

CompTox chemicals dashboard: Data and tools to support chemical and environmental risk assessment and the ENTACT project

***Antony Williams¹, Chris Grulke¹, Andrew McEachran²,
Emma Schymanski³, Jon Sobus⁴, Elin Ulrich⁴, Ann Richard¹,
Jeremy Dunne¹ and Jeff Edwards¹***

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*

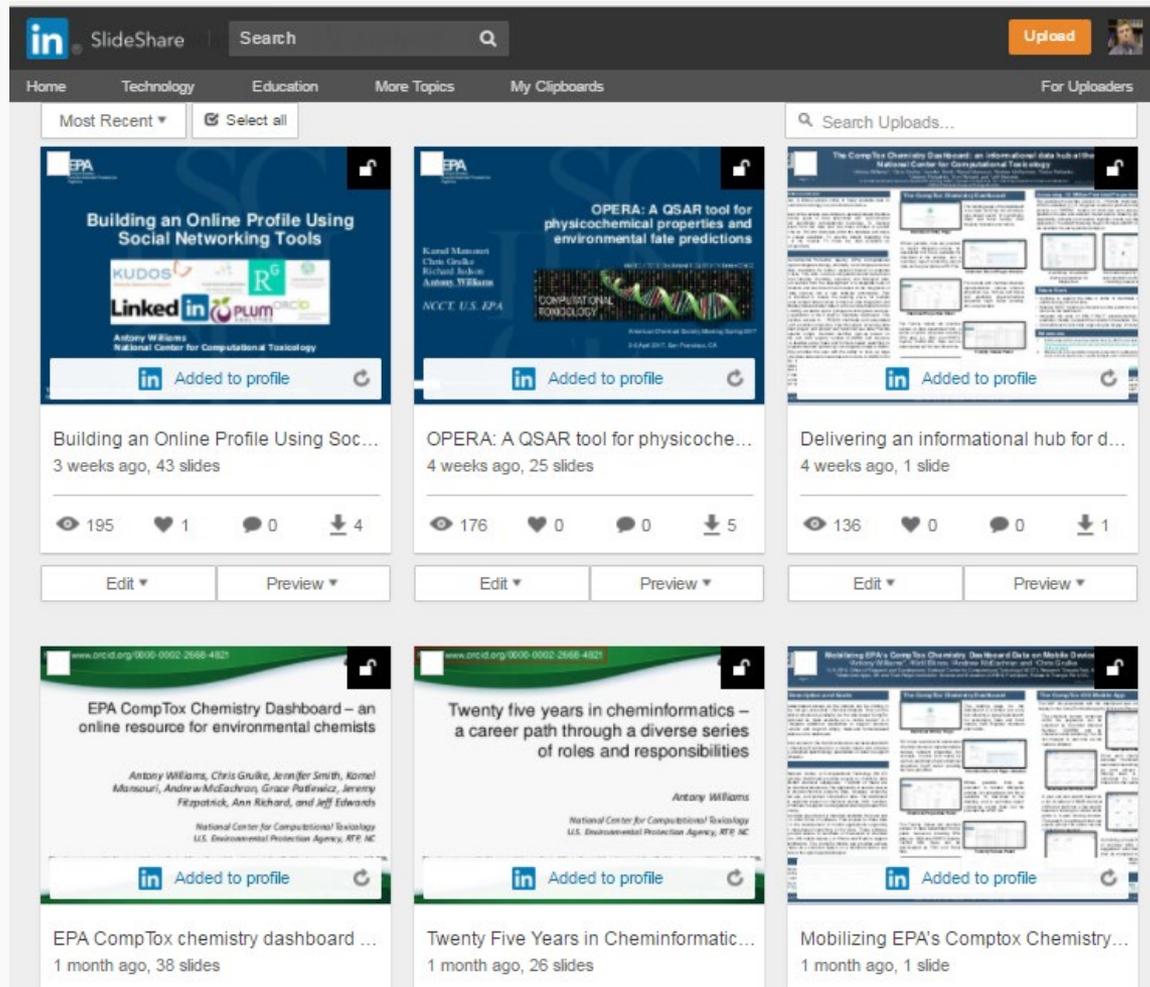
4) *National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*

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

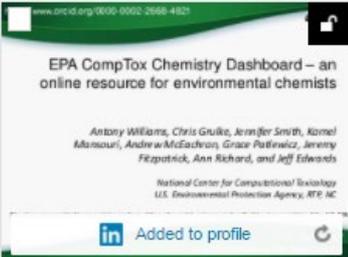
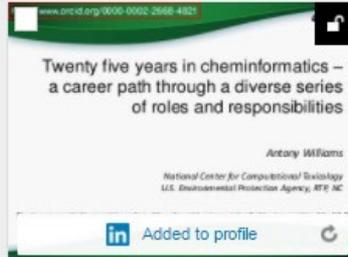
*Spring 2019
ACS Spring Meeting, Orlando*

Limit notetaking if you wish

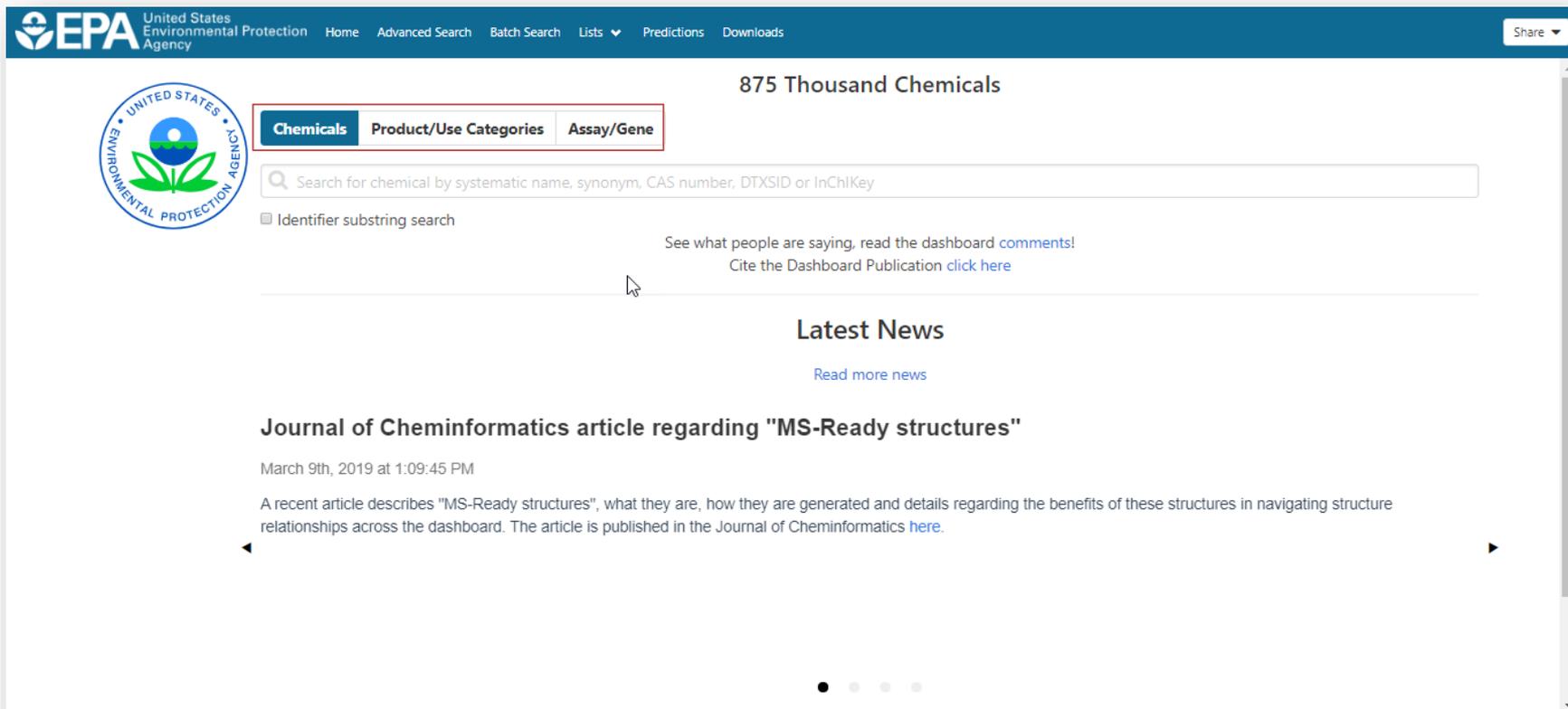
www.slideshare.net/AntonyWilliams



The screenshot displays the SlideShare profile of Antony Williams, a member of the National Center for Computational Toxicology at the EPA. The profile features a navigation bar with 'Home', 'Technology', 'Education', 'More Topics', 'My Clipboards', and 'For Uploaders'. Below the navigation bar, there are filters for 'Most Recent' and 'Select all'. The main content area shows a grid of six presentations, each with a thumbnail, title, description, and engagement metrics (views, likes, comments, downloads). Each presentation also has 'Edit' and 'Preview' buttons.

Thumbnail	Title	Description	Views	Likes	Comments	Downloads
	Building an Online Profile Using Social Networking Tools	3 weeks ago, 43 slides	195	1	0	4
	OPERA: A QSAR tool for physicochemical properties and environmental fate predictions	4 weeks ago, 25 slides	176	0	0	5
	Delivering an informational hub for d...	4 weeks ago, 1 slide	136	0	0	1
	EPA CompTox Chemistry Dashboard – an online resource for environmental chemists	1 month ago, 38 slides				
	Twenty five years in cheminformatics – a career path through a diverse series of roles and responsibilities	1 month ago, 26 slides				
	Mobilizing EPA's CompTox Chemistry Dashboard Data on Mobile Devices	1 month ago, 1 slide				

875k Chemical Substances



The screenshot displays the EPA CompTox Chemicals Dashboard. At the top, the EPA logo and navigation links (Home, Advanced Search, Batch Search, Lists, Predictions, Downloads) are visible. The main heading is "875 Thousand Chemicals". Below this, there are three tabs: "Chemicals", "Product/Use Categories", and "Assay/Gene". A search bar is present with the placeholder text "Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey". There is also a checkbox for "Identifier substring search". Below the search bar, there are links to "See what people are saying, read the dashboard comments!" and "Cite the Dashboard Publication [click here](#)". The "Latest News" section features an article titled "Journal of Cheminformatics article regarding 'MS-Ready structures'" dated March 9th, 2019. The article text states: "A recent article describes 'MS-Ready structures', what they are, how they are generated and details regarding the benefits of these structures in navigating structure relationships across the dashboard. The article is published in the Journal of Cheminformatics [here](#)."

Detailed Chemical Pages

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

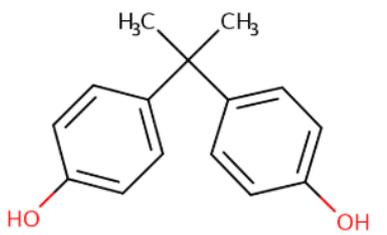
Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.



Wikipedia

Bisphenol A (BPA) is an organic synthetic compound with the chemical formula $(\text{CH}_3)_2\text{C}(\text{C}_6\text{H}_4\text{OH})_2$ belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water (0.344 wt % at 83 °C). BPA is a starting material for the synthesis of plastics, primarily certain polycarbonates

[Read more](#)

Intrinsic Properties

Molecular Formula: $\text{C}_{15}\text{H}_{16}\text{O}_2$ [Mol File](#) [Find All Chemicals](#)

Average Mass: 228.291 g/mol [Isotope Mass Distribution](#)

Monoisotopic Mass: 228.11503 g/mol

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

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

Search query

Product and Use Categories (PUCs)

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

First << < 1 2 3 4 5 6 7 8 9 10 > >> Last

EXPOSURE

PRODUCT & USE CATEGORIES

CHEMICAL WEIGHT FRACTION

CHEMICAL FUNCTIONAL USE

TOXICS RELEASE INVENTORY

MONITORING DATA

EXPOSURE PREDICTIONS

PRODUCTION VOLUME

- 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/Formula Searching and Metadata Ranking

Advanced Searches

Mass Search

Mass Search

\pm Min/Max

Adduct

Neutral



All Adducts



Choose adduct from dropdown

191.131

Da

\pm

5

Da

ppm

Search 

Advanced Searches

Mass Search

Search Results

Searched by Mass: 191.131 +/- 5.0 ppm.

329 of 329 chemicals visible

Select all

Download

Send to Batch Search

Mass Difference

DTXSID

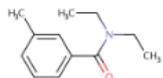
CASRN

TOXCAST

Mass Diff

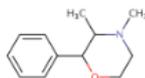
Multicomponent Chemicals

Filter by Name or CASRN



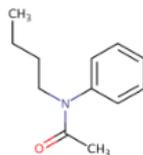
DEET

DTXSID: DTXSID2021995
CASRN: 134-62-3
TOXCAST: 12/768
Mass Diff: 0.000014



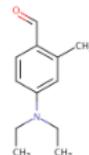
Phendimetrazine

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



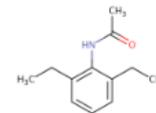
N-Butylacetanilide

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



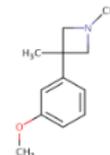
Benzaldehyde, 4-(diethylamino)-2-methyl-

DTXSID: DTXSID90168148
CASRN: 92-14-8
TOXCAST: -
Mass Diff: 0.000014



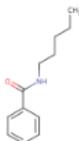
Acetanilide, 2',6'-diethyl-

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



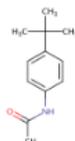
Azetidine, 1,3-dimethyl-3-(m-methoxyphenyl)-

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



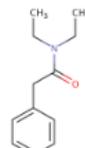
Benzamide, N-pentyl-

DTXSID: DTXSID20174196
CASRN: 20308-43-4
TOXCAST: -
Mass Diff: 0.000014



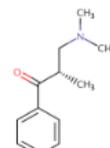
p-t-Butylacetanilide

DTXSID: DTXSID00174238
CASRN: 20330-45-4
TOXCAST: -
Mass Diff: 0.000014



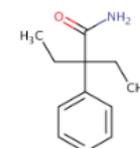
N,N-Diethylphenylacetamide

DTXSID: DTXSID00179048
CASRN: 2431-96-1
TOXCAST: -
Mass Diff: 0.000014



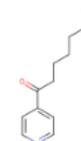
3-(Dimethylamino)-2-methylpropiofen...

DTXSID: DTXSID60180796
CASRN: 26171-50-6
TOXCAST: -
Mass Diff: 0.000014



Butyramide, 2-ethyl-2-phenyl-

DTXSID: DTXSID60184653
CASRN: 30568-39-9
TOXCAST: -
Mass Diff: 0.000014



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

DTXSID: DTXSID40186594
CASRN: 32941-30-3
TOXCAST: -
Mass Diff: 0.000014

MS-Ready Structures for Formula Search

Molecular Formula Search

MS Ready Formula  Exact Formula 

Formula

Please use the format of the following example: C₆H₈O₂ or C₆H(8-10)O(0-2)

Search 

“MS-Ready Structures”

<https://doi.org/10.1186/s13321-018-0299-2>

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

Journal of Cheminformatics

METHODOLOGY

Open Access

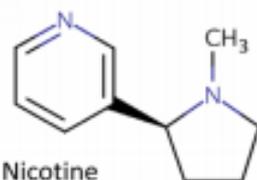


“MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies

Andrew D. McEachran^{1,2*}, Kamel Mansouri^{1,2,3}, Chris Grulke², Emma L. Schymanski⁴, Christoph Ruttkies⁵ and Antony J. Williams^{2*}

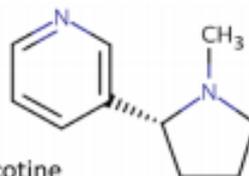
Open Science for Identifying “Known Unknown” Chemicals

Emma L. Schymanski*[†] and Antony J. Williams*[‡]



Nicotine

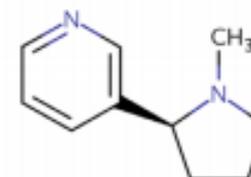
CN1CCC[C@H]1C1=CN=CC=C1
DTXSID1020930 | SNICXCGAKADSCV
54-11-5 | **162.1157** | 0.929 | **72**
Tox: **yes** | Expo: **yes** | Bioassay: **yes**



D-Nicotine

CN1CCC[C@@H]1C1=CN=CC=C1
DTXSID004635 | SNICXCGAKADSCV
25162-00-9 | **162.1157** | 0.929 | **20**
Tox: **no** | Expo: **yes** | Bioassay: **yes**

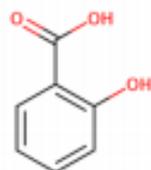
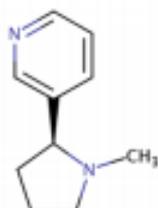
LEGEND: Name, SMILES
DTXSID | InChIKey 1st Block
CAS | Monoiso. Mass | logP | Sources
Data on: Toxicity | Exposure | Bioassays



HCl

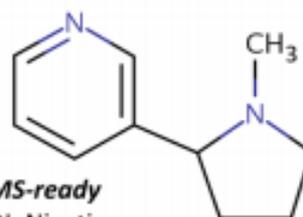
Nicotine hydrochloride

Cl.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID602093 | HDJBTCAJIMNXEW
2820-51-1 | **198.0924** | 0.929 | **9**
Tox: **no** | Expo: **yes** | Bioassay: **yes**



Benzoic acid, 2-hydroxy-, compd. with
3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)

OC(=O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID5075319 | AIBWPBUAKCMKNS
29790-52-1 | **300.1474** | 0.929 | **6**
Tox: **no** | Expo: **yes** | Bioassay: **no**



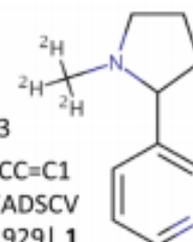
MS-ready

DL-Nicotine

CN1CCCC1C1=CN=CC=C1
DTXSID3048154 | SNICXCGAKADSCV
22083-74-5 | **162.1157** | 0.953 | **9**
Tox: **yes** | Expo: **no** | Bioassay: **yes**

DL-Nicotine-d3

[2H]C([2H])([2H])N1CCCC1C1=CN=CC=C1
DTXSID80442666 | SNICXCGAKADSCV
69980-24-1 | **165.1345** | 0.929 | **1**
Tox: **no** | Expo: **no** | Bioassay: **no**



MS-Ready Mappings

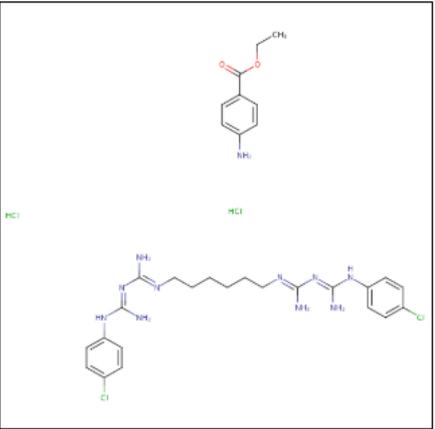
 United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Copy Share Submit Comment Search all data

Progaron

108532-15-6 | DTXSID20148579
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

Intrinsic Properties

Molecular Formula: $C_{31}H_{43}Cl_4N_1O_2$ Mol File Find All Chemicals

Average Mass: 743.56 g/mol Isotope Mass Distribution

Structural Identifier

Linked Substances

Same Connectivity: 1 (nChl)

Mixtures, Component records

MS-Ready Mappings: DTXCID301804: 12 records; DTXCID0013314: 11 records;

Similar Compounds: 0 records

Quality Control Notes

MS-Ready Mappings Set

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

Select all Download Send to Batch Search

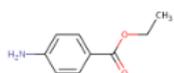
Default

DTXSID PubChem CPDAT

10 of 12 chemicals visible

Isotopes

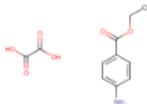
Filter by Name or CASRN



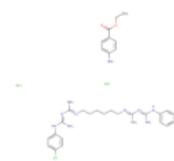
Benzocaine
DTXSID: DTXSID8021804
PubChem: 184
CPDAT: 42



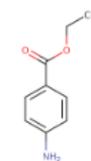
Anesthesine succinate
DTXSID: DTXSID60148336
PubChem: 10
CPDAT: 0



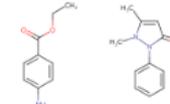
Anesthesine oxalate
DTXSID: DTXSID20148337
PubChem: 6
CPDAT: 0



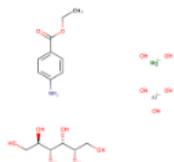
Progaron
DTXSID: DTXSID20148579
PubChem: 5
CPDAT: 0



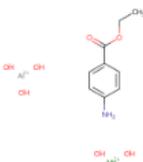
Benzocaine hydrochloride
DTXSID: DTXSID50177812
PubChem: 33
CPDAT: 0



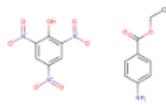
Antipyrine mixture with benzocaine
DTXSID: DTXSID80212866
PubChem: 9
CPDAT: 0



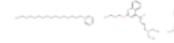
Almagel A-neo
DTXSID: DTXSID60227559
PubChem: 9
CPDAT: 0



Almagel
DTXSID: DTXSID70227560
PubChem: 9
CPDAT: 0



Ethyl 4-aminobenzoate--2,4,6-trinitroph...
DTXSID: DTXSID70787033
PubChem: 5
CPDAT: 0



1-Hexadecylpyridin-1-ium 2-butoxy-N-...
DTXSID: DTXSID50997335
PubChem: 0
CPDAT: 0

- EXACT Formula:** C₁₀H₁₆N₂O₈: 3 Hits

MS Ready Formula  Exact Formula 

Formula

C₁₀H₁₆N₂O₈

Select all

Download

Send to Batch Search

Default



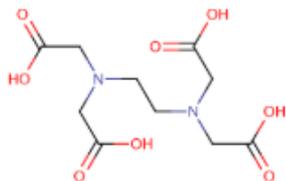
DTXSID 

PubChem 

CPDAT 



3 of 3 chemi

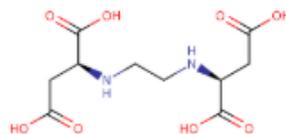


Ethylenediaminetetraacetic acid

DTXSID: DTXSID6022977

PubChem: 158

CPDAT: 387

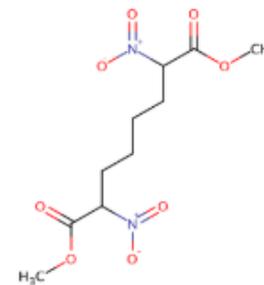


N,N'-Ethylenedi-L-aspartic acid

DTXSID: DTXSID1051852

PubChem: 25

CPDAT: 8



Dimethyl 2,7-dinitrooctanedioate

DTXSID: DTXSID20498864

PubChem: 5

CPDAT: 0

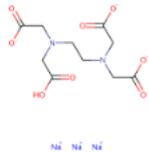
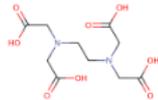
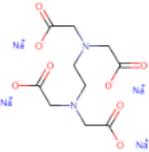
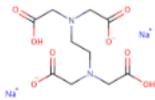
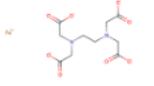
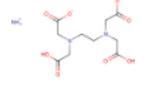
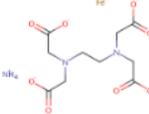
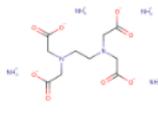
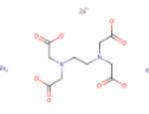
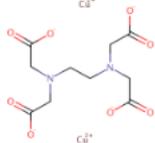
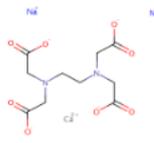
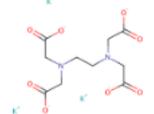
MS-Ready Mappings

- **Same Input Formula: C₁₀H₁₆N₂O₈**
- **MS Ready Formula Search: 125 Chemicals**

125 chemicals

Select all Download Send to Batch Search Default DTXSID PubChem CPDAT

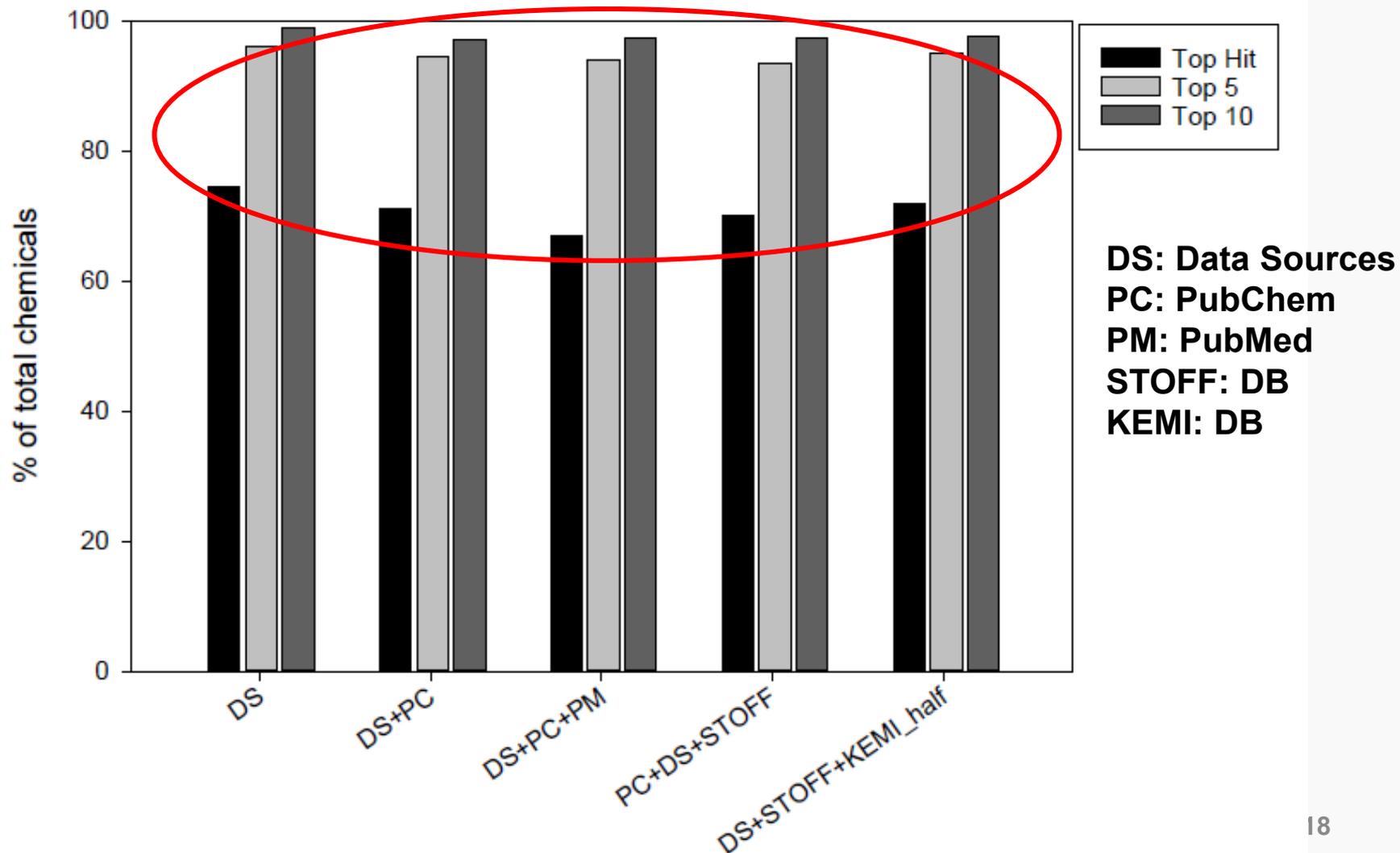
Hide chemicals that are: Filter by Name or CASRN

 <p>Trisodium ethylenediaminetetraacetate DTXSID: DTXSID7020556 PubChem: 33 CPDAT: 82</p>	 <p>Ethylenediaminetetraacetic acid DTXSID: DTXSID6022977 PubChem: 158 CPDAT: 387</p>	 <p>Ethylenediaminetetraacetic acid tetrasod... DTXSID: DTXSID3026350 PubChem: 57 CPDAT: 1227</p>	 <p>Ethylenediaminetetraacetic acid, disodiu... DTXSID: DTXSID9027073 PubChem: 56 CPDAT: 1359</p>	 <p>Ethylenediaminetetraacetic acid ferric so... DTXSID: DTXSID5027774 PubChem: 53 CPDAT: 62</p>	 <p>Diammonium dihydrogen ethylenediami... DTXSID: DTXSID9027813 PubChem: 12 CPDAT: 17</p>
 <p>Ferrate(1-), [[N,N'-1,2-ethanediy]bis[N-]]... DTXSID: DTXSID9027815 PubChem: 24 CPDAT: 20</p>	 <p>Tetraammonium ethylenediaminetetraac... DTXSID: DTXSID8027820 PubChem: 11 CPDAT: 12</p>	 <p>Zincate(2-), [[N,N'-1,2-ethanediy]bis[N-]]... DTXSID: DTXSID8028343 PubChem: 5 CPDAT: 10</p>	 <p>EDTA, copper salt DTXSID: DTXSID0034564 PubChem: 8 CPDAT: 10</p>	 <p>Calcium disodium ethylenediaminetetra... DTXSID: DTXSID2036409 PubChem: 42 CPDAT: 29</p>	 <p>Tetrapotassium ethylenediaminetetra... DTXSID: DTXSID3036442 PubChem: 25 CPDAT: 36</p>

Candidate ranking using public resources

- Use available metadata to rank candidates
 - Associated data sources
 - Associated lists in the underlying database
 - Associated data sources in PubChem
 - Specific types (e.g. water, surfactants, pesticides etc.)
 - Number of associated literature articles (Pubmed)
 - **Chemicals in the environment** – the number of products/categories containing the chemical is a very important source of data

Identification ranks for 1783 chemicals using multiple data streams



Is a bigger database better?



© American Society for Mass Spectrometry, 2011

J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

RESEARCH ARTICLE

Identification of “Known Unknowns” Utilizing Accurate Mass Data and ChemSpider

- ChemSpider was 26 million chemicals then
- Much BIGGER today
- Is bigger better??

71 Million
chemical structures

^{vs}
Anal Bioanal Chem (2017) 49:1729–1735
DOI 10.1007/s00216-016-0139-z



RAPID COMMUNICATION

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

Andrew D. McEachran¹ · Jon R. Sobus² · Antony J. Williams³

- Dashboard content *was* 720k chemicals
- Only 3% of ChemSpider size
- What was the comparison in performance?

SAME dataset for comparison

Compound class	Number in class	Average rank	Number of compounds in each position rank-ordered				
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Steroid hormones	7	1.0	7				
Perfluorochemicals	6	1.2	5	1			
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 ^a	1.3	2	1			

How did performance compare?

Summary statistics and rank-ordered position in the CompTox Chemistry Dashboard and ChemSpider of the 89 compound subset from the Little et al. [7] study

		Average rank	Number in each position rank-ordered				
		(±SD)	#1	#2	#3	#4	#5+
Mass-based	Dashboard	1.2 ± 0.7	77 ^a	5	3	3	
	ChemSpider	2.2 ± 6.1 ^b	68	8	7	1	5
Formula-based	Dashboard	1.1 ± 0.4	78 ^a	8	2		
	ChemSpider	1.3 ± 1.0	77	8	2	1	2

^aOne chemical (tephrosin) not present in the Dashboard

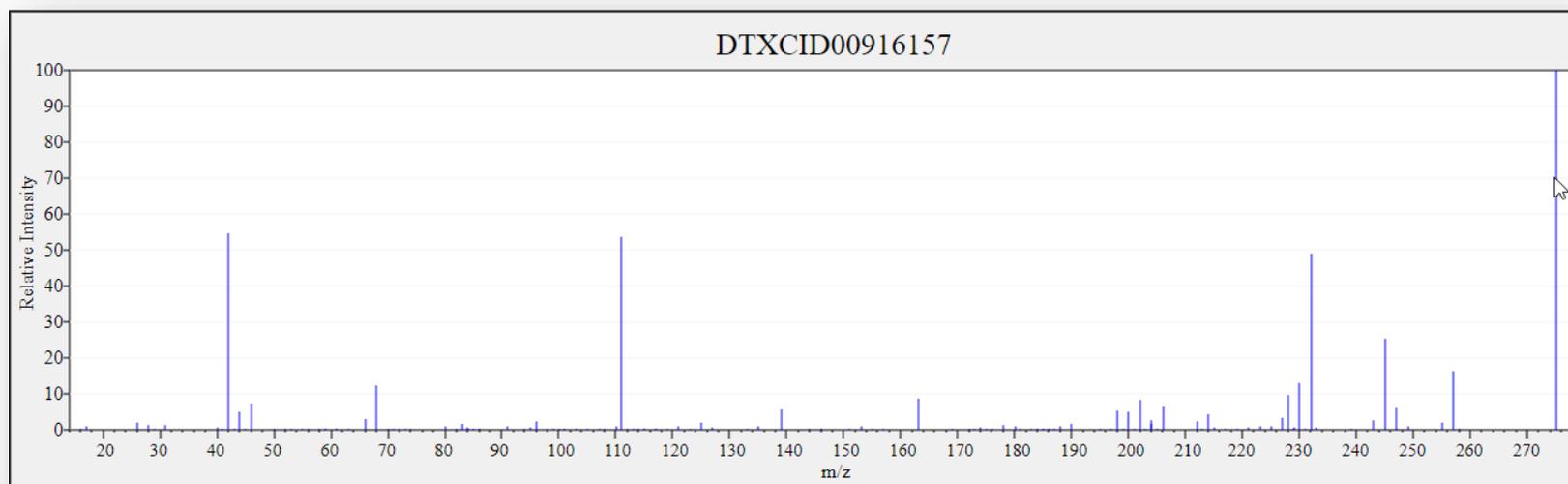
WORK IN PROGRESS

Predicted Mass Spectra

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



- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra

Non Target Analysis Prototype

Mass Search

Min/Max

Da Da ppm

Molecular Formula Search

Mass or Formula must be entered before searching spectrum

Ionization Type

ESI+ ▼
ESI+
ESI-
EI

Spectra Input

Single Energy Multiple

304.1332052 11.6199475
198.0913404 7.308439699
123.0440559 6.538348292
198.0756904 5.269463115
216.1019051 4.700461978
300.1080005 4.800144384

Peak Match Window: Da ppm

Search Expt. vs. Predicted Spectra

Spectra Input

Single Ener

304.1332052 11.61
198.0913404 7.30
123.0440559 6.53
196.0756904 5.28
216.1019051 4.70
200.4080005 4.80

Peak Match

Search

TSV CSV Excel

Chemical Structure ID

DTXCID101048191

DTXCID101181567

DTXCID50879086

DTXCID60686349

DTXCID00830900

DTXCID10971176

DTXCID60301242

DTXCID40703048

DTXCID60349982

DTXCID10316649

Showing 1 to 10 of 38 entries

Chemical Structure ID

Score (10eV)

DTXCID101048191

0.22

DTXCID101181567

0.19

DTXCID50879086

0.17

DTXCID60686349

0.14

DTXCID00830900

0.13

DTXCID10971176

0.12

DTXCID60301242

0.12

DTXCID40703048

0.11

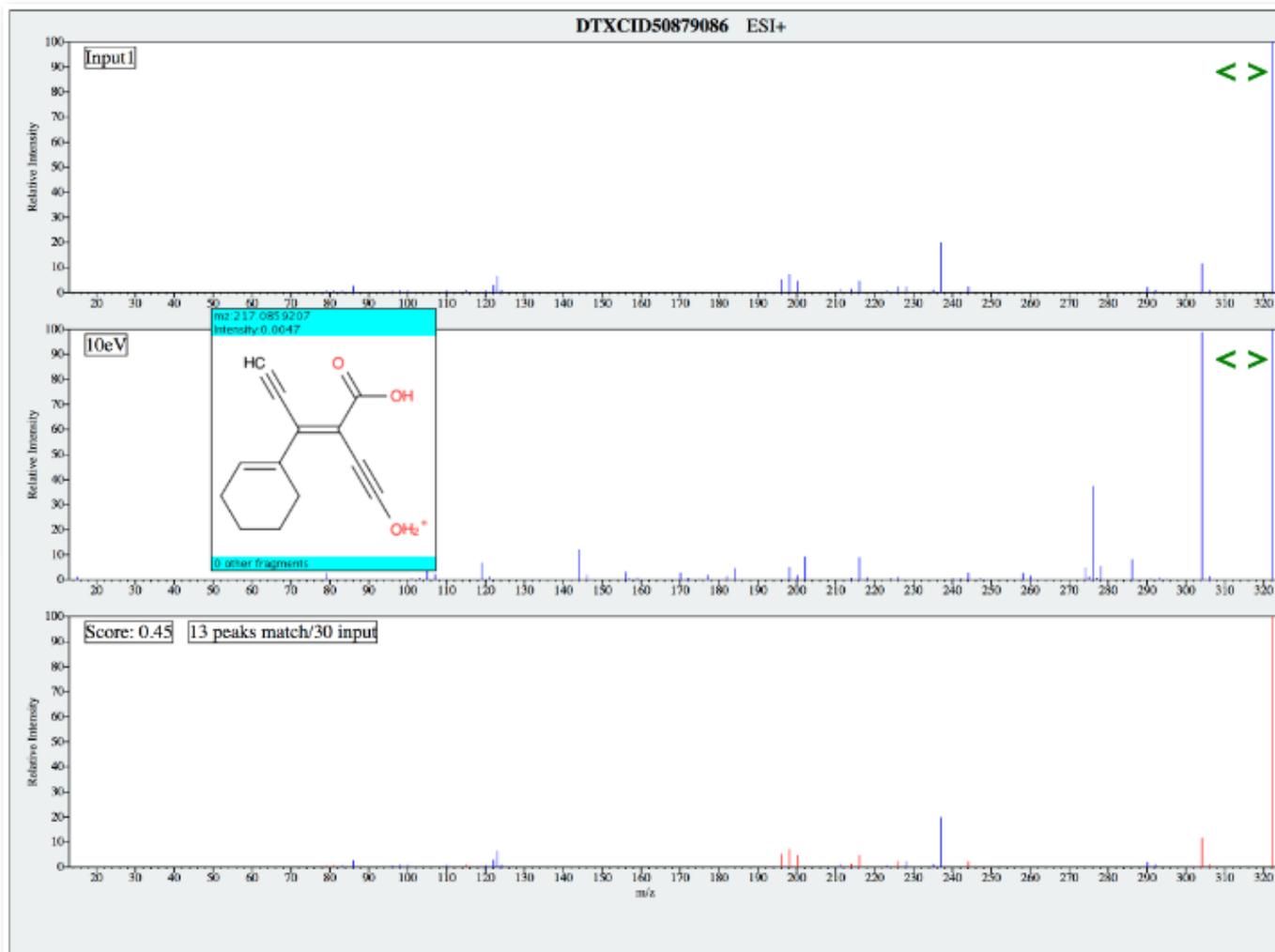
DTXCID60349982

0.11

DTXCID10316649

0.09

Spectral Viewer Comparison



Peak Window: 2.0

ppm Update Score

Input	10eV	20eV	40eV
1	0.45	0.16	0.029
2	0.45	0.16	0.029
3	0.45	0.16	0.029

Prototype Development

AADashboard

atrazine

Search

Select properties to predict

H
C
N
O

T.E.S.T. 18

OPERA

Search

- Exact
- Substructure

Search result 2540

Show Isotopically Labeled Charged Salts or Mixtures

Sort Similarity

<chem>Nc1ncncn1</chem> 1	<chem>Nc1ncncn1</chem> 0.62	<chem>Cc1ncncn1</chem> 0.57	<chem>Nc1ncncn1</chem> 0.57	<chem>Nc1ncncn1</chem> 0.57	<chem>Cc1ncncn1</chem> 0.53	<chem>Clc1ncncn1</chem> 0.53	<chem>Cc1ncncn1</chem> 0.53	<chem>Oc1ncncn1</chem> 0.5	<chem>Fc1ncncn1</chem> 0.5	<chem>Clc1ncncn1</chem> 0.5
<chem>Clc1ncncn1</chem> 0.5	<chem>Brn1cncn1</chem> 0.5	<chem>Brn1cncn1</chem> 0.5	<chem>Cc1ncncn1</chem> 0.47	<chem>Cc1ncncn1</chem> 0.44	<chem>Cc1ncncn1</chem> 0.44	<chem>Cc1ncncn1</chem> 0.44	<chem>Cc1ncncn1</chem> 0.42	<chem>Cc1ncncn1</chem> 0.42	<chem>Cc1ncncn1</chem> 0.42	<chem>Cc1ncncn1</chem> 0.42
<chem>Nc1ncncn1</chem> 0.42	<chem>Cc1ncncn1</chem> 0.42	<chem>Oc1ncncn1</chem> 0.42	<chem>Nc1ncncn1</chem> 0.42	<chem>Cc1ncncn1</chem> 0.42	<chem>Nc1ncncn1</chem> 0.4	<chem>Clc1ncncn1</chem> 0.4	<chem>Cc1ncncn1</chem> 0.4	<chem>Cc1ncncn1</chem> 0.4	<chem>Nc1ncncn1</chem> 0.4	<chem>Nc1ncncn1</chem> 0.4
<chem>Cc1ncncn1</chem> 0.4	<chem>Cc1ncncn1</chem> 0.4	<chem>Cc1ncncn1</chem> 0.4	<chem>Cc1ncncn1</chem> 0.4	<chem>Cc1ncncn1</chem> 0.4						
<chem>Cc1ncncn1</chem> 0.38	<chem>Nc1ncncn1</chem> 0.38	<chem>Nc1ncncn1</chem> 0.38	<chem>Nc1ncncn1</chem> 0.38	<chem>Cc1ncncn1</chem> 0.38	<chem>Cc1ncncn1</chem> 0.38	<chem>Cc1ncncn1</chem> 0.38				

Search result 2540

Show Isotopically Labeled

Prototype Development

atrazine Search

Select properties to predict

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

C

N

O

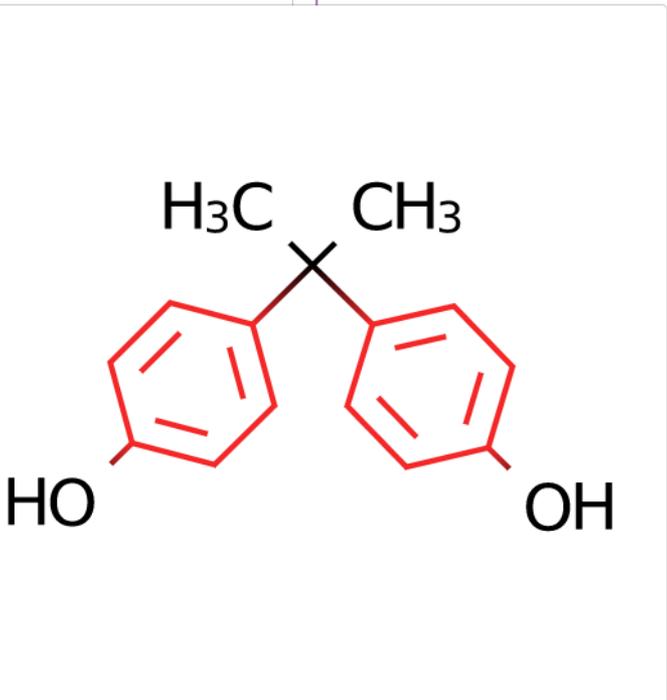
S

P

F

Cl Input formula (e.g. C6 H6):

Br C15H16O2

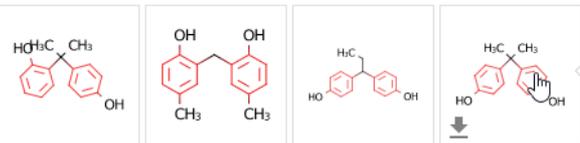


HO

OH

Search

Search result 5 Show Isotopically Labeled Cl



Elements per page 50

1

<https://comptox.epa.gov/dashboard/DTXSID7020182>

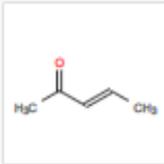
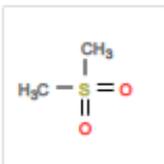
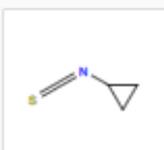
Collision Cross Section Prediction

PNNL Collision Cross Section Database

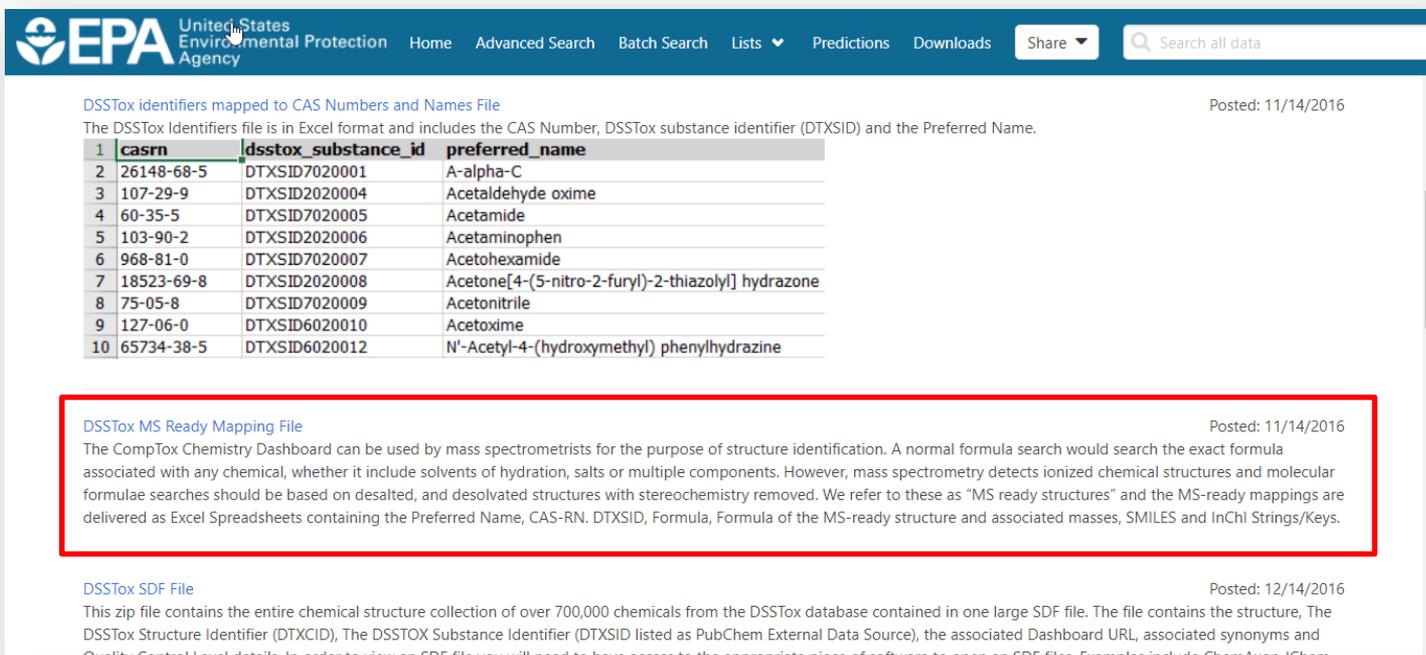


Showing 1 to 25 of 1000 rows rows per page

< 1 2 3 4 5 ... 40 >

Chemical	SMILES	InChI	Formula	Mass 	CCS (Å ²)
 <p>(3E)-pent-3-en-2-one</p>	*	*	C ₅ H ₈ O	84.0575	[M-H] ⁻ 112.1 ISiCLE Lite v0.1.0 [M+Na] ⁺ 112.6 ISiCLE Lite v0.1.0 [M+H] ⁺ 113.1 ISiCLE Lite v0.1.0
 <p>Dimethyl sulfone</p>	*	*	C ₂ H ₆ O ₂ S	94.0089	[M-H] ⁻ 106.9 ISiCLE Lite v0.1.0 [M+Na] ⁺ 107.3 ISiCLE Lite v0.1.0 [M+H] ⁺ 108.1 ISiCLE Lite v0.1.0
 <p>isothiocyanatocyclopropane</p>	*	*	C ₄ H ₅ NS	99.0143	[M-H] ⁻ 111.9 ISiCLE Lite v0.1.0 [M+Na] ⁺ 112.1 ISiCLE Lite v0.1.0 [M+H] ⁺ 110.0 ISiCLE Lite v0.1.0

- Groups waiting on our API and web services
- Mass Spec companies instrument integration
- Release will be in iterations but for now our data are available



The screenshot shows the EPA website interface. At the top, there is a navigation bar with the EPA logo, "United States Environmental Protection Agency", and links for Home, Advanced Search, Batch Search, Lists, Predictions, Downloads, and a Share button. A search bar is also present. Below the navigation bar, there are three main sections:

- DSSTox identifiers mapped to CAS Numbers and Names File** (Posted: 11/14/2016)
The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name.

1	casrn	dsstox_substance_id	preferred_name
2	26148-68-5	DTXSID7020001	A-alpha-C
3	107-29-9	DTXSID2020004	Acetaldehyde oxime
4	60-35-5	DTXSID7020005	Acetamide
5	103-90-2	DTXSID2020006	Acetaminophen
6	968-81-0	DTXSID7020007	Acetohexamide
7	18523-69-8	DTXSID2020008	Acetone[4-(5-nitro-2-furyl)-2-thiazolyl] hydrazone
8	75-05-8	DTXSID7020009	Acetonitrile
9	127-06-0	DTXSID6020010	Acetoxime
10	65734-38-5	DTXSID6020012	N'-Acetyl-4-(hydroxymethyl) phenylhydrazine
- 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 file. Examples include ChemAxon JChem

Benefiting the community with Open Data

NORMAN Suspect List Exchange

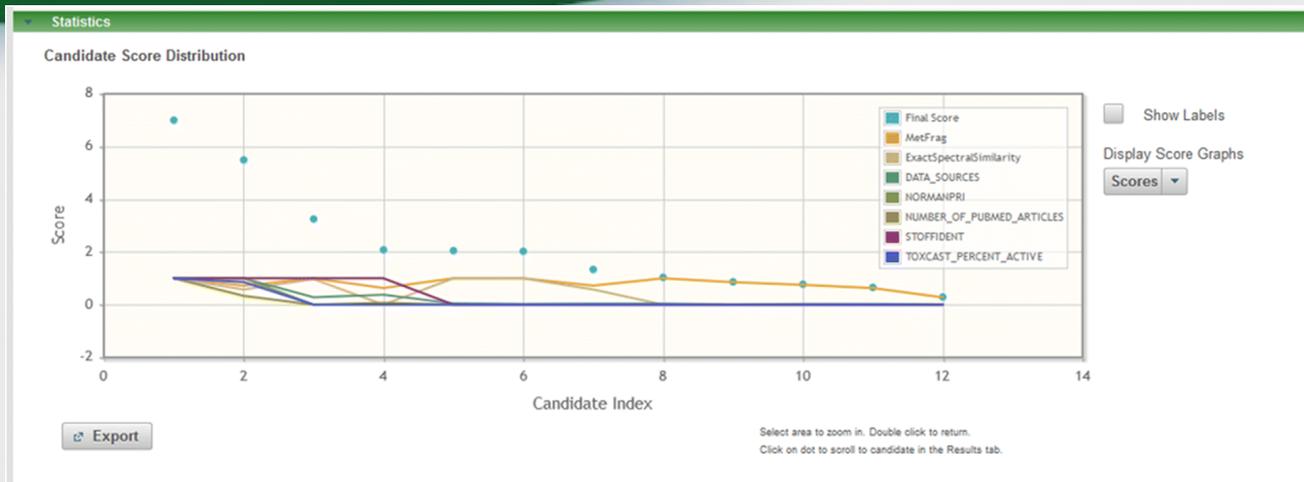
<https://www.norman-network.com/?q=node/236>

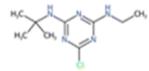
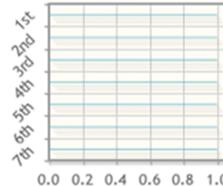
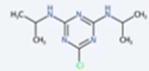
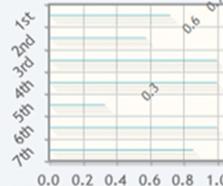
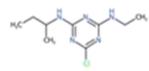
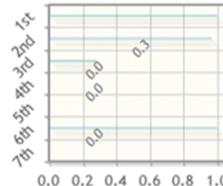


Wastewater Suspect List based on Swedish Product Data	Wastewater Suspect List Original File with Mapped DTXSIDs (12/02/2019)	KEMIWWSUS InChIKeys (12/02/2019)	A prioritized list of 1,123 substances relevant for wastewater based on Swedish product registry data, including scores. Provided by Stellan Fischer,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

Integration to MetFrag in place

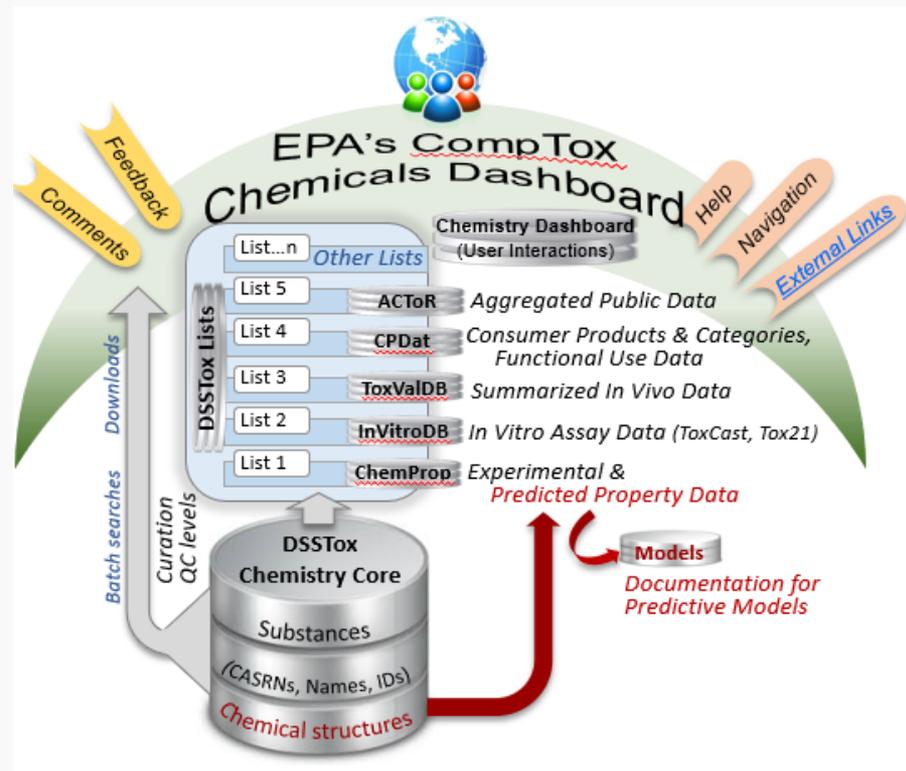
<https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0299-2>



#	Molecule	Identifier	Mass	Formula	Normalized Scores	FinalScore	Details
1	 Terbutylazine	DTXSID4027608 InChIKeyBlock1 = FZXISNSWEXTPMF	229.10948	C ₉ H ₁₆ ClN ₅		7.0	Peaks: 10 / 14 Fragments Scores Download
2	 Propazine	DTXSID3021196 InChIKeyBlock1 = WJNRPIHGGKWCK	229.10948	C ₉ H ₁₆ ClN ₅		5.4894	Peaks: 7 / 14 Fragments Scores Download
3	 Sebuthylazine	DTXSID7058171 InChIKeyBlock1 = BZRUVKZGXNSXMB	229.10948	C ₉ H ₁₆ ClN ₅		3.2476	Peaks: 10 / 14 Fragments Scores Download

Conclusion

- Dashboard access to data for ~875,000 chemicals
- “MS-Ready” structures dramatically enable searches
- Ongoing list curation expands data access
- Candidate ranking supported by integrated meta-data
- Integration of predicted spectra (ESI+/- and EI) will assist candidate ranking
- API and services will be delivered





Credit: the Research Triangle Foundation

EPA-RTP

- *An enormous team of contributors from NCCT, especially the IT software development team*
- *Our curation team for their care and focus on data quality*
- *Multiple centers and laboratories across the EPA*
- *Many public domain databases and open data contributors*

Antony Williams

NCCT, US EPA Office of Research and Development,

Williams.Antony@epa.gov

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

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

 Journal of Cheminformatics

DATABASE

Open Access

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



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

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