstreuer et al <a href="http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00347">http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00347</a>
ATHENSSUS	<a href="#">University of Athens Surfactant and Suspect List</a>	2017-07-14	60	ATHENSSUS is a compilation of suspects, predicted transformation products and surfactants screened in wastewater by University of Athens, as described in Gago-Ferrero et al 2015, DOI: 10.1021/acs.est.5b03454
ATSDRLST	<a href="#">ATSDR Toxic Substances Portal Chemical List</a>	2017-03-11	200	The Agency for Toxic Substances and Disease Registry (ATSDR) is a federal public health agency of the U.S. Department of Health and Human Services.
ATSDRMRLS	<a href="#">ATSDR Minimal Risk Levels (MRLs) for Hazardous Substances</a>	2018-05-02	188	The ATSDR Minimal Risk Levels (MRLs) were developed as an initial response to the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA)

# Eleven PFAS Lists

[http://comptox-prod.epa.gov/dashboard/chemical\\_lists](http://comptox-prod.epa.gov/dashboard/chemical_lists)



List Acronym	List Name	Last Updated	Number of Chemicals	List Description
EPAPFAS75S1	<a href="#">PFAS: EPA List of 75 Test Samples (Set 1)</a>	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.
EPAPFASCAT	<a href="#">PFAS: Registered DSSTox "category substances" representing Per- and Polyfluoroalkyl Substances (PFAS) categories</a>	2018-06-29	64	List of registered DSSTox "category substances" representing PFAS categories created using ChemAxon's Markush structure-based query representations.
EPAPFASINSOL	<a href="#">PFAS in EPA's Chemical Inventory Insoluble in DMSO</a>	2018-06-29	43	PFAS chemicals included in EPA's expanded ToxCast chemical inventory found to be insoluble in DMSO above 5mM.
EPAPFASINV	<a href="#">PFAS in EPA's ToxCast Chemical Inventory</a>	2018-06-29	430	PFAS chemicals included in EPA's expanded ToxCast chemical inventory and available for testing.
EPAPFASRL	<a href="#">PFAS: EPA Cross-Agency Research List</a>	2018-08-12	194	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.
PFASEUOECD	<a href="#">PFAS Listed in OECD Global Database</a>	2018-07-26	4725	OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing approximately 4700 new PFAS
PFASKEMI	<a href="#">PFAS List from the Swedish Chemicals Agency (KEMI) Report</a>	2017-02-09	2397	Perfluorinated substances from a Swedish Chemicals Agency (KEMI) Report on the occurrence and use of highly fluorinated substances.
PFASMASTER	<a href="#">PFAS Master List of PFAS Substances</a>	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.



Port

HOME



## The OECD releases a new list of PFASs

The OECD releases a new list of Per- and Polyfluoroalkyl Substances (PFASs) based on a comprehensive analysis of information available in the public domain. In total, 4730 PFAS-related CAS numbers have been identified and categorised in this study, including several new groups of PFASs that fulfil the common definition of PFASs (i.e. they contain at least one perfluoroalkyl moiety) but have not yet been commonly regarded as PFASs.

This work has been conducted under the OECD/UN Environment Global PFC Group in support of the Strategic Approach to International Chemicals Management (SAICM) and shifting to safer alternatives for PFASs.

The [New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances \(PFASs\)](#) comes with a [methodology report](#) also detailing the major findings with respect to the total numbers and types of PFASs identified, the limitations, gaps and challenges identified in the development of the new list, and opportunities for improving the future understanding of PFASs production, use on the global market, and presence in the environment, biota, and other matrices.



INARS



# The OECD List of PFAS

<http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals/>

## PFAS Listed in OECD Global Database

Search PFASEUOECD Chemicals



Substring search

### List Details

**Description:** OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances (PFASs) listing approximately 4700 new PFAS, including several new groups of PFASs that fulfill the common definition of PFASs (i.e. they contain at least one perfluoroalkyl moiety) but have not yet been commonly regarded as PFASs. The list can be used in conjunction with the methodology report summarising the major findings with respect to the total numbers and types of PFASs identified, the limitations, gaps and challenges identified, and opportunities for improving the future understanding of PFASs production, use on the global market, and presence in the environment, biota, and other matrices.

Source website: <http://www.oecd.org/chemicalsafety/portal-perfluorinated-chemicals>

A major effort was undertaken to register this list within DSSTox, adding chemical structures for as many PFAS entries as possible using both manual and auto-mapping (structures using CAS-matching) curation methods. The result is that approximately 1/3 of the list is curated at the highest two curation levels (DSSTox\_High or DSSTox\_Low) currently, whereas more than half of this list is registered at the Public\_Low curation level (based on PubChem content). The PFASOECD list is undergoing continuous registration and curation.

Number of Chemicals: 4725

4725 chemicals

Download / Send ▼

Show info:

DTXSID ×

CASRN ×

TOXCAST ×

Select all



Sort by: DTXSID ▼



Filter by: Name or CASRN

Hide ▼

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

Posted: 11/14/2016

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

## [DSSTox SDF File](#)

Posted: 12/14/2016

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

## [PHYSROP Analysis File](#)

Posted: 12/14/2016

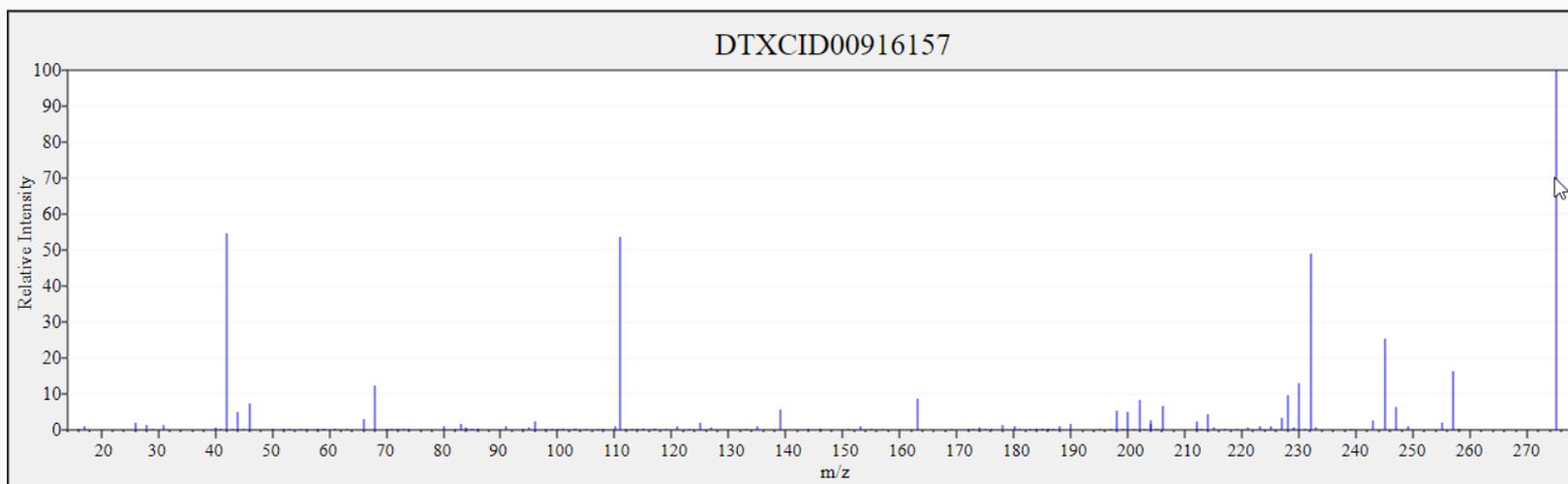
- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- MetFrag integration (beta already done)
- Investigating Retention Time Index Prediction for candidate ranking
- Structure/substructure/similarity search

# Predicted Mass Spectra

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



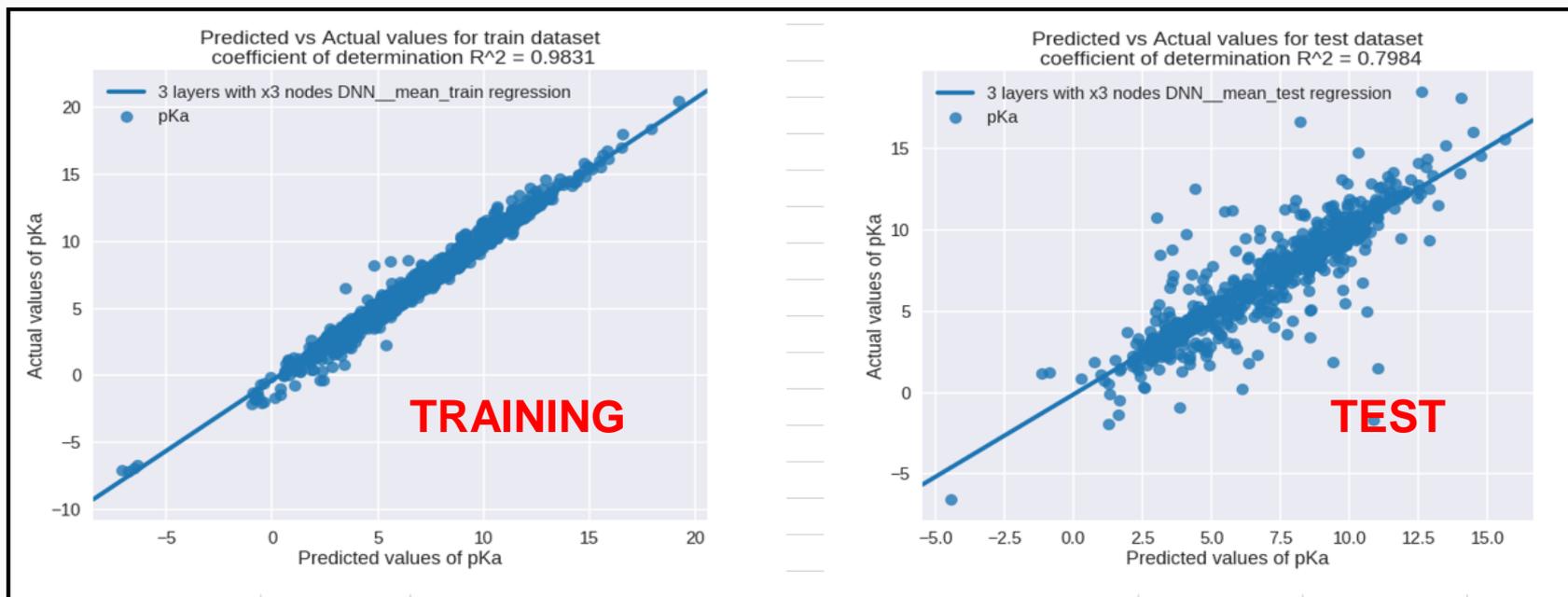
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >700,000 structures, to be accessible via Dashboard



# What structures are amenable to LC vs GC? Prediction models

- For many of our studies we would like to know whether chemicals are amenable to LC vs GC, +ve vs -ve ion etc?
- Public databases are being harvested in order to build training sets for prediction models and availability of public data

- pKa prediction models based on Open Data Set of 8000 chemicals – acidic, basic and amphoteric chemicals
- pKa for prediction of LC/GC amenable??



# Prototype Development

## AADashboard

atrazine  Search

100%

### Select properties to predict

T.E.S.T. 18 OPERA Search

- H
- C
- N
- O
- S
- P
- F
- Cl

- Exact
- Substructure
- Similarity
- Molecular Formula
- Molecular Weight

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

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

1	0.62	0.57	0.57	0.57	0.53	0.53	0.53	0.5	0.5	0.5
0.5	0.5	0.5	0.47	0.44	0.44	0.44	0.42	0.42	0.42	0.42
0.42	0.42	0.42	0.42	0.42	0.4	0.4	0.4	0.4	0.4	0.4
0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.38
0.38	0.38	0.38	0.38	0.38	0.38					

Search result **2540** Show  Isotopically

- The CompTox Chemistry Dashboard provides access to data for ~762,000 chemicals
- High quality data from ongoing curation efforts
- An integration hub for multiple “modules”
- The dashboard serves many purposes but has functionality to support Mass Spectroscopy
- MS-Ready processing of chemicals has high value
- With much work to do, we are well on the way...

# Acknowledgments

- The CompTox Dashboard Development Team
- NCCT scientists providing data and modules
- The NERL Mass Spectrometry Team
- Our wonderful team of curators
- Tommy Cathey, Tom Transue and Ilya Balyabin (CFM-ID fragmentation work)
- Multiple external collaborators via NORMAN
- Emma Schymanski (so many contributions)
- Christoph Ruttkies (MetFrag)



## **Antony Williams**

US EPA Office of Research and Development

National Center for Computational Toxicology (NCCT)

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

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