

What chemicals constitute the Exposome? Accessing data via the US EPA's CompTox Chemicals Dashboard

Antony Williams

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

...and an enormous cast of characters

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

CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard



United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

875 Thousand Chemicals

Chemicals

Product/Use Categories

Assay/Gene

SEARCH

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

Identifier substring search

United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

BATCH SEARCH

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

Select Input Type(s)

☒ Identifiers

☒ Chemical Name(s)

☐ CASRN(s)

☐ InChIkey(s)

☐ DDTXSID Substring(s)

☐ DDTXSID Compound(s)

☐ InChIkey Substring(s)

☐ MS-Ready Formula(s)

☐ Exact Formula(s)

☐ Monoisotopic Mass(s)

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

107-26-2
88-39-2
14024-55-6
33468-78-5
7784-40-9
1625-40-9
60332-96-5
2122-76-5
126833-17-8
7786-81-4

Display All Chemicals

Download Chemical Data

Select Output Format

Excel

Download

Customize Results

☐ Select All

☐ Select All in Lists

Chemical Identifiers

☒ Chemical Name(s)

☐ DDTXSID

☐ Chemical Name(s)

☐ DDTXSID

Presence in Lists:

☐ Pharmaceuticals from DDTXSID

☐ ACRFSSS Extremely Hazardous Substance List and Threshold Planning Quantities

☐ AEDS Acute Exposure Guideline Levels

☐ Amphibols Minerals

☐ AROGENS Androgen Receptor Chemistry

United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

TOX DATA

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

Hazard

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. RATE/TRANSPORT

HAZARD

ACME

EXPOSURE

BIODIVERSITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

Data Type

Toxicity Value

Download

Columns

Max	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species	Substance	Source
MES	Short-term	Critical Air	short-term	500	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	OOD
MES	Short-term	Marginal Air	short-term	100	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	OOD
MES	Short-term	Negligible Air	short-term	15	mg/m3	-	inhalation	-	TG 230 Military Exposure Guidelines Table	OOD
MES	Soil	Negligible Soil	chronic	100000	mg/kg	-	soil	-	TG 230 Military Exposure Guidelines Table	OOD
MES	Long-term	Soil Negligible Water	chronic	7	mg/L	-	oral	-	TG 230 Military Exposure Guidelines Table	OOD

United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

TOX DATA

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

Chemical Activity Summary

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. RATE/TRANSPORT

HAZARD

ACME

EXPOSURE

BIODIVERSITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

TOXCAST DATA

Assay Description: 742
Gene Symbol: E2F1
Organism: human
Tissue: Liver
Assay Format: Type: cell-based
Biological Process: Target: protein stabilization
Detection Technology: Vision: Target Complementarity
Analysis Direction: positive
Intended Target Family: nuclear receptor
Description: Data from this assay component OT_ER_RaRa_D400 was analyzed into 1 assay endpoint.
OT_ER_RaRa_D400 was analyzed in the positive filtering direction relative to DMSO as the negative control and baseline of activity. Using a type of binding receptor, measure of receptor for gene of signal activity can be used to understand the binding at the pathway level as this relates to the gene ER. 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 suitable targets, this assay endpoint is associated to the "nuclear receptor" intended target family, where the subfamily is "nuclear".

United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

SIMILARITY

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

Searched with a similarity threshold of 0.8

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. RATE/TRANSPORT

HAZARD

ACME

EXPOSURE

BIODIVERSITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

376 of 390 chemicals visible

4-Compound
CASRN: 80-05-7
DTXSID: DTXSID7020182
TOXCAST: 7020182

4-(1,1-diphenyl)ethanol
CASRN: 80-05-7
DTXSID: DTXSID7020182
TOXCAST: 7020182

4-(1,1-diphenyl)ethanol
CASRN: 80-05-7
DTXSID: DTXSID7020182
TOXCAST: 7020182

4-(1,1-diphenyl)ethanol
CASRN: 80-05-7
DTXSID: DTXSID7020182
TOXCAST: 7020182

4-(1,1-diphenyl)ethanol
CASRN: 80-05-7
DTXSID: DTXSID7020182
TOXCAST: 7020182

United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

TOX DATA

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

Chemical Activity Summary

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. RATE/TRANSPORT

HAZARD

ACME

EXPOSURE

BIODIVERSITY

SIMILAR COMPOUNDS

GENRA (BETA)

RELATED SUBSTANCES

TOXCAST DATA

Assay Description: 742
Gene Symbol: E2F1
Organism: human
Tissue: Liver
Assay Format: Type: cell-based
Biological Process: Target: protein stabilization
Detection Technology: Vision: Target Complementarity
Analysis Direction: positive
Intended Target Family: nuclear receptor
Description: Data from this assay component OT_ER_RaRa_D400 was analyzed into 1 assay endpoint.
OT_ER_RaRa_D400 was analyzed in the positive filtering direction relative to DMSO as the negative control and baseline of activity. Using a type of binding receptor, measure of receptor for gene of signal activity can be used to understand the binding at the pathway level as this relates to the gene ER. 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 suitable targets, this assay endpoint is associated to the "nuclear receptor" intended target family, where the subfamily is "nuclear".

United States Environmental Protection Agency

HomeAdvanced SearchBatch SearchLists▼PredictionsDownloads

TOX DATA

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

Chemical Activity Summary

DETAILS

EXECUTIVE SUMMARY

PROPERTIES

ENV. RATE/TRANSPORT

HAZARD

ACME

EXPOSURE

BIODIVERSITY

SIMILAR COMPOUNDS

GENRA (BETA)

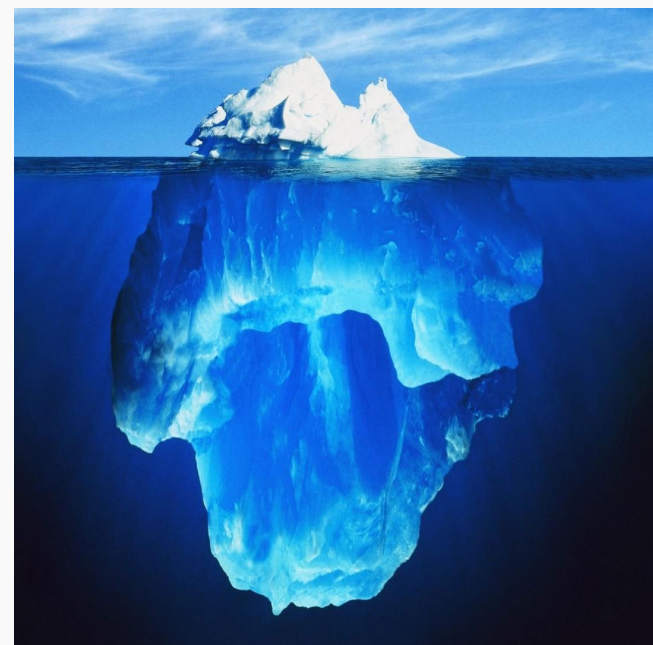
RELATED SUBSTANCES

TOXCAST DATA

Assay Description: 742
Gene Symbol: E2F1
Organism: human
Tissue: Liver
Assay Format: Type: cell-based
Biological Process: Target: protein stabilization
Detection Technology: Vision: Target Complementarity
Analysis Direction: positive
Intended Target Family: nuclear receptor
Description: Data from this assay component OT_ER_RaRa_D400 was analyzed into 1 assay endpoint.
OT_ER_RaRa_D400 was analyzed in the positive filtering direction relative to DMSO as the negative control and baseline of activity. Using a type of binding receptor, measure of receptor for gene of signal activity can be used to understand the binding at the pathway level as this relates to the gene ER. 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 suitable targets, this assay endpoint is associated to the "nuclear receptor" intended target family, where the subfamily is "nuclear".

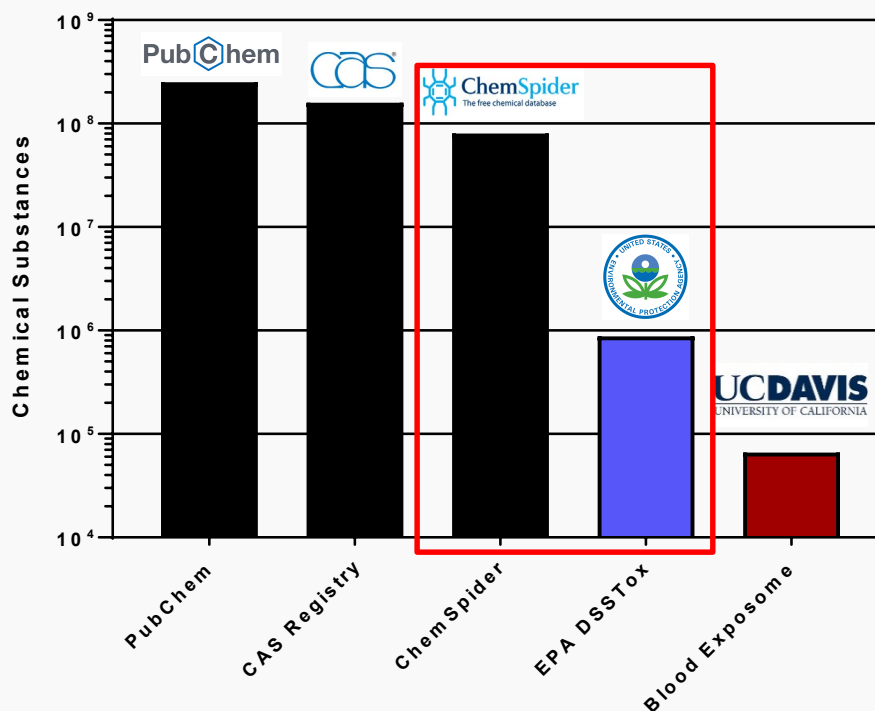
How many chemicals to consider?

- Dashboard content is small but focused curated content
- **Dashboard** ~900,000
- ChemSpider ~90,000,000
- PubChem ~111,000,000
- CAS ~ 165,000,000



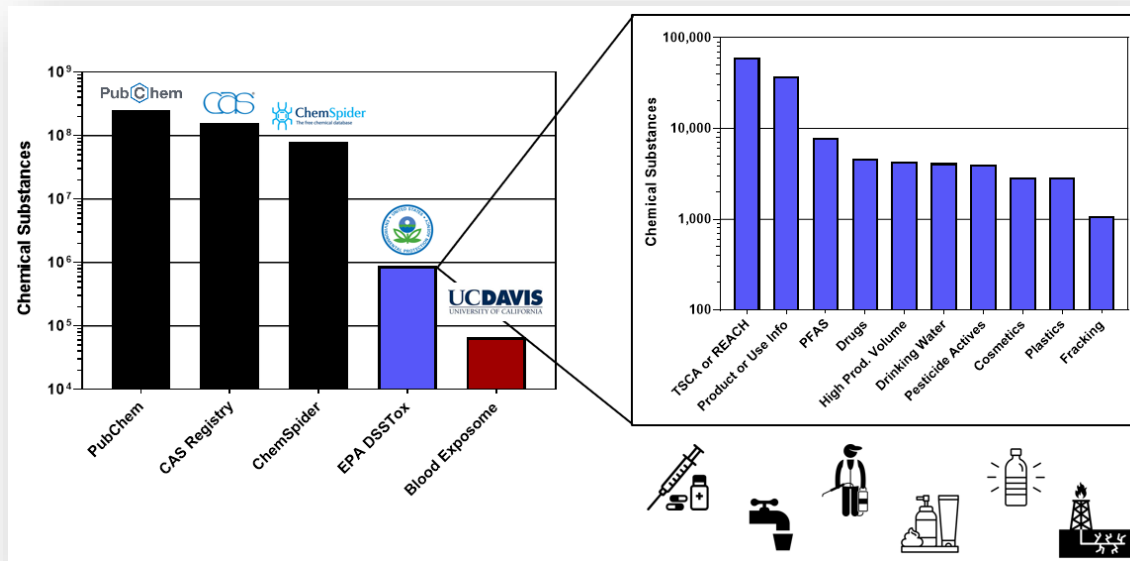
BIG databases are GREAT!

- Thanks to all of the public database efforts
- So much benefit from what's been done
- There are hundreds of them at this point...



What chemicals constitute the Exposome?

- What constitutes the exposome is, of course, difficult to answer.
- Focused chemical list expansion highlights chemical classes of interest – chemicals in commerce, pesticides, cosmetics, PFAS



No “large library” dilution

- The top 6 data sources on ChemSpider are vendor collections

ChemSpider

Search and share chemistry

Search ChemSpider

Simple Structure Advanced History

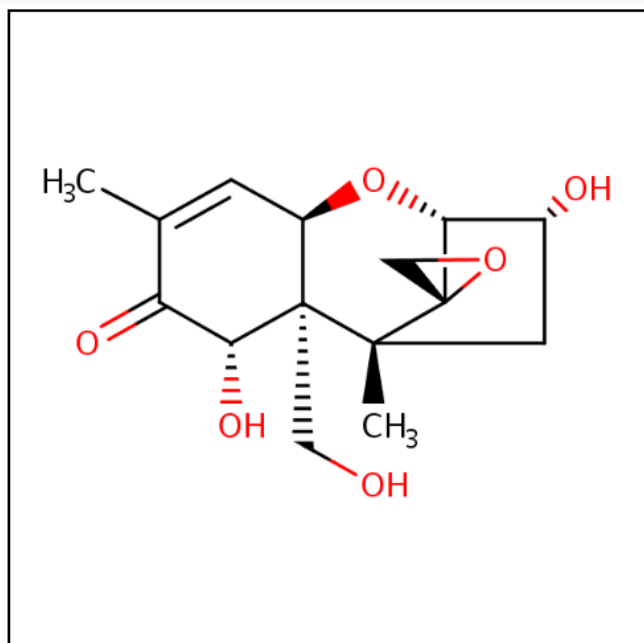
<

Data Sources

<u>Data Source</u>	<u>Count</u>	<u>Date Created</u>	<u>Last Updated</u>
Aurora Fine Chemicals	61257202	13/04/2009	01/09/2020
Chemspace	14283334	30/11/2016	04/12/2018
AKos	12326390	15/04/2008	09/10/2017
Mcule	9299742	21/01/2014	26/10/2018
Molport	8389921	09/02/2010	15/06/2020
LabNetwork	3078080	28/04/2016	03/03/2020

Data Quality Dilution

Vomitoxin



Structural Identifiers

 **IUPAC Name:** (3 α ,7 α)-3,7,15-Trihydroxy-12,13-epoxytrichothec-9-en-8-one

 **SMILES:** CC1=C[C@H]2O[C@@H]3[C@H](O)C[C@@](C)([C@]33CO3)[C@@]2(CO)[C@H]1O

 **InChI String:** InChI=1S/C15H20O6/c1-7-3-9-14(5-16,11(19)10(7)18)13(2)4-8(17)12(20)2H3/t8-,9-,11-,12-,13-,14-,15+/m1/s1

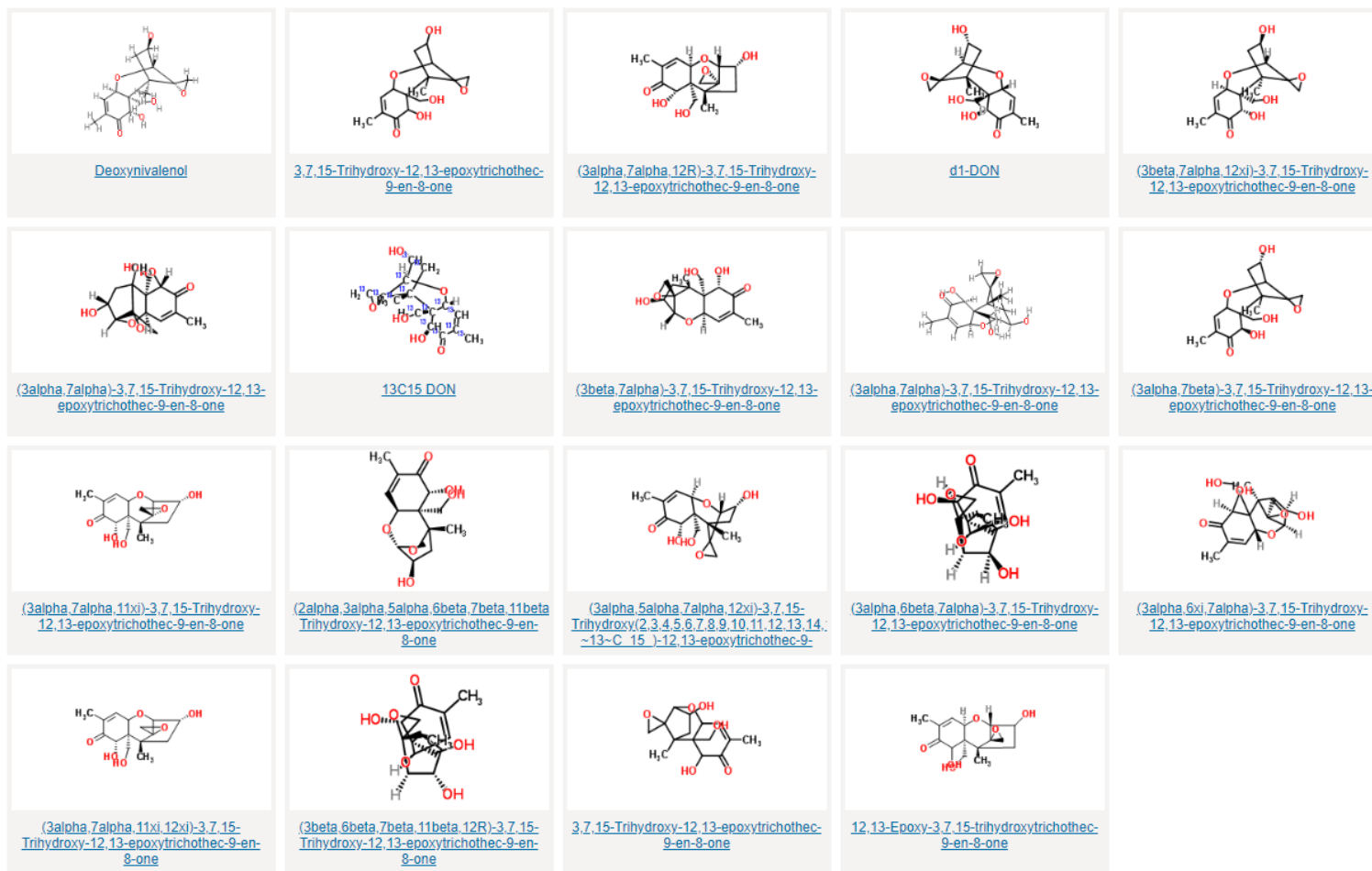
 **InChIKey:** LINOMUASTDIRTM-QGRHZQQGSA-N

Search Google for:

 Copy All

- 19 “Vomitoxins” – 3 isotopically labeled

Search term: **LINOMUASTDIRTM** (Found by InChIKey (skeleton match))



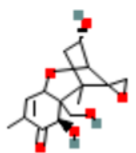
- 33 unique InChI Keys

Compounds
(33)

Substances
(10)

Searching chemical names and synonyms including IUPAC names and InChIKeys across the compound collection. Note that pages is not searched. [Read More...](#)

33 results Filters SORT BY Relevance



Vomitoxin; Trichothec-9-En-8-One, 12,13-Epoxy-3,7,15-Trihydroxy-, (3.Α.Α.,7.Α.Α.)-; 12,13-Epoxy-3.Α.Α.,7.Α.Α.,15-Trihydroxy-9-Trichothecen-8-One; LINOMUASTDIRTM-LMJBVPRVSA-N; Spiro[2,5-Methano-1-Benzoxepin-10,2'-Oxirane], Trichothec-9-En-8-One Deriv.

Compound CID: 6432495

MF: C₁₅H₂₀O₆ MW: 296.31g/mol

InChIKey: LINOMUASTDIRTM-LMJBVPRVSA-N

IUPAC Name: (3R,10S)-3,10-dihydroxy-2-(hydroxymethyl)-1,5-dimethylspiro[8-oxatricyclo[7.2.1.02,7]dodec-5-ene-12,2'-oxirane]-4-one

Create Date: 2006-04-28

Not Vomitoxin

PubChem – “virtual chemistry”

Large Databases – collapse

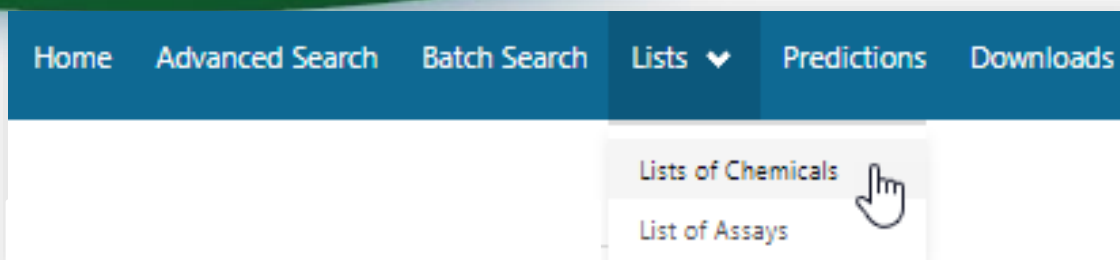
- Lots of “virtual chemistry” and “MODs”
- Vomitoxin has 7 ZINC stereoforms.

PUBCHEM_CID	Compound_Name	Compound_Synonym	InChIKey
98043267	(1R,2S,3R,7R,9S,10R,12S)	ZINC100006545	LINOMUASTDIRTM-DOZBXCHUSA-N
98051113	(1R,2R,3S,7S,9S,10R,12S)	ZINC100066010	LINOMUASTDIRTM-KCWNRFLPSA-N
100853641	(1R,2R,3S,7S,9S,10R,12R)	ZINC229762267	LINOMUASTDIRTM-OMTHLLQNSA-N
98043268	(1R,2S,3R,7S,9S,10R,12S)	ZINC100006546	LINOMUASTDIRTM-UBTIPYQWSA-N
95566296	(1R,2S,3R,7R,9R,10R,12S)	ZINC71789640	LINOMUASTDIRTM-WYQUPHEGSA-N
100853642	(1R,2R,3S,7R,9S,10R,12R)	ZINC229762273	LINOMUASTDIRTM-XFRIDARHSA-N
95566297	(1R,2S,3R,7S,9R,10R,12S)	ZINC71789642	LINOMUASTDIRTM-XGQZSAOASA-N

- Care is needed to choose subsets of interest and Emma Schymanski *et al* are doing this

Chemical Lists

~275 Chemical Lists (and growing)



Download Columns mass Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
HDXEXCH	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - Under HDX Conditions	2018-11-07	592	Observed species (deuterated and undeuterated) from the HDXNOEX list under hydrogen deuterium exchange conditions (Ruttkies, Schymanski et al. in prep.)
HDXNOEX	MASSPECDB: Hydrogen Deuterium Exchange Standard Set - No Exchange	2018-11-07	765	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule high resolution mass spectrometry (Ruttkies, Schymanski et al. in prep.)
MASSBANKEUSP	MASSPECDB: MassBank.EU Collection: Special Cases	2017-07-16	263	The MassBank.EU list contains curated chemicals (Schymanski/Williams) associated with the literature/tentative/unknown/SI spectra available on MassBank.EU that are not available as part of the full MassBank collection of reference standard spectra.
MASSBANKREF	MASSPECDB: MassBank Reference Spectra Collection	2017-07-13	1267	This MassBank list contains chemicals associated with the full MassBank collection of reference standard spectra available on MassBank.EU, MassBank.JP and MassBank of North America as well as the Open Data collection, curated by Williams/Schymanski.
MYCOTOXINS	MASSPECDB: Mycotoxins from MassBank.EU	2017-08-02	88	This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to MassBank.EU by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada

Public TSCA Inventory on Dashboard 33,364 Chemicals (9/6/2020)

EPA|TSCA: TSCA Inventory, active non-confidential portion

☐ Identifier substring search

List Details

Description: Section 8 (b) of the Toxic Substances Control Act (TSCA) requires EPA to compile, keep current and publish a list of each chemical substance that is manufactured or processed, including imports, in the United States for uses under TSCA. Information about what types of substances are on the TSCA inventory can be found here. The Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, requires EPA to designate chemical substances on the TSCA Chemical Substance Inventory as either "active" or "inactive" in U.S. commerce. To accomplish this, EPA finalized a rule requiring industry reporting of chemicals manufactured (including imported) or processed in the U.S.. This reporting is used to identify which chemical substances on the TSCA Inventory are active in U.S. commerce and help inform the prioritization of chemicals for risk evaluation. The list contained in the dashboard includes the active TSCA inventory based on notifications through Feb. 7th 2018 and substances reported from Feb 8, 2018 – March 30, 2018 that have been unambiguously mapped to DSSTox using CASRN and chemical names. The curation of the non-confidential portion of active TSCA inventory is an ongoing process involving trained chemists to validate the correctness of DSSTox structural and identifier data. The content of the list will change over time as the non-confidential active TSCA inventory is updated and more substances are curated. (Updated January 5th 2020)

Number of Chemicals: 31460

2250 of 31460 chemicals loaded

Select all

Download

Send to Batch Search

Default

↑

CASRN

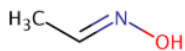
DTXSID

▼

Hide chemicals that are:

Filter by Name or CASRN

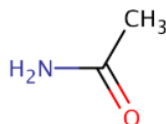
≡



Acetaldehyde oxime

CASRN:107-29-9

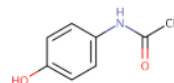
DTXSID:DTXSID2020004



Acetamide

CASRN:60-35-5

DTXSID:DTXSID7020005



Acetaminophen

CASRN:103-90-2

DTXSID:DTXSID2020006



Acetonitrile

CASRN:75-05-8

DTXSID:DTXSID7020009

Hydraulic Fracturing (1640)

EPA's Study of Hydraulic Fracturing and Its Potential Impact on Drinking Water Resources

[Contact Us](#)

Hydraulic Fracturing Study
Home

Final Assessment

EPA Published Research

Fact Sheets

Questions & Answers about
the final assessment

Multi-agency collaboration
on unconventional oil and
gas research

EPA Hydraulic Fracturing -
Agency Main Page

Hydraulic Fracturing For Oil And Gas: Impacts From The Hydraulic

WATER|EPA; Chemicals associated with hydraulic fracturing

Search EPAHFR Chemicals

Identifier substring search

List Details

Description: Chemicals used in hydraulic fracturing fluids and/or identified in produced water from 2005-2013, corresponding to chemicals listed in Appendix H of EPA's Hydraulic Fracking Drinking Water Assessment Final Report (Dec 2016). Citation: U.S. EPA, Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report). U.S. Environmental Protection Agency, Washington, D.C. EPA/600/R-16/236F, 2016. <https://www.epa.gov/hfstudy>

*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances (as of Jan 2018), both were included in the current EPAHFR chemical listing for completeness; additionally, 34 previously unmapped chemicals in Table H-5 are now registered in DSSTox (all but 2 assigned CASRN) and, thus, have been added to the current EPAHFR listing.

Number of Chemicals: 1640

Select all

Download

Send to Batch Search

Default

↑

CASRN

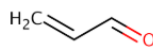
DTXSID

1640 chemicals

Hide chemicals that are:

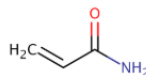
Filter by Name or CASRN

☰



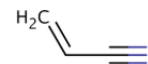
Acrolein

CASRN:107-02-8
DTXSID:DTXSID5020023



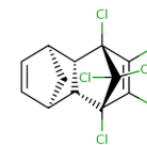
Acrylamide

CASRN:79-06-1
DTXSID:DTXSID5020027



Acrylonitrile

CASRN:107-13-1
DTXSID:DTXSID5020029



Aldrin

CASRN:309-00-2
DTXSID:DTXSID8020040

Tire Crumb Rubber (298)

Related Topics: [Safer Chemicals Research](#)

[CONTACT US](#)

SHARE



July 2019 Report: Tire Crumb Rubber Characterization

Key Takeaways:

- EPA is releasing a new report that addresses exposure (that is, chemicals and how people come in contact with these) to tire crumb rubber on synthetic turf fields. **This report is not a risk assessment** nor can the information be used to identify a level above which health effects could occur.
- In general, the findings from the study suggest that human exposure appears to be low.
- Only Part 1 is being released for public comment and assessment.
- Part 1 of this report presents the chemical constituents identified in the State-of-Science Literature Review/Gaps Analysis, White Paper Summary of Results. Eleven sources of publicly available toxicity reference information were searched. It is important to recognize that not all potential chemical constituents identified through the literature search were confirmed through measurements made under the Federal Research Action Plan.
- The scope of this study was limited to chemicals identified in the literature search.

Tire Crumb Rubber

Search TIRECRUMB Chemicals

☐ Identifier substring search

List Details

Description: This chemical list is based on data contained within the [Federal Research Action Plan \(FRAP\) on Recycled Tire Crumb Used on Playing Fields and Playgrounds](#). The chemical list is obtained from the [Toxicity reference information spreadsheet](#) compiled for the potential tire crumb rubber chemical constituents identified in the State-of-Science Literature Review/Gaps Analysis, White Paper Summary of Results. Eleven sources of publicly available toxicity reference information were searched. It is important to recognize that not all potential chemical constituents identified through the literature search were confirmed through measurements made under the Federal Research Action Plan.

Number of Chemicals: 298

Select all

Download

Send to Batch Search

Default



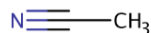
CASRN

DTXSID

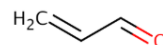
298 chemicals

Hide chemicals that are:

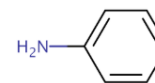
Filter by Name or CASRN



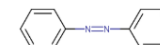
Acetonitrile
CASRN:75-05-8
DTXSID:DTXSID7020009



Acrolein
CASRN:107-02-8
DTXSID:DTXSID5020023



Aniline
CASRN:62-53-3
DTXSID:DTXSID8020090



Azobenzene
CASRN:103-33-3
DTXSID:DTXSID8020123

Disinfection By-Products



TrAC Trends in Analytical Chemistry

Volume 22, Issue 10, November 2003, Pages 666-684



Trends

Disinfection by-products and other emerging contaminants in drinking water

Susan D. Richardson



Home Advanced Search Batch Search Lists Predictions Downloads

Share

Search all data

LIST: Disinfection By-products (Richardson et al)

Search DBP Chemicals

Identifier substring search

List Details

Description: A list of disinfection by-products associated with RICHARDSON, S. D. Disinfection By-Products: Formation and Occurrence in Drinking Water. Chapter 2, J.O. Nriagu (ed.), Encyclopedia of Environmental Health, Elsevier Science Inc., Burlington, MA, 2:110-136, (2011).

Number of Chemicals: 619

619 chemicals

Select all

Download

Send to Batch Search

Default

↑

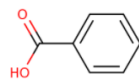
DTXSID

CASRN

TOXCAST

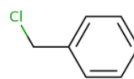
Hide chemicals that are:

Filter by Name or CASRN



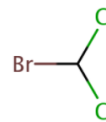
Benzoic acid

DTXSID:DTXSID6020143
CASRN:65-85-0
TOXCAST:13/857



Benzyl chloride

DTXSID:DTXSID0020153
CASRN:100-44-7
TOXCAST:3/235



Bromodichloromethane

DTXSID:DTXSID1020198
CASRN:75-27-4
TOXCAST:0/235



2-Bromo-1-ethanol

DTXSID:DTXSID8020200
CASRN:540-51-2
TOXCAST:5/235

PFAS lists of Chemicals

Select List

Download

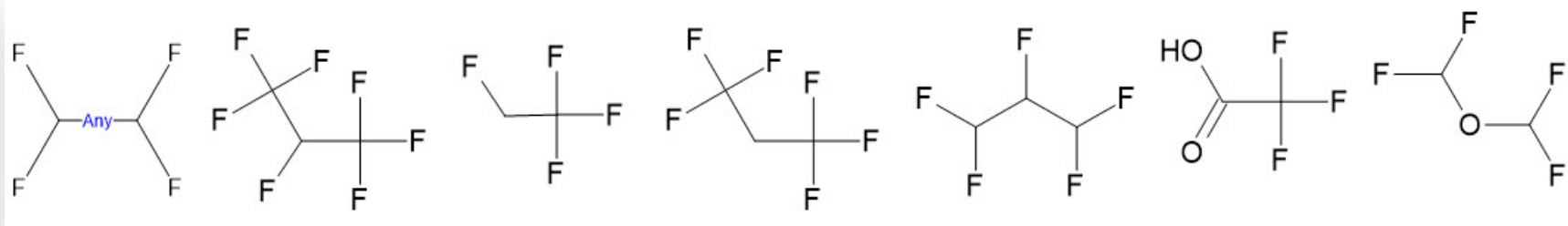
Columns

PFAS

Copy Filtered Lists URL

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

PFAS Structure List (8163)



EPA United States Environmental Protection Agency

Home Advanced Search Batch Search Lists Predictions Downloads

Share Search all data

PFAS|EPA: PFAS structures in DSSTox (update August 2020)

Search PFASSTRUCT Chemicals
☐ Identifier substring search

List Details

Description: List consists of all DTXSID records with a structure assigned, and using a set of substructural filters based on community input. The substructural filters ([visible here](#)) are designed to be simple, reproducible and transparent, yet general enough to encompass the largest set of structures having sufficient levels of fluorination to potentially impart PFAS-type properties.

Number of Chemicals: 8163

Select all Download Send to Batch Search Name

8163 chemicals

DTXSID CASRN TOXCAST

Hide chemicals that are: Filter by Name or CASRN

(((Perfluorodecyl)ethylsulphonyl)methyl)oxirane
DTXSID:DTXSID70880957
CASRN:72276-05-2
TOXCAST:-

((2,2,3,3-Tetrafluoropropoxy)methyl)oxirane
DTXSID:DTXSID70880230
CASRN:19932-26-4
TOXCAST:-

((Perfluoro-11-methyldodecyl)ethyl) dihydroxyacetate
DTXSID:DTXSID60240985
CASRN:94200-57-4
TOXCAST:-

((Perfluoro-13-methyltetradecyl)ethyl) acetate
DTXSID:DTXSID80238647
CASRN:91615-22-4
TOXCAST:-

Blood Exposome Curation in Process

Vol. 127, No. 9 | Research

Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach

Dinesh Kumar Barupal  and Oliver Fiehn 

Published: 26 September 2019 | CID: 097008 | <https://doi.org/10.1289/EHP4713> | Cited by: 1

- Blood exposome data collection from Barupal and Fiehn. We are reviewing and curating.

LIST: BLOODEXPOSOME

☐ Identifier substring search

List Details

Description: Barupal et al report on [Generating the Blood Exposome Database Using a Comprehensive Text Mining and Database Fusion Approach](#). The database can be used for prioritizing chemicals for systematic reviews, developing target assays in exposome research, identifying compounds in untargeted mass spectrometry, and biological interpretation in metabolomics data.

Number of Chemicals: 19867

5000 of 19867 chemicals loaded

Select all Download Send to Batch Search Default DTXSID CASRN TOXCAST Hide chemicals that are: Filter by N

“UVCB” Chemicals

Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB Substance) on the TSCA Inventory

This paper is a compendium of information related to the broad class of chemical substances referred to as UVCBs for the Toxic Substances Control Act (TSCA) Chemical Substance Inventory. These chemical substances cannot be represented by unique structures and molecular formulas.

Public TSCA Inventory on Dashboard 31,460 Chemicals (1/24/2020)

EPA|TSCA: TSCA Inventory, active non-confidential portion

Search TSCAACTIVENONCONF Chemicals

☐ Identifier substring search

List Details

Description: Section 8 (b) of the Toxic Substances Control Act (TSCA) requires EPA to compile, keep current and publish a list of each chemical substance that is manufactured or processed, including imports, in the United States for uses under TSCA. Information about what types of substances are on the TSCA inventory can be found here. The Toxic Substances Control Act (TSCA), as amended by the Frank R. Lautenberg Chemical Safety for the 21st Century Act, requires EPA to designate chemical substances on the TSCA Chemical Substance Inventory as either "active" or "inactive" in U.S. commerce. To accomplish this, EPA finalized a rule requiring industry reporting of chemicals manufactured (including imported) or processed in the U.S.. This reporting is used to identify which chemical substances on the TSCA Inventory are active in U.S. commerce and help inform the prioritization of chemicals for risk evaluation. The list contained in the dashboard includes the active TSCA inventory based on notifications through Feb. 7th 2018 and substances reported from Feb 8, 2018 – March 30, 2018 that have been unambiguously mapped to DSSTox using CASRN and chemical names. The curation of the non-confidential portion of active TSCA inventory is an ongoing process involving trained chemists to validate the correctness of DSSTox structural and identifier data. The content of the list will change over time as the non-confidential active TSCA inventory is updated and more substances are curated. (Updated January 5th 2020)

Number of Chemicals: 31460

2250 of 31460 chemicals loaded

Select all

Download

Send to Batch Search

Default

↑

CASRN

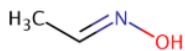
DTXSID

▼

Hide chemicals that are:

Filter by Name or CASRN

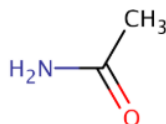
≡



Acetaldehyde oxime

CASRN:107-29-9

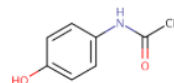
DTXSID:DTXSID2020004



Acetamide

CASRN:60-35-5

DTXSID:DTXSID7020005



Acetaminophen

CASRN:103-90-2

DTXSID:DTXSID2020006



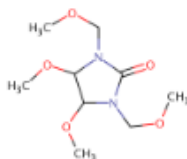
Acetonitrile

CASRN:75-05-8

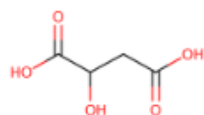
DTXSID:DTXSID7020009

Many Chemicals are “Complex”

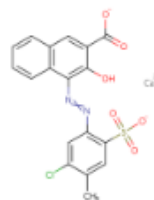
>14000 chemicals are UVCBs



2-Imidazolidinone, 4,5-dimethoxy-1,3-bis(methoxymethyl)-
DTXSID: DTXSID0027569
PubChem: 24
CASRN: 4356-80-9



Malic acid
DTXSID: DTXSID0027640
PubChem: 273
CASRN: 6915-15-7



C.I. Pigment Red 48, calcium salt (1:1)
DTXSID: DTXSID0027642
PubChem: 0
CASRN: 7023-81-2

0 related chemical
structures with this
substance

Lard, oil
DTXSID: DTXSID0027690
PubChem: 0
CASRN: 8016-28-2

0 related chemical
structures with this
substance

Tall-oil pitch
DTXSID: DTXSID0027692
PubChem: 0
CASRN: 8016-81-7

0 related chemical
structures with this
substance

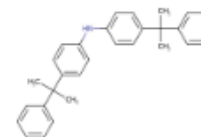
Palm kernel oil
DTXSID: DTXSID0027694
PubChem: 0
CASRN: 8023-79-8

0 related chemical
structures with this
substance

Tallow, hydrogenated
DTXSID: DTXSID0027696
PubChem: 0
CASRN: 8030-12-4

0 related chemical
structures with this
substance

Quaternary ammonium compounds, tri...
DTXSID: DTXSID0027698
PubChem: 0
CASRN: 8030-78-2



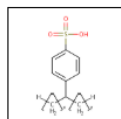
4-(2-Phenylpropan-2-yl)-N-[4-(2-phenyl...
DTXSID: DTXSID0027721
PubChem: 50
CASRN: 10081-67-1

1 related chemical
structure with this
substance

Isomethyltetrahydrophthalic anhydride
DTXSID: DTXSID0027729
PubChem: 0
CASRN: 11070-44-3

Representing complexity

Markush representations



Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

Searched by DSSTox Substance Id.

15 of 25 chemicals selected

Deselect all

Download

Send to Batch Search

Relationship

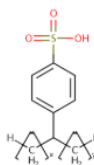
CASRN

DTXSID

Unselected

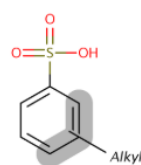
Filter by Name or CASRN

Searched Chemical



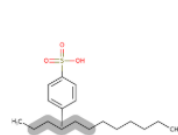
Alkylbenzenesulfonate, linear
CASRN:42615-29-2
DTXSID:DTXSID3020041

Component



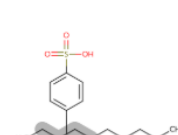
(C10-C16) Alkylbenzenesulfonic acid
CASRN:68584-22-5
DTXSID:DTXSID2028723

Component



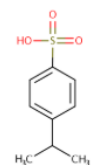
C12-linear alkyl benzene sulfonate
CASRN:NOCAS_891641
DTXSID:DTXSID90891641

Component



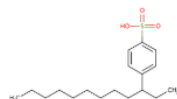
C10-linear alkylbenzenesulfonate
CASRN:NOCAS_891689
DTXSID:DTXSID70891689

Markush Child



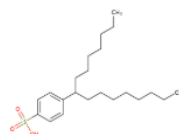
4-Isopropylbenzenesulfonic acid
CASRN:16066-35-6
DTXSID:DTXSID1044932

Markush Child



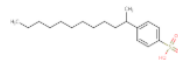
4-(3-Dodecyl)benzenesulfonic acid
CASRN:18777-54-3
DTXSID:DTXSID7058670

Markush Child



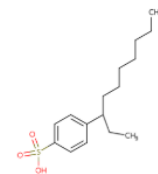
4-(1-Heptylnonyl)benzenesulfonic acid
CASRN:80233-94-9
DTXSID:DTXSID40273953

Markush Child



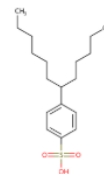
2-Phenyldodecane-p-sulfonate
CASRN:18777-53-2
DTXSID:DTXSID40274021

Markush Child



4-(Decan-3-yl)benzene-1-sulfonic acid
CASRN:65186-00-7
DTXSID:DTXSID20859618

Markush Child

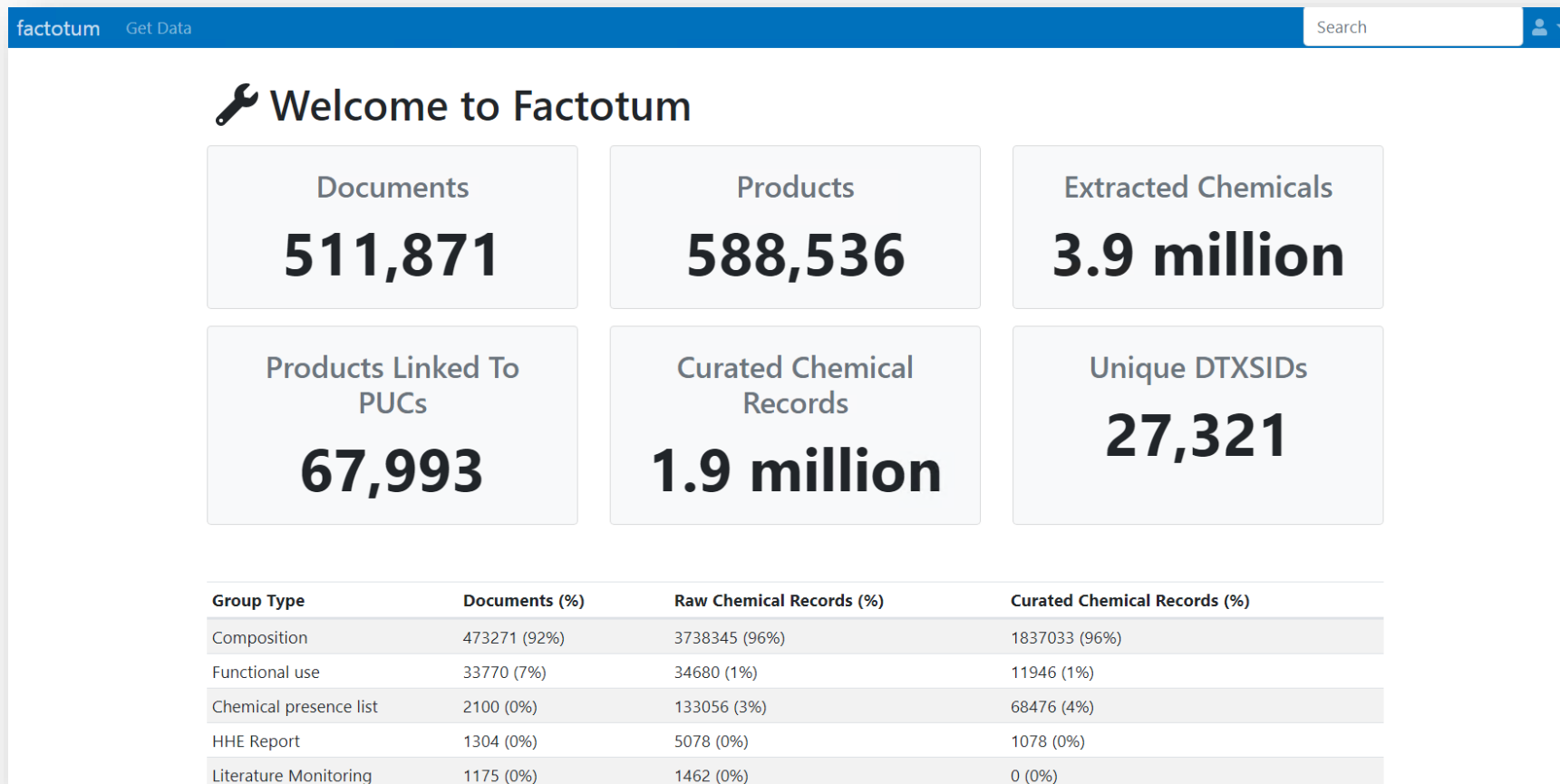


4-(Dodecan-6-yl)benzene-1-sulfonic acid
CASRN:23003-92-1
DTXSID:DTXSID30860093

Chemicals in Commerce

Factotum Database

Consumer Products Database



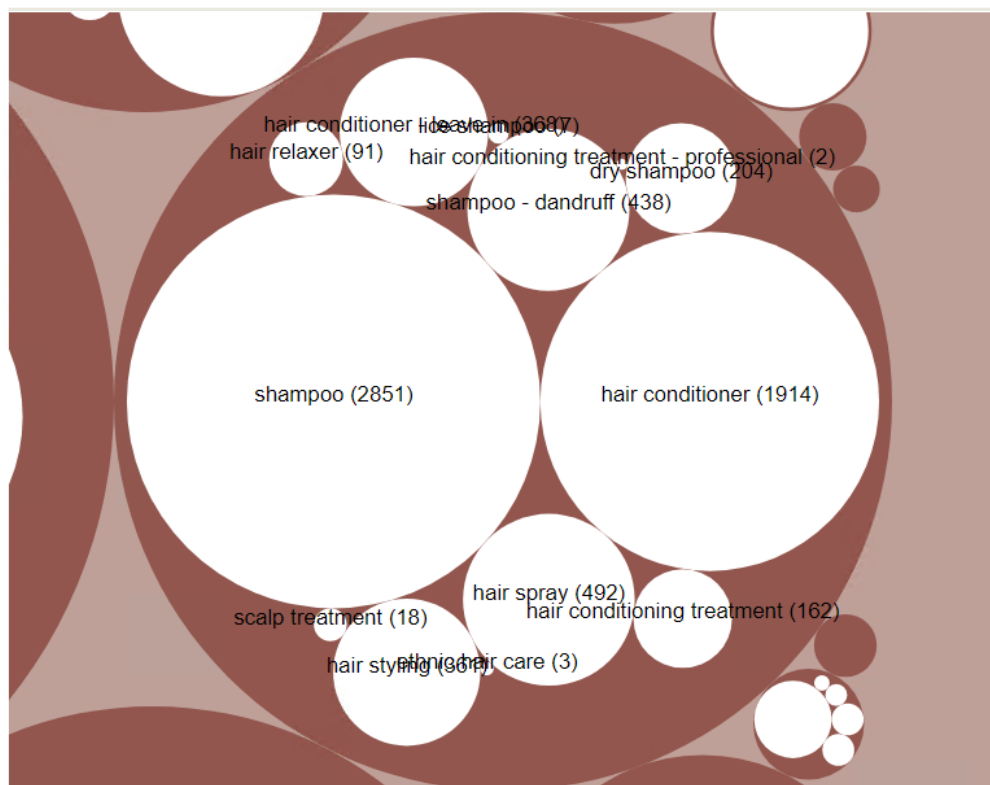
Factotum Database

Consumer Products Database

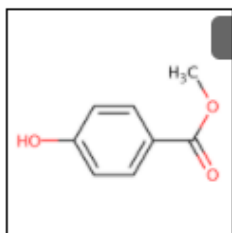
Formulation PUCs

General Category - Product Family - Product Type

> Arts and crafts/Office supplies <i>i</i>	4297
> Cleaning products and household care <i>i</i>	7917
> Electronics/small appliances <i>i</i>	2267
> Home maintenance <i>i</i>	5105
> Landscape/Yard <i>i</i>	1113
> Personal care <i>i</i>	38631
> Pesticides <i>i</i>	1472
> Pet care <i>i</i>	1076
> Sports equipment <i>i</i>	139
> Vehicle <i>i</i>	1743



Sources of Exposure to Chemicals



Methylparaben

99-76-3 | DTXSID4022529

Searched by DSSTox Substance Id.

Download

Columns

10

Product Name

rp super filter coat no. 418

rp super filter cost #421 bulk

sbs-30 waterless skin cleaner

sbs-40 medicated skin cream

schilling test kit

schilling test kt(cyanocobalamin cobalt co 57)_reference std

solution_ a e r pads

super filter coat 412

wondermask p peelable mask_ 2211 bulk

YOUTOPIA- PET SHAMPOO

ns **i**

Search query

action	Maximum Weight Fraction	Data Type	Source
	1.00e-2	MSDS	SIRI
	1.00e-2	MSDS	SIRI
	1.00e-3	MSDS	SIRI
	1.00e-3	MSDS	SIRI
	1.00e-2	MSDS	SIRI
	1.00e-2	MSDS	SIRI
	1.00e-3	MSDS	SIRI
	1.00e-2	MSDS	SIRI
	1.00e-2	MSDS	SIRI
		MSDS	CPCPdb

9 10 > >> Last

Candidate ranking using metadata



© American Society for Mass Spectrometry, 2011

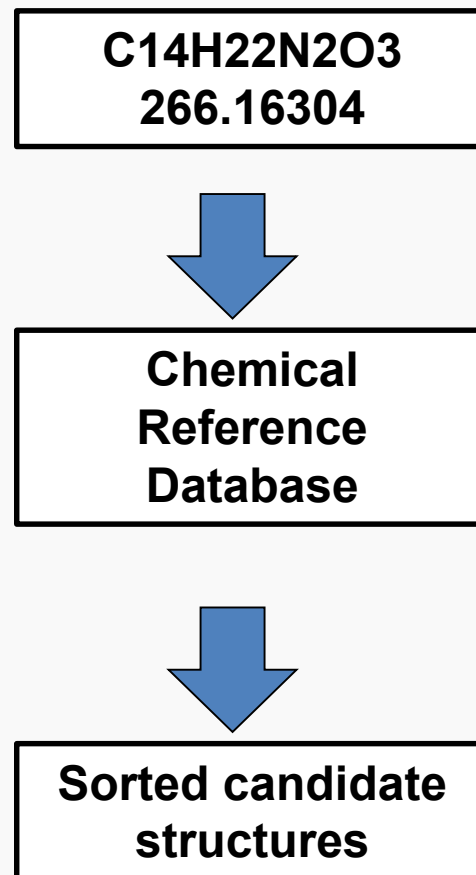
J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

RESEARCH ARTICLE

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

Data Source Ranking of “*known unknowns*”

- A mass and/or formula search is for an ***unknown*** chemical but it is a ***known*** chemical contained within a reference database
- **Most likely** candidate chemicals have the **most** associated data sources, **most** associated literature articles or both



Data Streams for Ranking

- CompTox Dashboard Data Sources
-  Data Source Count
-  Reference Count
- Toxcast *in vitro* bioactivity
- Presence in CPDat database
- OPERA PhysChem Properties
- Other possibilities – predicted media occurrence, frequency of InChIs online

³
Anal Bioanal Chem (2017) 409:1729–1735
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

- When dashboard contained 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			

EXACTLY THE SAME DATASET

How did performance compare?

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

^a 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

**For the same 162 chemicals,
Dashboard outperforms
ChemSpider for both Mass and
Formula Ranking**

“MS-ready” structures supporting searches

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*}

"MS-Ready" Structures

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

Journal of Cheminformatics

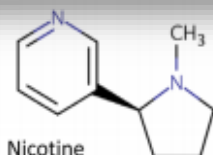
METHODOLOGY

Open Access

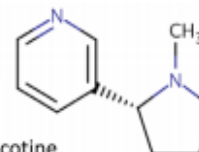


"MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies

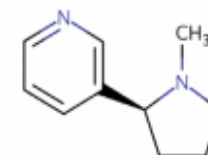
Andrew D. McEachran^{1,2*}, Kamel Mansouri^{1,2,3}, Chris Grulke², Emma L. Schymanski⁴, Christoph Ruttkies⁵ and Antony J. Williams^{2*}



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

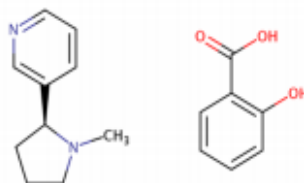


D-Nicotine
CN1CCC[C@@H]1C1=CN=CC=C1
DTXSID004635 | SNICXGAKADSCV
25162-00-9 | **162.1157** | 0.929 | **20**
Tox: **no** | Expo: **yes** | Bioassay: **yes**

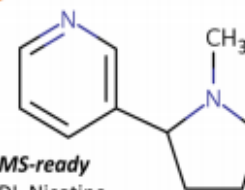


LEGEND: Name, SMILES
DTXSID | InChIKey 1st Block
CAS | Monoiso. Mass | logP | Sources
Data on: Toxicity | Exposure | Bioassays

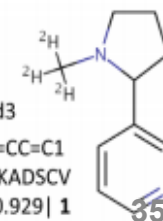
HCl
Nicotine hydrochloride
Cl.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID602093 | HDJBTCAJIMNWEW
2820-51-1 | **198.0924** | 0.929 | **9**
Tox: **no** | Expo: **yes** | Bioassay: **yes**



Benzoic acid, 2-hydroxy-, compd. with
3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)
OC(=O)C1=CC(=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 | SNICXGAKADSCV
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 | SNICXGAKADSCV
69980-24-1 | **165.1345** | 0.929 | **1**
Tox: **no** | Expo: **no** | Bioassay: **no**

Batch Searching

Batch Search Names

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

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

Step Five: Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

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

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

Buprenorphine
 Codeine
 Dextromethorphan
 Dihydrocodeine
 Dihydromorphine
 Ethylmorphine
 Fentanyl
 Heroin
 Hydrocodone
 Hydromorphone

Excel
Download

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

Add Other Data of Interest

Chemical Identifiers

- ☒ DTXSID 
- ☒ Chemical Name 
- ☐ DTXCID 
- ☒ CAS-RN 
- ☒ InChIKey 
- ☐ IUPAC Name 

Structures

- ☐ Mol File 
- ☐ SMILES 
- ☐ InChI String 
- ☒ MS-Ready SMILES 
- ☐ QSAR-Ready SMILES 

Intrinsic And Predicted Properties

- ☒ Molecular Formula 
- ☐ Average Mass 
- ☒ Monoisotopic Mass 
- ☐ TEST Model Predictions 
- ☐ OPERA Model Predictions 

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

Batch Searching Formula/Mass

Batch Search

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

Step Five: Choose Data Fields to Download

Please enter one identifier per line 

Select Input Type(s)

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

 Display All Chemicals

+/- ppm

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

41.0265
56.02621
53.0265
58.0418|
93.0578
113.9639
151.8754
69.9377
77.9872

This search is based on what we refer to as "Mass Spec (MS) Ready" structures. All chemicals within the database are treated in a manner such that all are desalted, mixtures are separated, and stereochemistry is removed as Mass Spectrometry detects the major components of a salt or mixture and is insensitive to stereochemistry. As an example, a search for the monoisotopic mass of phenol will return phenol, sodium phenolate and calcium phenoxide. See the publication for more details: <https://doi.org/10.1186/s13321-018-0299-2>.

Searching batches using MS-Ready Formula (or mass) searching

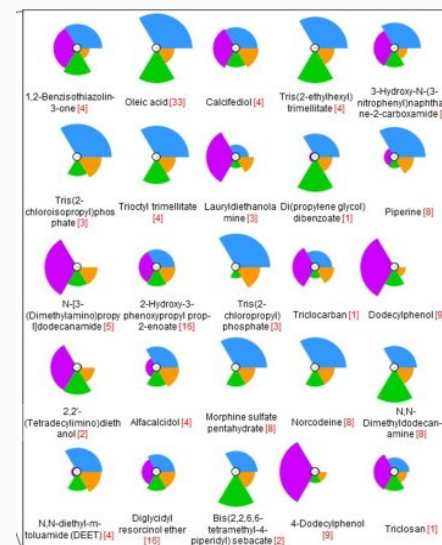
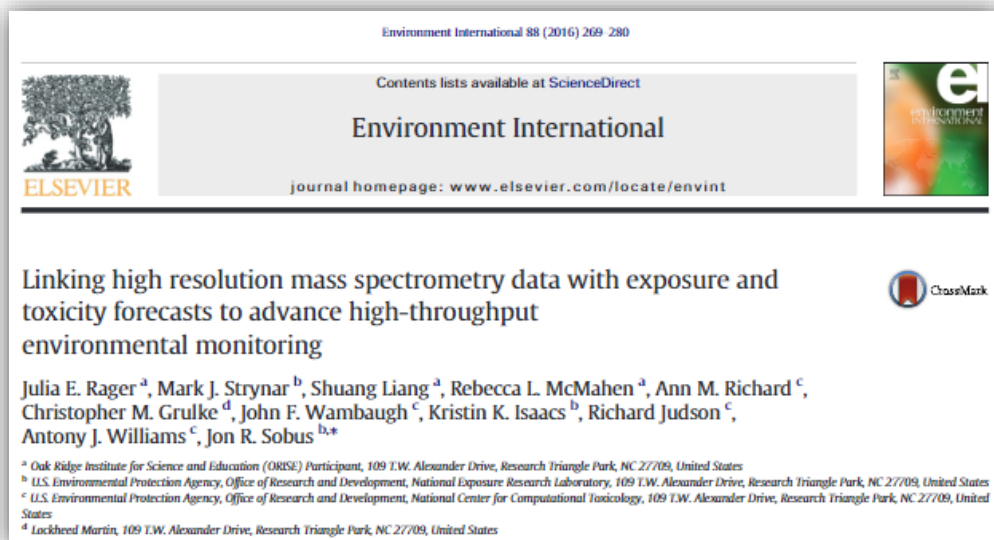
	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	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-phenoxypyrimidine	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

Batch Search in specific lists

☐	INPUT	DTXSID	MASSBANKREF	NEMILIST	WRTMSD	NORMANPRI	SUSDAT
☐	Buprenorph	DTXSID202	-	-	Y	-	Y
☐	Codeine	DTXSID202	Y	Y	Y	Y	Y
☐	Dextrometh	DTXSID302	Y	Y	Y	-	Y
☐	Dihydrocod	DTXSID502	Y	-	Y	Y	Y
☐	Dihydromor	DTXSID704	-	-	-	-	Y
☐	Ethylmorph	DTXSID104	-	-	Y	-	Y
☐	Fentanyl	DTXSID902	Y	-	Y	-	Y
☒	Heroin	DTXSID604	Y	-	Y	Y	Y
☒	Hydrocodor	DTXSID802	Y	Y	Y	Y	Y
☐	Hydromorph	DTXSID802	-	-	Y	-	Y
☐	Ketamine	DTXSID802	Y	-	Y	-	Y
☒	Meperidine	DTXSID902	Y	-	Y	-	Y
☐	Methadone	DTXSID702	Y	Y	Y	-	Y
☒	Morphine	DTXSID902	Y	Y	Y	Y	Y
☐	Morphinone	DTXSID501	-	-	-	-	Y
☒	Naloxone	DTXSID802	-	-	Y	-	Y
☐	Naltriben	-	-	-	-	-	-
☐	Oxycodone	DTXSID502	Y	Y	Y	Y	Y
☐	Oxymorpho	DTXSID502	-	-	Y	-	Y
☐	Propoxyph	DTXSID102	Y	Y	Y	-	Y
☐	Sufentanil	DTXSID602	-	-	Y	-	Y
☐	Tramadol	DTXSID908	Y	Y	Y	Y	Y

Benefits of bringing it all together

- The true dashboard benefit is integration
- Rank potential candidates for toxicity using available data – hazard, exposure, *in vitro*



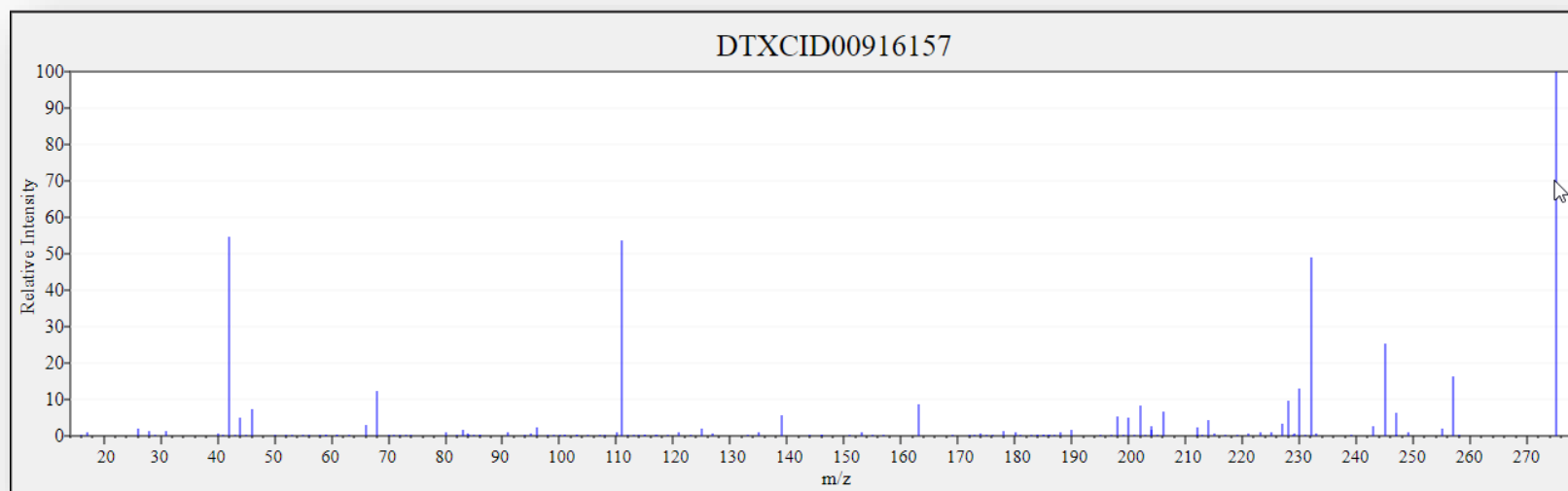
Dashboard Data for in silico spectral prediction

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

321.136493476

Da

±

0.0000002

Da

ppm

Molecular Formula Search

Molecular Formula

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 7.306439699
123.0440559 6.538348292
196.0756904 5.269463115
216.1019051 4.700461978
300.1080005 4.800442384

Peak Match Window:

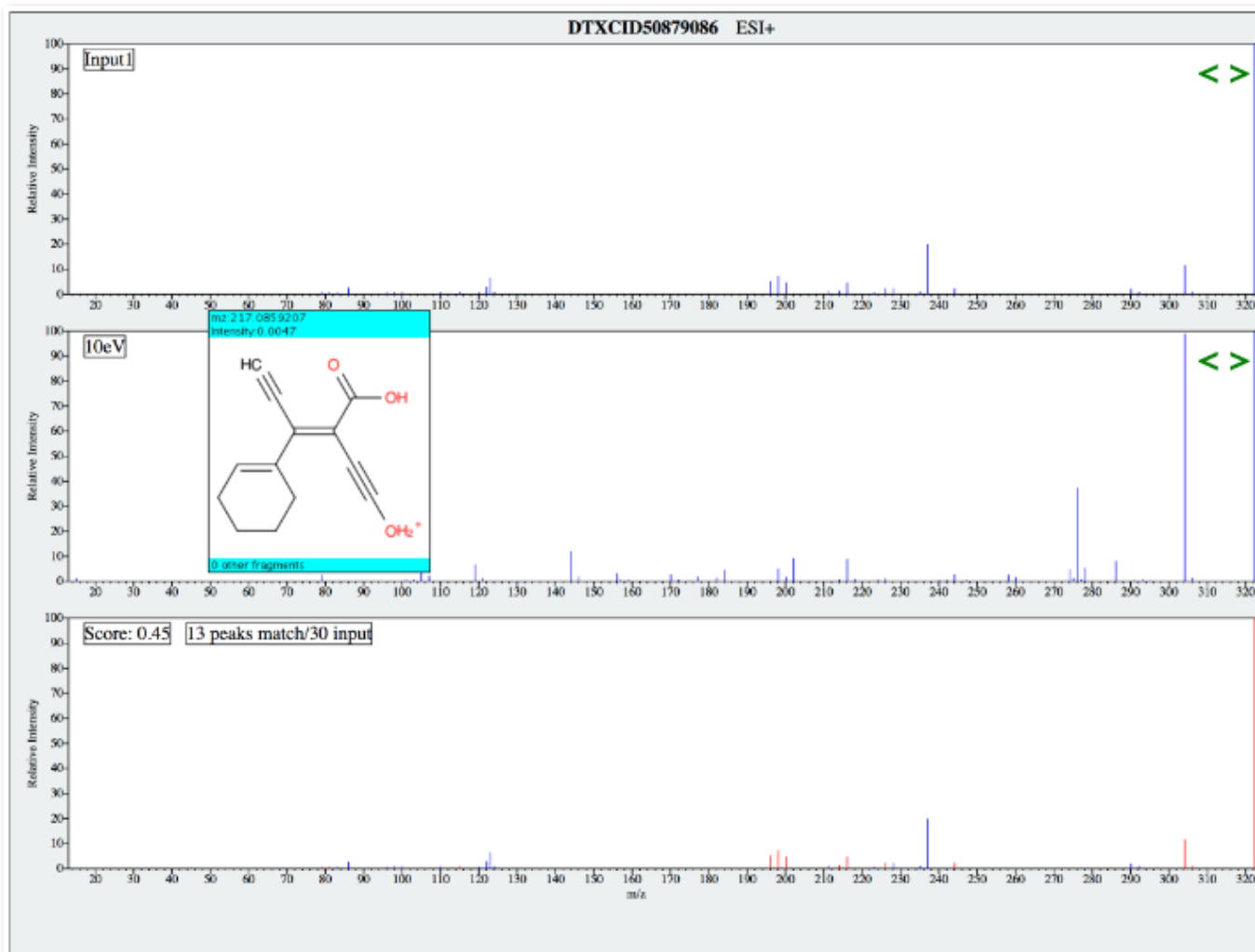
0.02

Da

ppm

Search

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

Analytical and Bioanalytical Chemistry

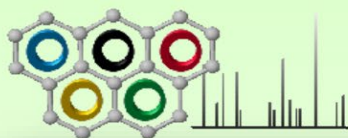
RESEARCH PAPER

In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples

Alex Chao^{1,2} • Hussein Al-Ghoul^{1,2} • Andrew D. McEachran^{1,3} • Ilya Balabin⁴ • Tom Transue⁴ • Tommy Cathey⁴ • Jarod N. Grossman^{2,3} • Randolph Singh^{1,5} • Elin M. Ulrich² • Antony J. Williams⁶ • Jon R. Sobus²

Received: 4 October 2019 / Revised: 27 November 2019 / Accepted: 11 December 2019

© The Author(s) 2019

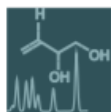


CASMI

Critical Assessment of Small Molecule Identification

The experimental and computational mass spectrometry communities are invited to participate in the fifth round of an open contest on the identification of small molecules from mass spectrometry data.

This year the contest will test the applicability of MS and MS/MS on natural products chemistry identifications. With 45 (Category 1) and up to 243 (Categories 2&3) natural products challenges - including a few tricky ones - there's something for everyone!



metabolites



Article

Revisiting Five Years of CASMI Contests with EPA Identification Tools

Andrew D. McEachran ^{1,*}, Alex Chao ¹, Hussein Al-Ghoul ¹, Charles Lowe ²,
Christopher Grulke ², Jon R. Sobus ² and Antony J. Williams ^{2,*}

Predicted Data are Public *Publication and Data Files*

Data Descriptor | OPEN | Published: 02 August 2019

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

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

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

CFM-ID Paper Data

[Dataset](#) posted on 01.03.2019, 08:38 by EPA's National Center for Computational Toxicology

88
views

17
downloads

0
citations

This upload is a zip containing the following files:

Predicted EI-MS Spectra of CompTox Chemicals Dashboard Structures:

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

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

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

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

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



CATEGORIES

• Toxicology

KEYWORD(S)

Computational Toxicology

DSSTox Chemical Database

Chemicals Dashboard

Non-targeted analysis

CFM-ID

LICENCE



EXPORT

RefWorks

BibTeX

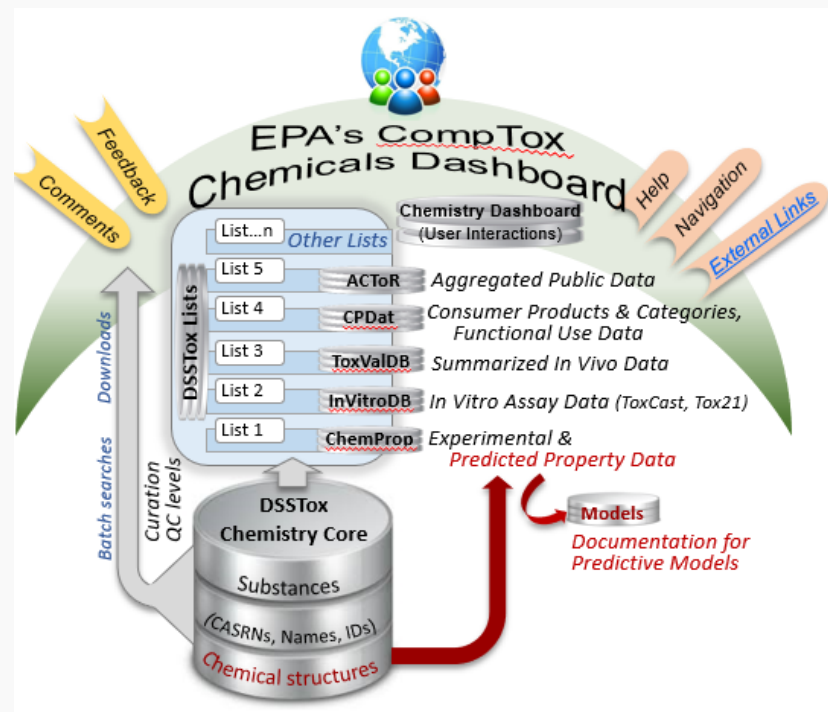
Ref. manager

Endnote

https://epa.figshare.com/articles/CFM-ID_Paper_Data/7776212/1

Conclusion

- Dashboard access to data for ~882,000 chemicals
- Chemical Lists support exposome analysis
- Related metadata facilitates candidate ranking
- Relationship mappings and chemical lists of great utility
- Curation and mutual sharing of chemical lists is important (e.g. NORMAN)



Acknowledgements



Credit: the Research Triangle Foundation

EPA ORD

Ann Richard
Chris Grulke
John Wambaugh
Jeremy Dunne
Jeff Edwards
Grace Patlewicz
Alex Chao
Kristin Isaacs
Charles Lowe
James McCord
Seth Newton
Katherine Phillips
Tom Purucker
Jon Sobus
Mark Strynar
Elin Ulrich
Joach Pleil

TEAMS

IT Development Team
Curation Team

ILS

Kamel Mansouri

GDIT

Ilya Balabin
Tom Transue
Tommy Cathey

Collaborators

Emma Schymanski
NORMAN Network
Andrew McEachran
Jerry Zweigenbaum