http://www.orcid.org/0000-0002-2668-4821

The US-EPA CompTox Chemicals Dashboard to support Non-Targeted Analysis

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

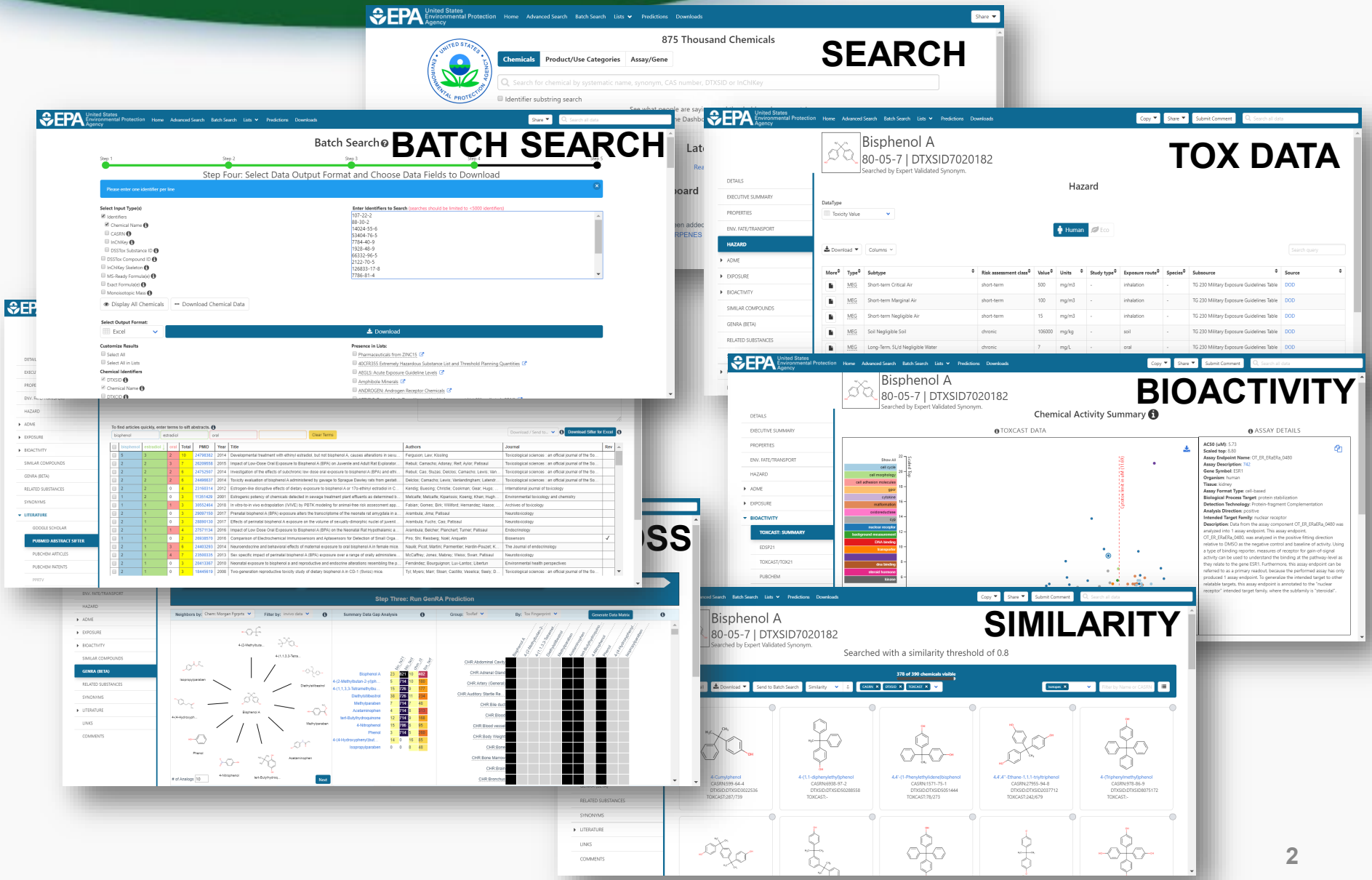
Triangle Area Mass Spectrometry
RTP, January 2020

Outline

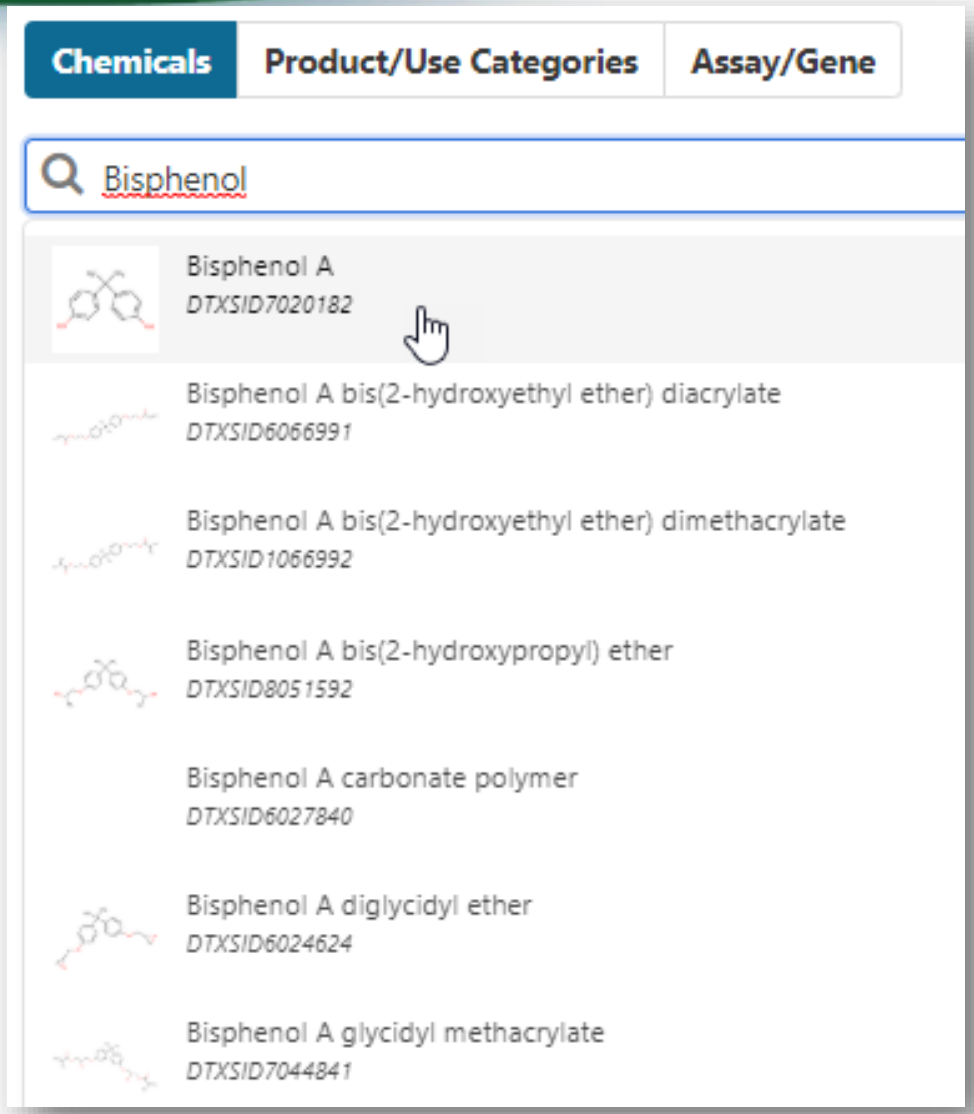
- Quick overview of the dashboard
- Specific data of interest to this audience (it's not just Computational Toxicology)
- Support for Mass Spectrometry
- Data ***quality*** in the public domain
- Work in progress – prototypes
- A request for help

CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard



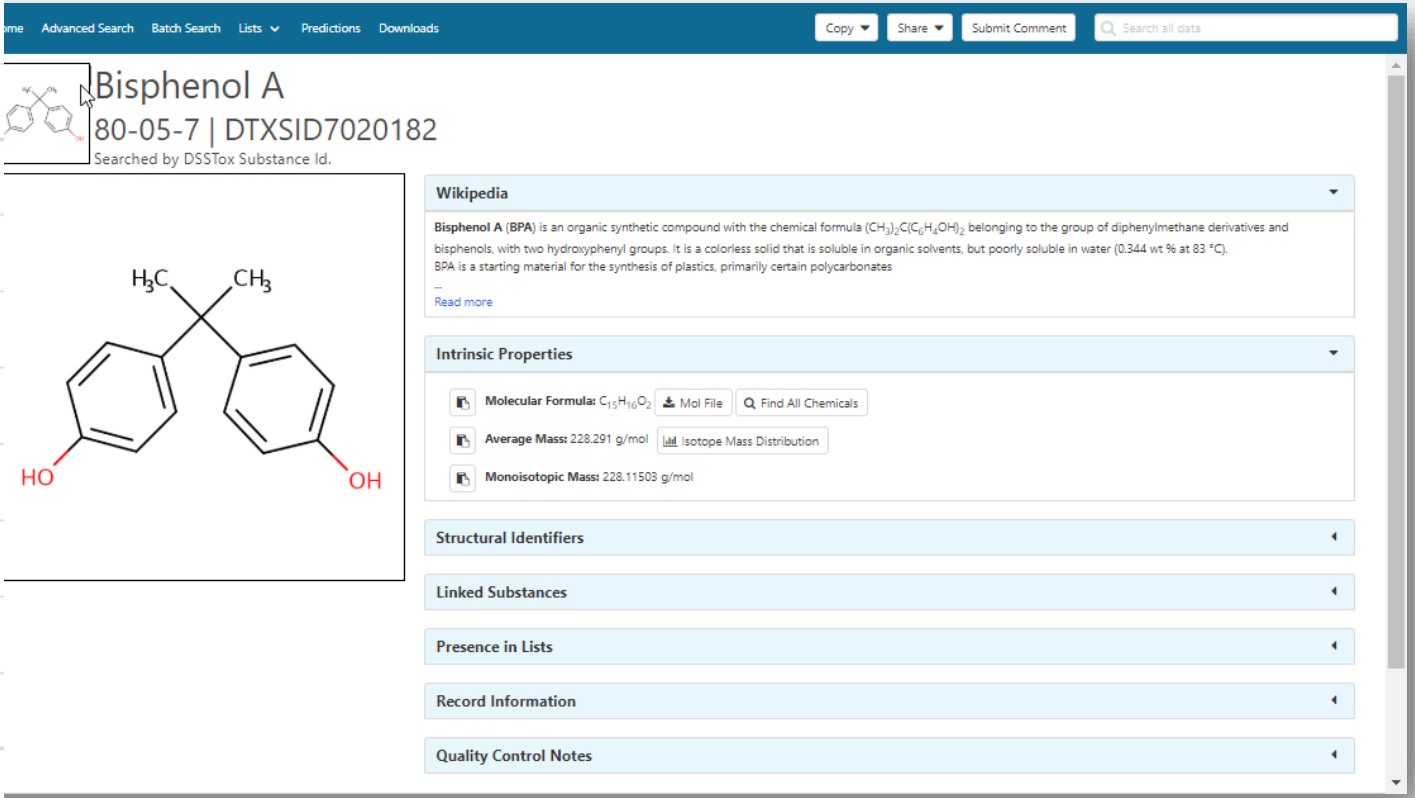
BASIC Search



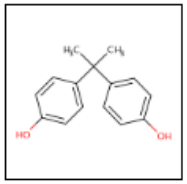
Detailed Chemical Pages

EPA United States Environmental Protection Agency

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



Properties, Fate and Transport



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Property

Summary

Summary

Download Columns

Property	Experimental average	Predicted average	Experimental median	Predicted median
LogKow: Octanol-Water	3.32 (1)	3.30		3.39
Melting Point	155 (7)	140	156	144
Boiling Point	200 (1)	360		355
Water Solubility	8.55e-4 (3)	8.78e-4	5.26e-4	7.56e-4
Vapor Pressure	-	6.83e-7		1.51e-7
Flash Point	-	190		190
Surface Tension	-	46.0		
Index of Refraction	-	1.60		
Molar Refractivity	-	68.2		

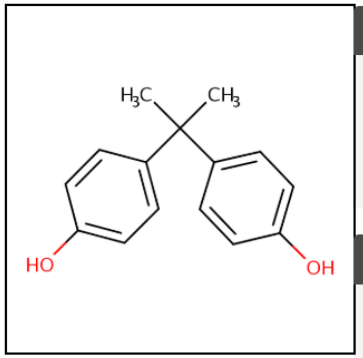
Properties, Fate and Transport
e.g. Solubility

Download Experimental Data

Source	Result
PhysPropNCCT	5.26e-4
Tetko et al. J. Chem. Inf. and Comp. Sci. 41.6 (2001): 1488-1493	1.51e-3
Kovdienko, et. al. Molecular informatics 29.5 (2010): 394-406.	5.25e-4

Source	Result	Calculation Details
EPISUITE	7.56e-4	Not Available
NICEATM	1.31e-3	Not Available
TEST	1.24e-3	TEST Report
OPERA	5.44e-4	OPERA Model Report [Inside AD]
OPERA2	5.35e-4	Not Available

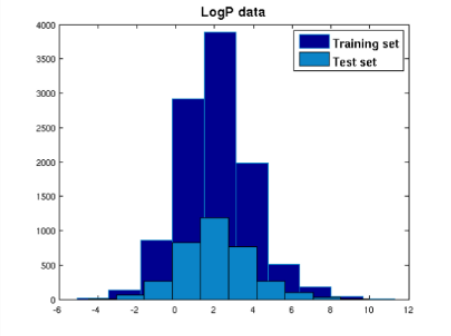
Properties, Fate and Transport
e.g. logP

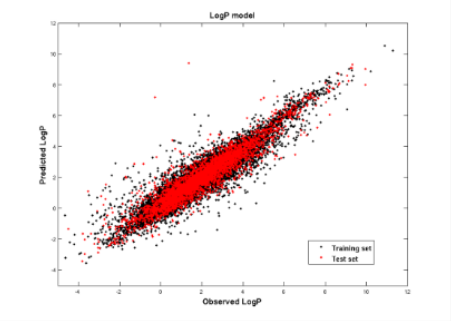


Model Results

Predicted value: 3.35
Global applicability domain: Inside
Local applicability domain index: 0.877
Confidence level: 0.748

Model Performance





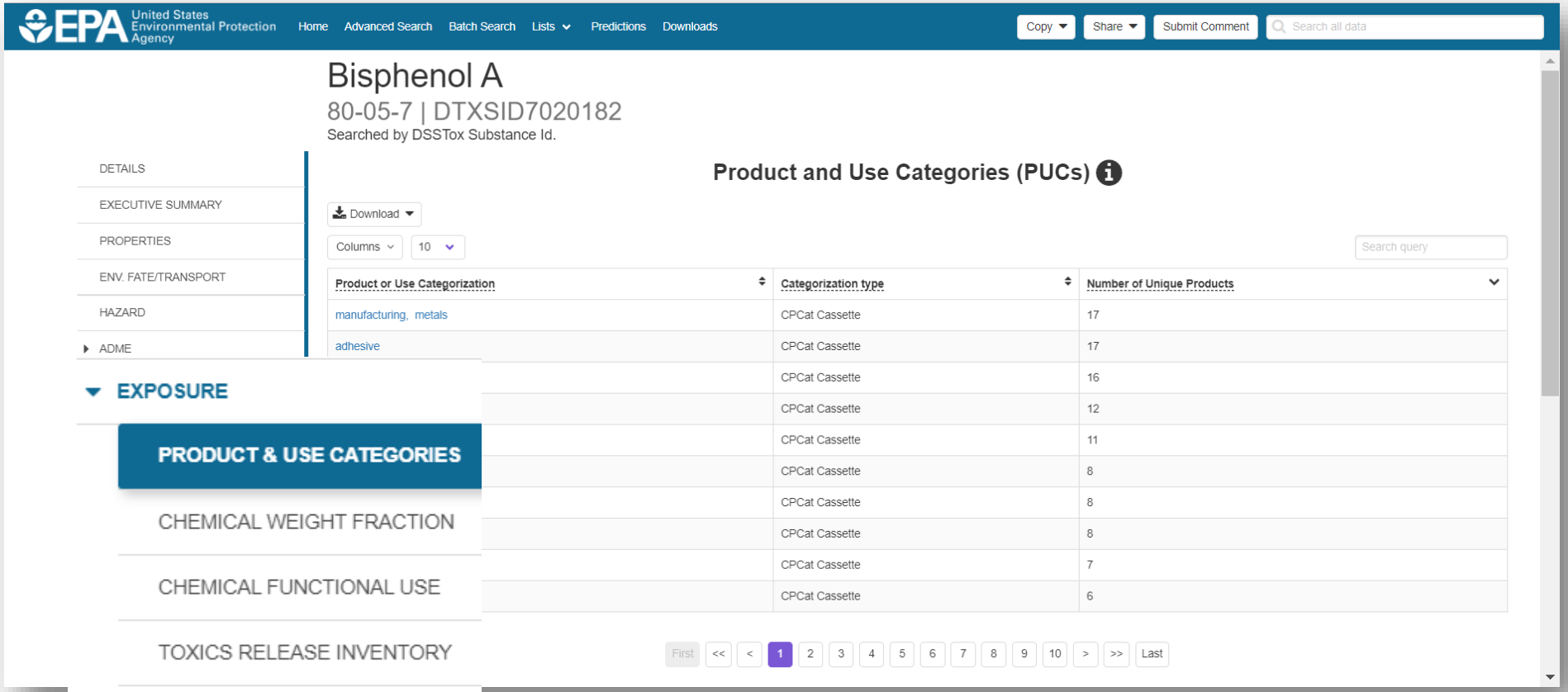
QMRF

Weighted KNN model

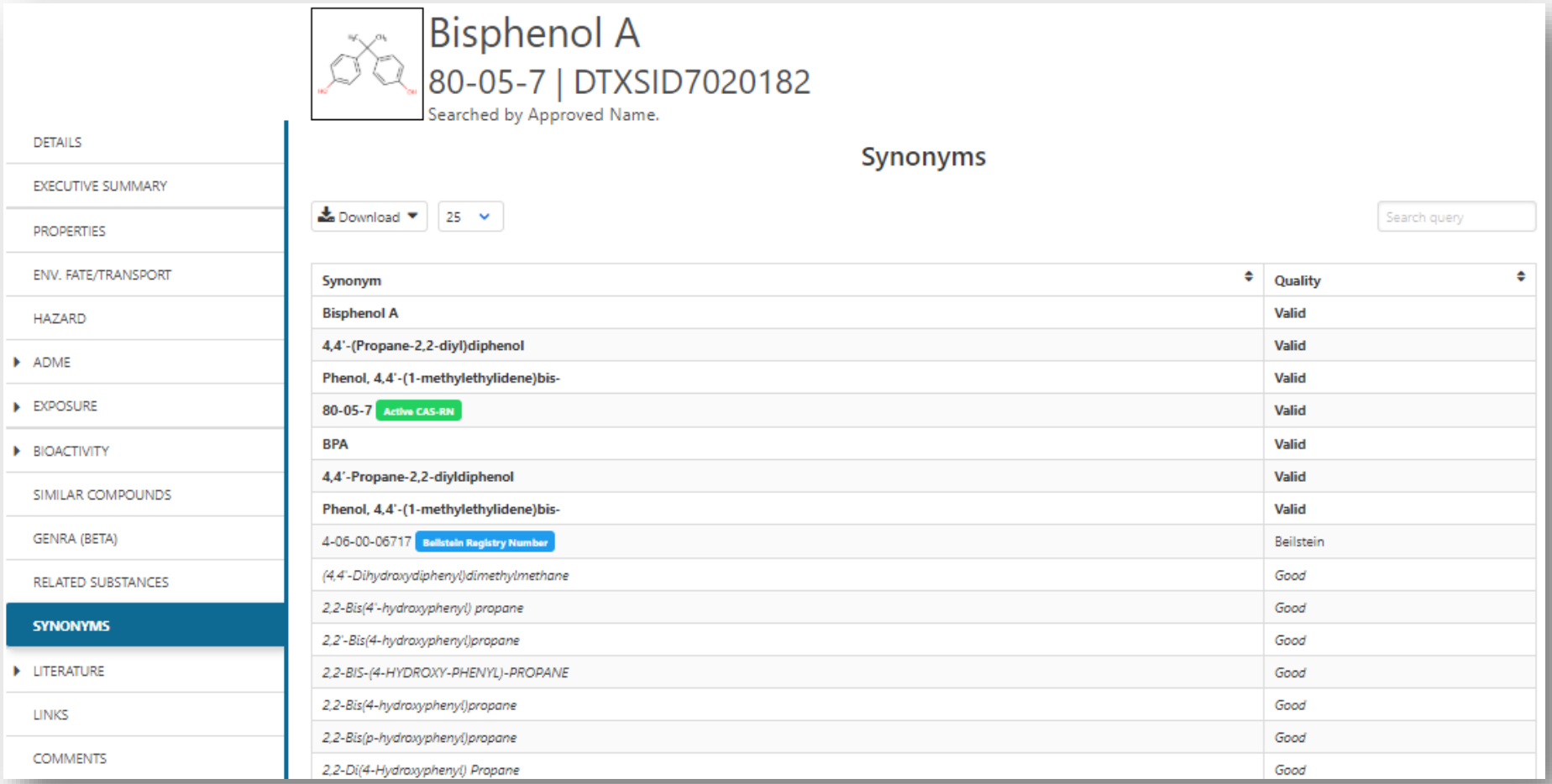
5-fold CV (75%)		Training (75%)		Test (25%)	
Q2	RMSE	R2	RMSE	R2	RMSE
0.850	0.690	0.860	0.670	0.860	0.780

od.epa.gov/dashboard/advanced_search/index

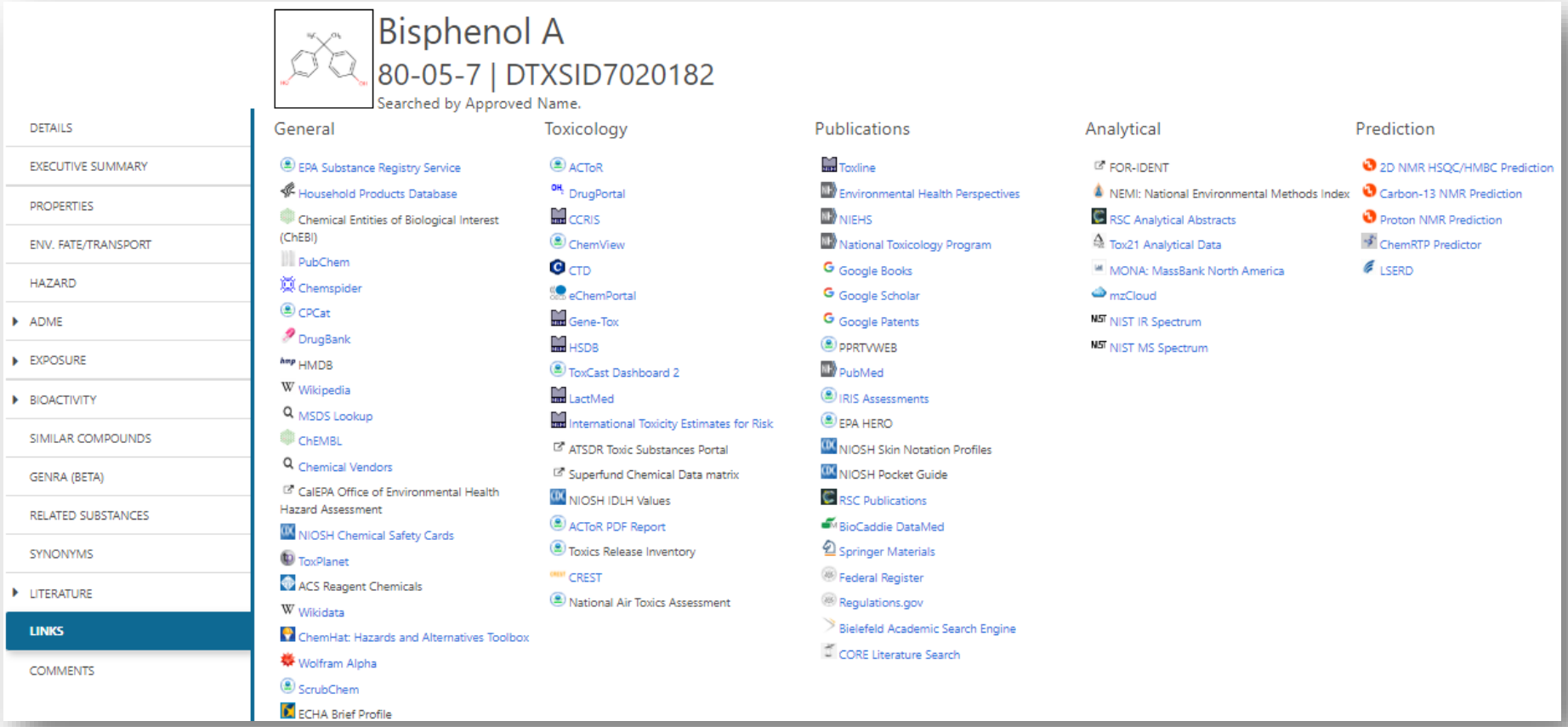
Sources of Exposure to Chemicals



Identifiers to Support Searches



Link Access

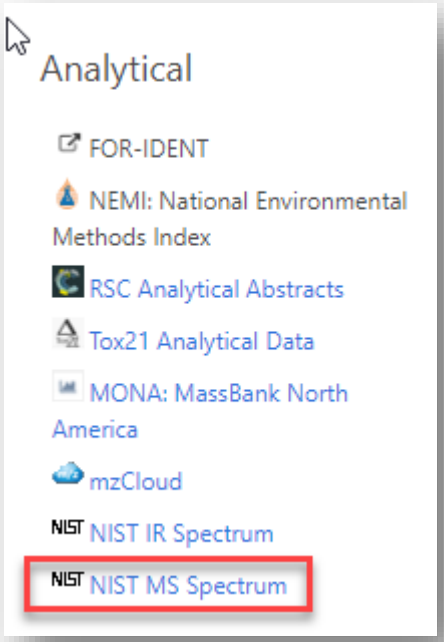

Bisphenol A
80-05-7 | DTXSID7020182
Searched by Approved Name.
DETAILS
EXECUTIVE SUMMARY
PROPERTIES
ENV. FATE/TRANSPORT
HAZARD
ADME
EXPOSURE
BIOACTIVITY
SIMILAR COMPOUNDS
GENRA (BETA)
RELATED SUBSTANCES
SYNONYMS
LITERATURE
LINKS
COMMENTS
General
EPA Substance Registry Service
Household Products Database
Chemical Entities of Biological Interest (ChEBI)
PubChem
Chemspider
CPCat
DrugBank
HMDB
Wikipedia
MSDS Lookup
ChEMBL
Chemical Vendors
CalEPA Office of Environmental Health Hazard Assessment
NIOSH Chemical Safety Cards
ToxPlanet
ACS Reagent Chemicals
Wikidata
ChemHat: Hazards and Alternatives Toolbox
Wolfram Alpha
ScrubChem
ECHA Brief Profile
Toxicology
ACToR
DrugPortal
CCRIS
ChemView
CTD
eChemPortal
Gene-Tox
HSDB
ToxCast Dashboard 2
LactMed
International Toxicity Estimates for Risk
ATSDR Toxic Substances Portal
Superfund Chemical Data matrix
NIOSH IDLH Values
ACToR PDF Report
Toxics Release Inventory
CREST
National Air Toxics Assessment
Publications
Toxline
Environmental Health Perspectives
NIEHS
National Toxicology Program
Google Books
Google Scholar
Google Patents
PPRTVWEB
PubMed
IRIS Assessments
EPA HERO
NIOSH Skin Notation Profiles
NIOSH Pocket Guide
RSC Publications
BioCaddie DataMed
Springer Materials
Federal Register
Regulations.gov
Bielefeld Academic Search Engine
CORE Literature Search
Analytical
FOR-IDENT
NEMI: National Environmental Methods Index
RSC Analytical Abstracts
Tox21 Analytical Data
MONA: MassBank North America
mzCloud
NIST IR Spectrum
NIST MS Spectrum
Prediction
2D NMR HSQC/HMBC Prediction
Carbon-13 NMR Prediction
Proton NMR Prediction
ChemRTP Predictor
LSERD

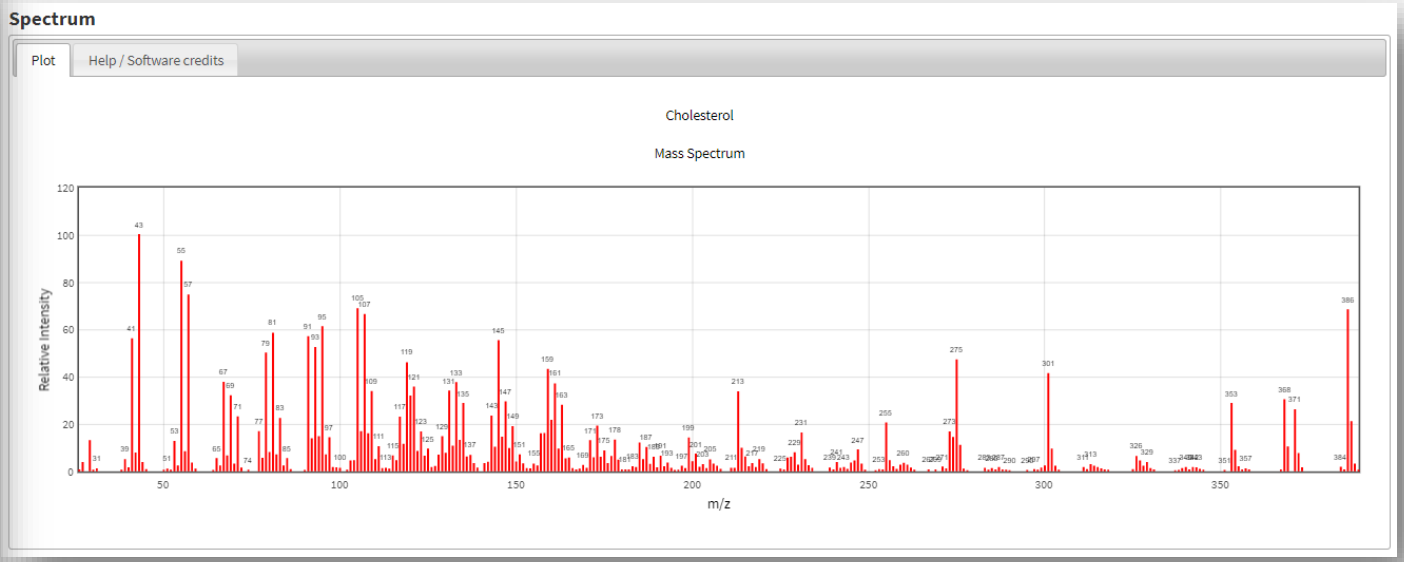
Mass Spec Links

Analytical

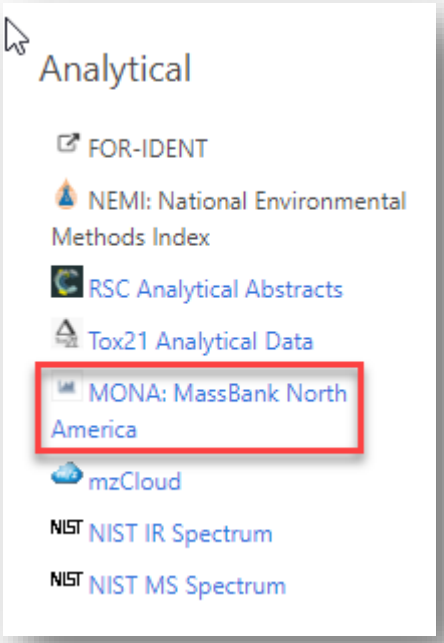
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America
- mzCloud
- NIST IR Spectrum
- NIST MS Spectrum
- MassBank
- NEMI: National Environmental Methods Index
- NIST Antoine Constants
- IR Spectra on PubChem
- NIST Kovats Index values

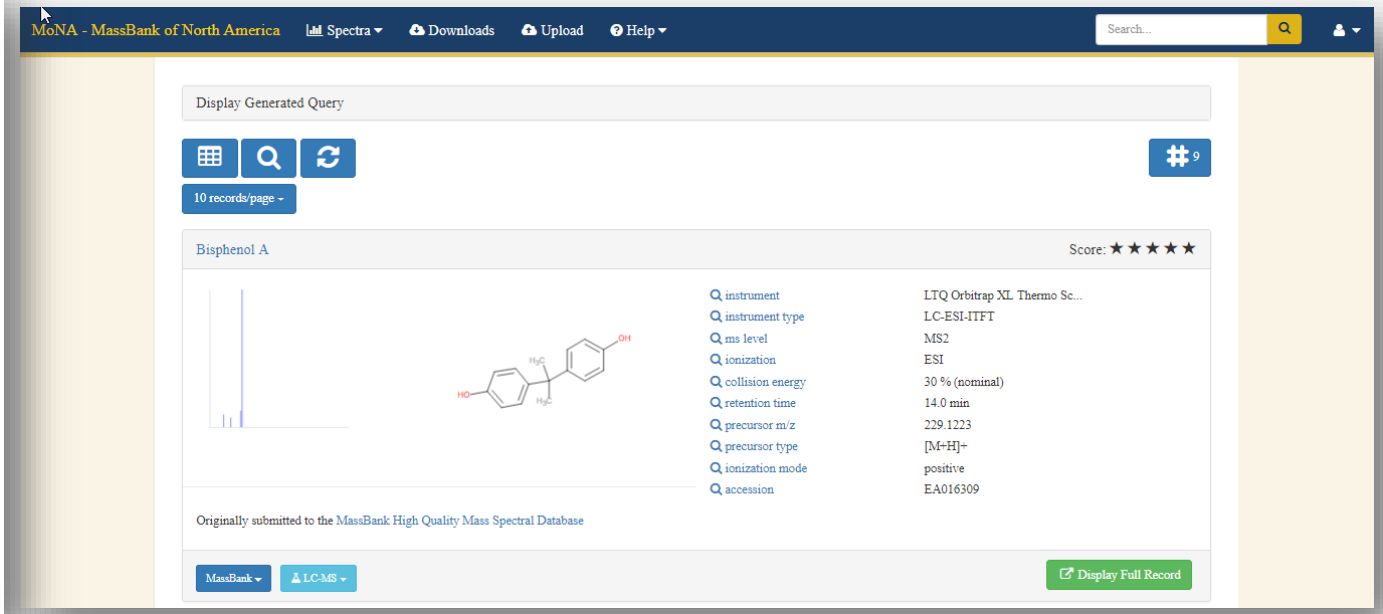
NIST WebBook
https://webbook.nist.gov/chemistry/





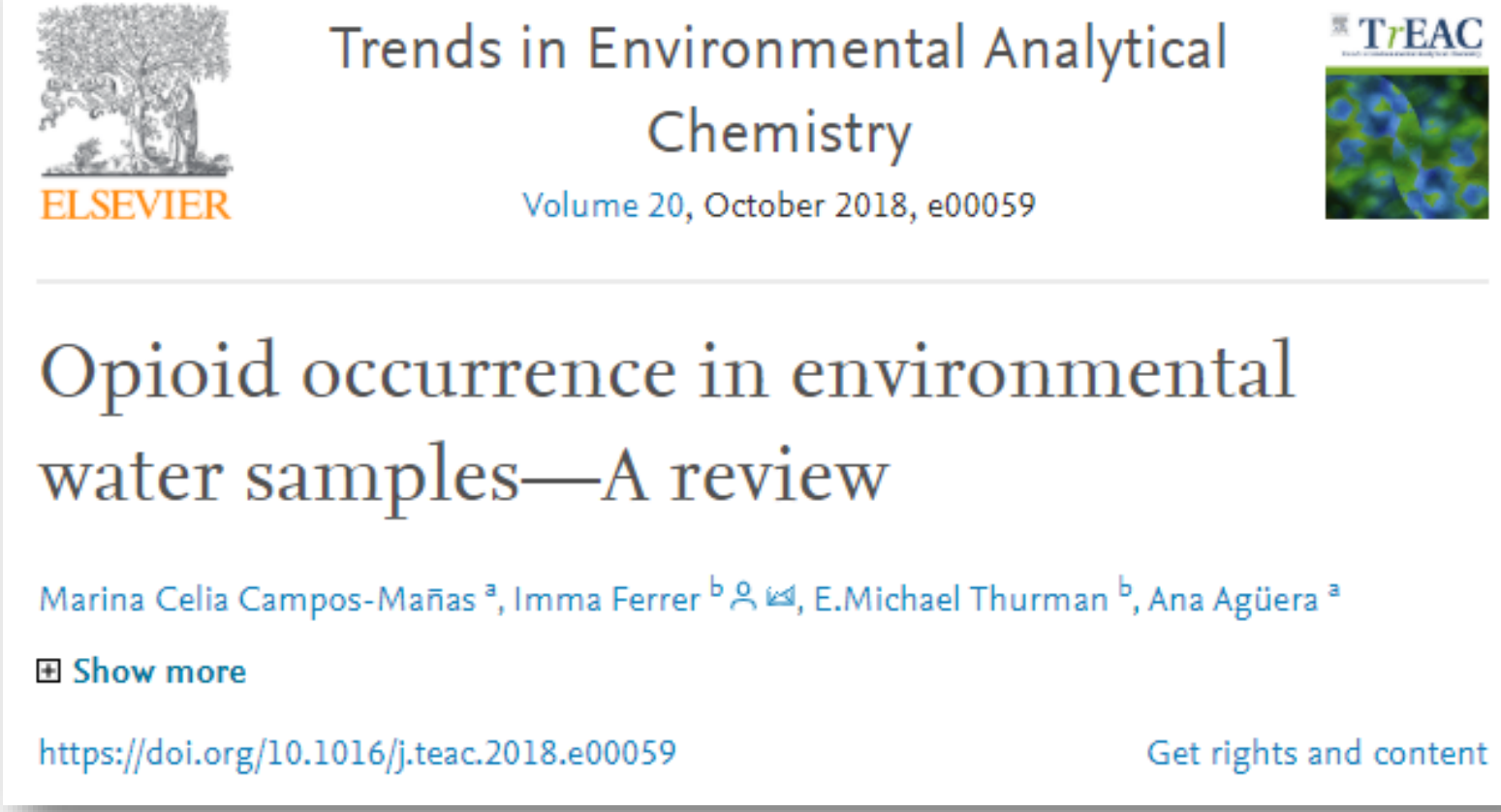
MassBank of North America
https://mona.fiehnlab.ucdavis.edu





Batch Searching

Aggregate data for a list of chemicals



ELSEVIER

Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059

TrEAC

Opioid occurrence in environmental water samples—A review

Marina Celia Campos-Mañas [a], Imma Ferrer [b], E.Michael Thurman [b], Ana Agüera [a]

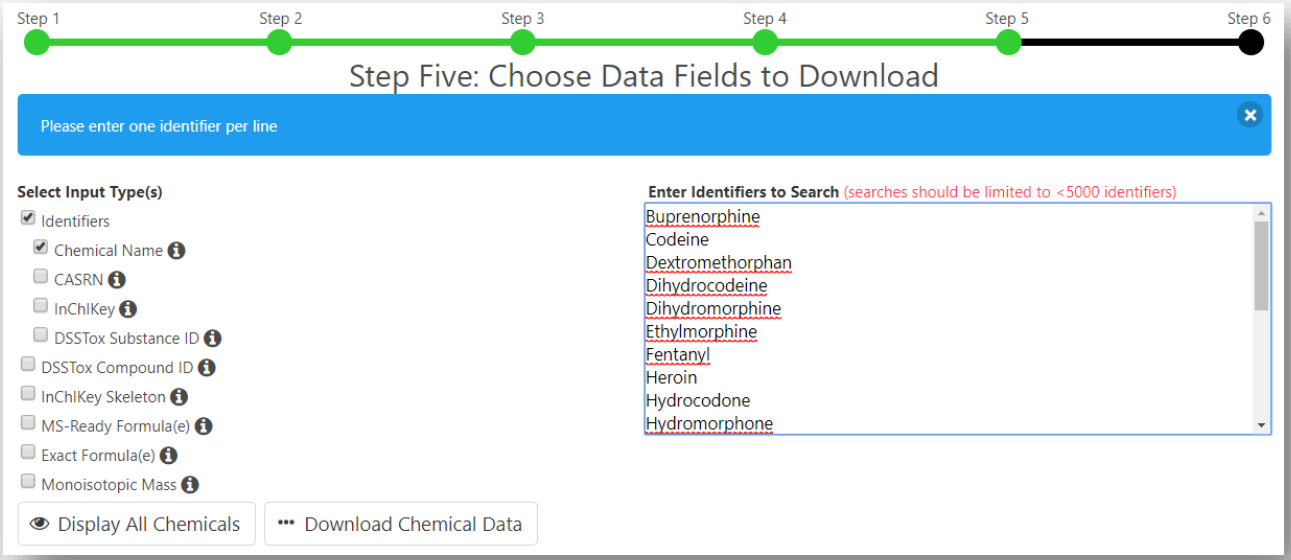
Show more

https://doi.org/10.1016/j.teac.2018.e00059

Get rights and content

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



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

- [x] DTXSID
- [x] Chemical Name
- [ ] DTXCID
- [x] CAS-RN
- [x] InChIKey
- [ ] IUPAC Name

Structures

- [ ] Mol File
- [ ] SMILES
- [ ] InChI String
- [x] MS-Ready SMILES
- [ ] QSAR-Ready SMILES

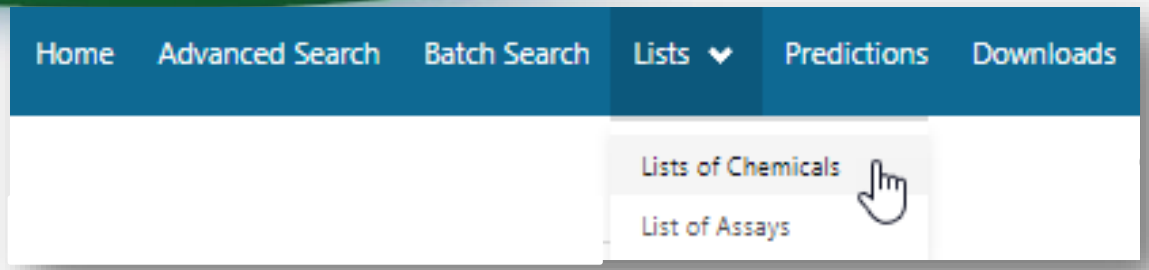
Intrinsic And Predicted Properties

- [x] Molecular Formula
- [ ] Average Mass
- [x] Monoisotopic Mass
- [ ] TEST Model Predictions
- [ ] OPERA Model Predictions

INPUT	DTXSID	CASRN	MOLECULAR_F	MONOISOTOPIC	MS_READY_SMI
Buprenorph	DTXSID202	52485-79-7	C29H41NO4	467.3035588	[H]C12CC3=C4C(
Codeine	DTXSID202	76-57-3	C18H21NO3	299.1521435	[H]C12CC3=C4C(
Dextrometh	DTXSID302	125-71-3	C18H25NO	271.1936144	[H]C12CC3=C(C=
Dihydrocod	DTXSID502	125-28-0	C18H23NO3	301.1677936	[H]C12CC3=C4C(
Dihydromor	DTXSID704	509-60-4	C17H21NO3	287.1521435	[H]C12CC3=C4C(
Ethylmorph	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(
Hydrocodor	DTXSID802	125-29-1	C18H21NO3	299.1521435	[H]C12CC3=C4C(
Hydromorph	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(
Naltriben	-	-	-	-	-
Oxycodone	DTXSID502	76-42-6	C18H21NO4	315.1470582	[H]C12CC3=C4C(
Oxymorpho	DTXSID502	76-41-5	C17H19NO4	301.1314081	[H]C12CC3=C4C(
Propoxyphe	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(=

Chemical Lists of Interest…

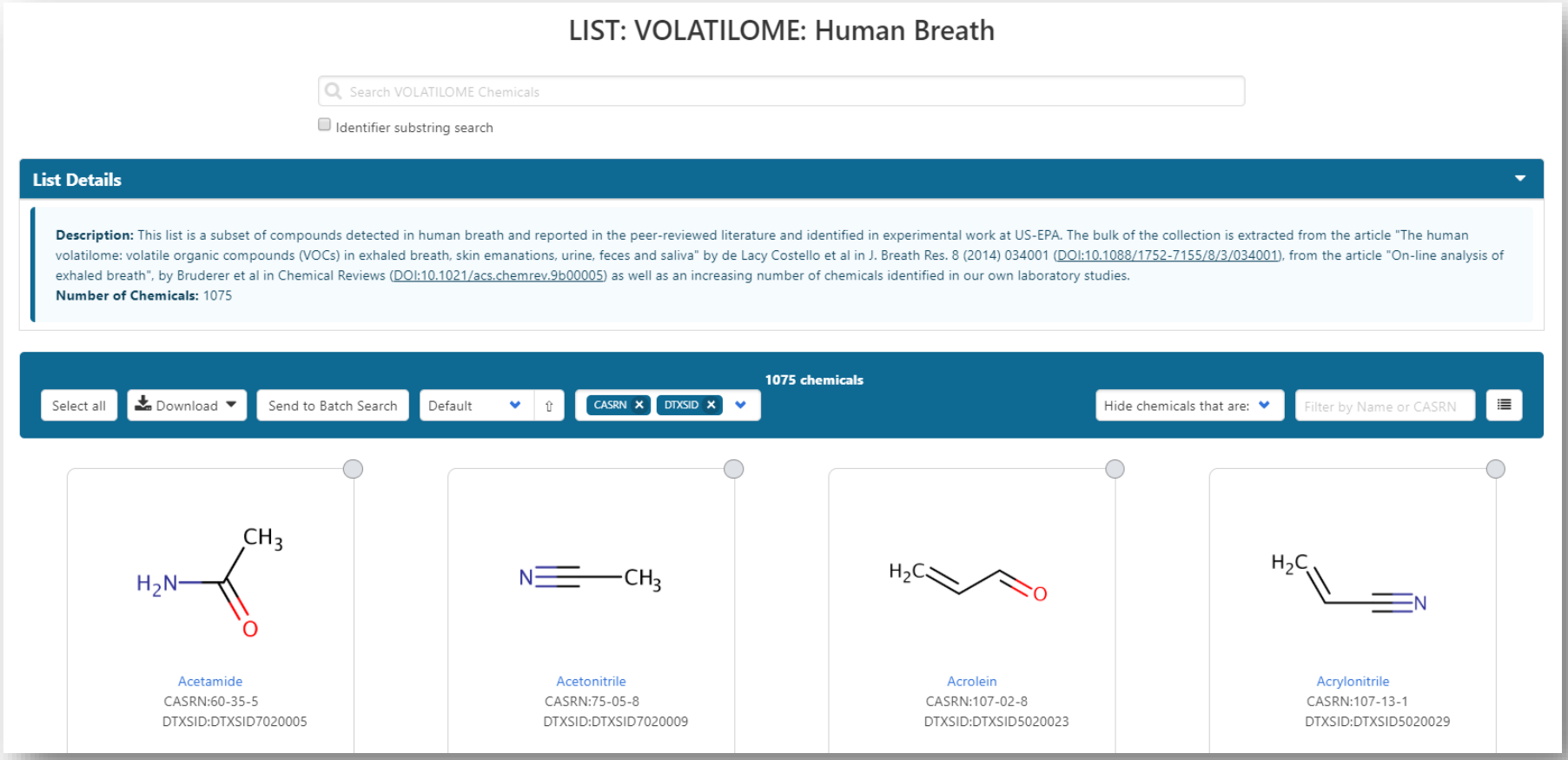
225 Chemical Lists (and growing)



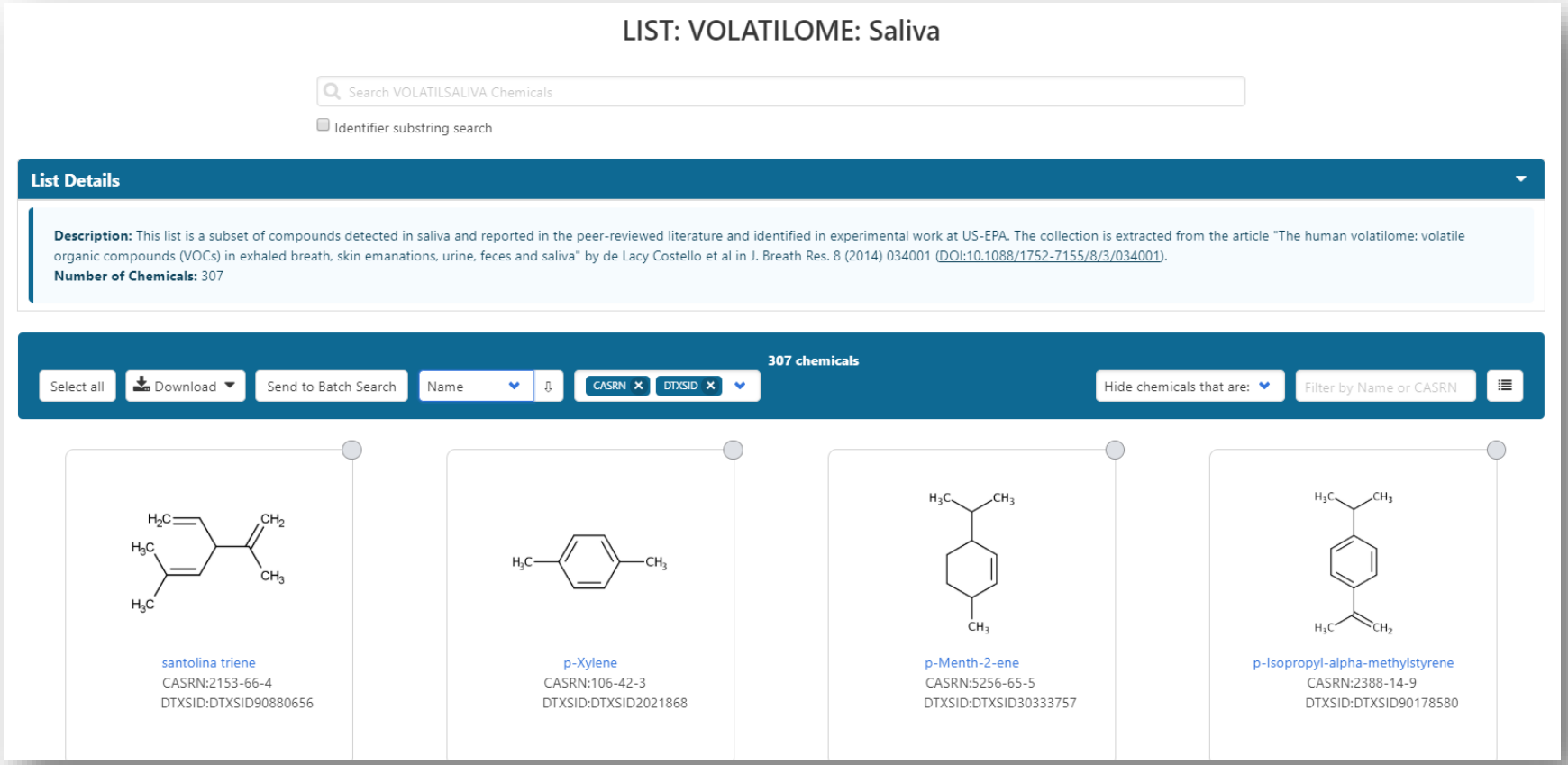
Download | Columns | mass | Copy Filtered Lists URL

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

“Volatilome” Human Breath



“Volatilome” Saliva



PFAS lists of Chemicals

EPA United States Environmental Protection Agency

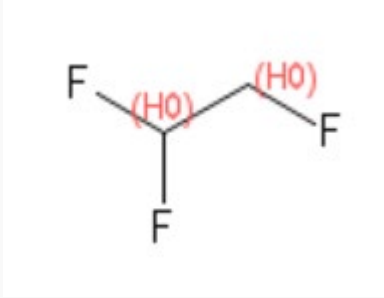
Select List

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)

Building a “reference” PFAS list

- PFAS structure list (PFASSTRUCT) is expanded from public databases, agency lists and literature
- Approaching ~7000 structures – 98.8% have associated CAS Numbers
- Compare with PubChem 220,720 structures



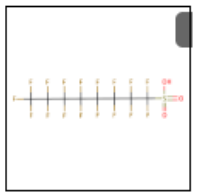
DATA SOURCES > CONTRIBUTOR INFORMATION

SureChEMBL

Share Tweet Email

SureChEMBL automatically extracts chemistry from the full text patent documents provided by the three major patent authorities (WIPO, USPTO, EPO). Compounds are derived from the chemical names found in text, images and attached MOL files, where available.

Formula Search can find isomers



Perfluorooctanesulfonic acid

1763-23-1 | DTXSID3031864

Searched by Synonym from Valid Source.

Wikipedia

Perfluorooctanesulfonic acid (conjugate base **perfluorooctanesulfonate**) (**PFOS**) is an a
was the key ingredient in Scotchgard, a fabric protector made by 3M, and numerous stain r
Convention on Persistent Organic Pollutants in May 2009. PFOS can be synthesized in indus
precursors. PFOS levels that have been detected in wildlife

...

Read more

Quality Control Notes

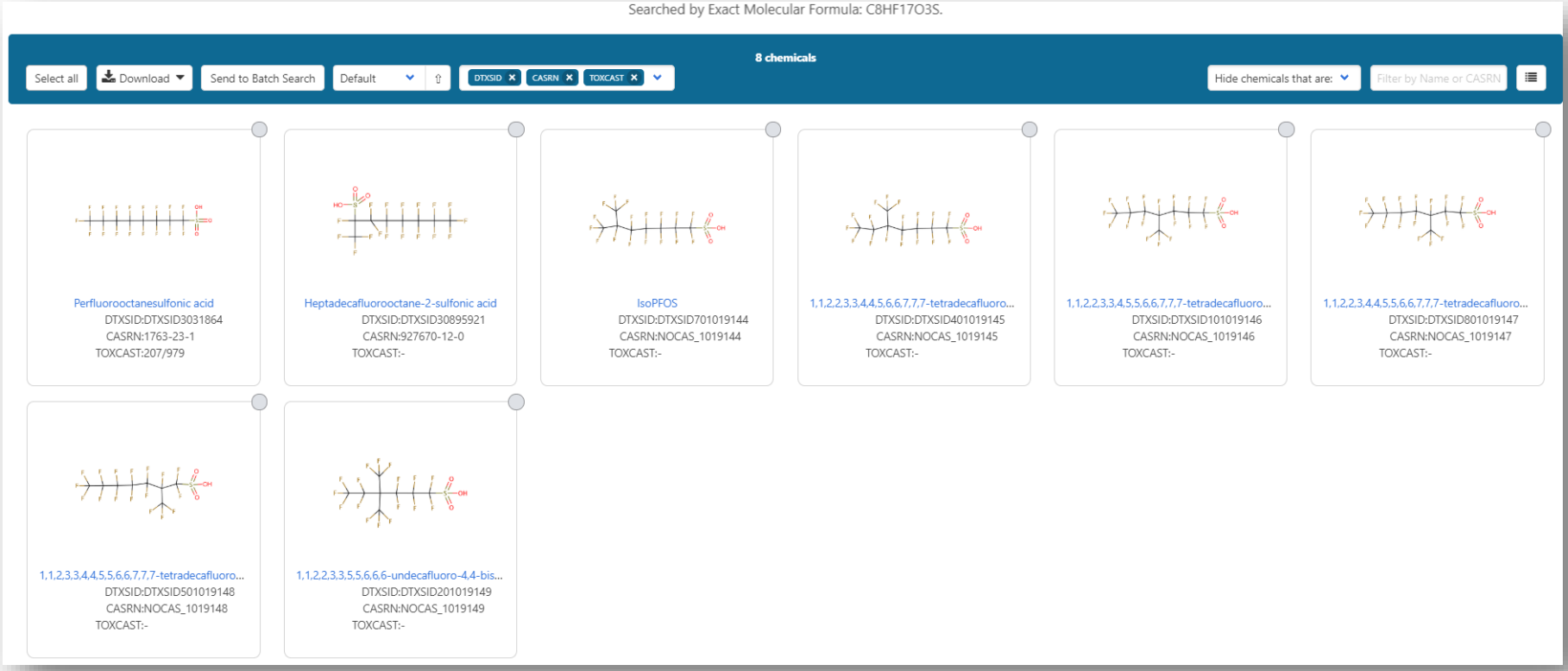
Intrinsic Properties

Molecular Formula: $C_8HF_{17}O_3S$ Mol File Find All Chemicals

Average Mass: 500.13 g/mol Isotope Mass Distribution

Monoisotopic Mass: 499.937494 g/mol

Active expansion of the PFAS list
From 2 to 8 variants of PFOS



Disinfection By-Products

LIST: Disinfection By-Products

Search DBPRODUCTS Chemicals

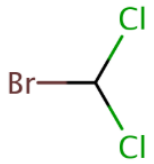
Identifier substring search

List Details

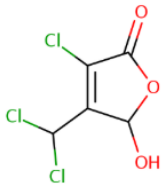
Description: Disinfection by-products (DBPs) result from chemical reactions between organic and inorganic matter in water with chemical treatment agents during the water disinfection process. DBPs are present in most drinking water supplies that have been subject to chlorination, chloramination, ozonation, or treatment with chlorine dioxide.
Number of Chemicals: 87

87 chemicals

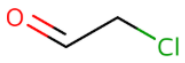
Select all | Download | Send to Batch Search | Default | CASRN | DTXSID | Hide chemicals that are: | Filter by Name or CASRN



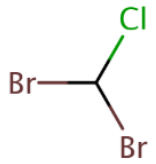
Bromodichloromethane
CASRN:75-27-4
DTXSID:DTXSID1020198



3-Chloro-4-(dichloromethyl)-5-hydroxy-...
CASRN:77439-76-0
DTXSID:DTXSID6020276



Chloroacetaldehyde
CASRN:107-20-0
DTXSID:DTXSID4020292



Chlorodibromomethane
CASRN:124-48-1
DTXSID:DTXSID1020300

Mycotoxins

MASSPECDB: Mycotoxins from MassBank.EU

Search MYCOTOXINS Chemicals

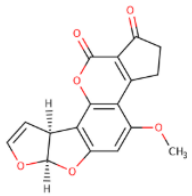
Identifier substring search

List Details

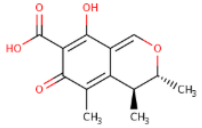
Description: This is a set of mycotoxins, initiated by the contribution of spectra of 90 mycotoxins to MassBank.EU by Justin Renaud and colleagues from Agriculture and Agri-Food Canada, Government of Canada. This list is also a part of the MASSBANKREF list and the NORMAN Suspect Exchange and will be expanded as new contributions arrive.
Number of Chemicals: 88

88 chemicals

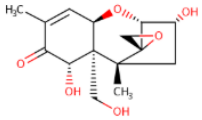
Select all | Download | Send to Batch Search | Default | CASRN | DTXSID | Hide chemicals that are: | Filter by Name or CASRN



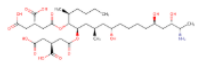
Aflatoxin B1
CASRN:1162-65-8
DTXSID:DTXSID9020035



Citrinin
CASRN:518-75-2
DTXSID:DTXSID8020333

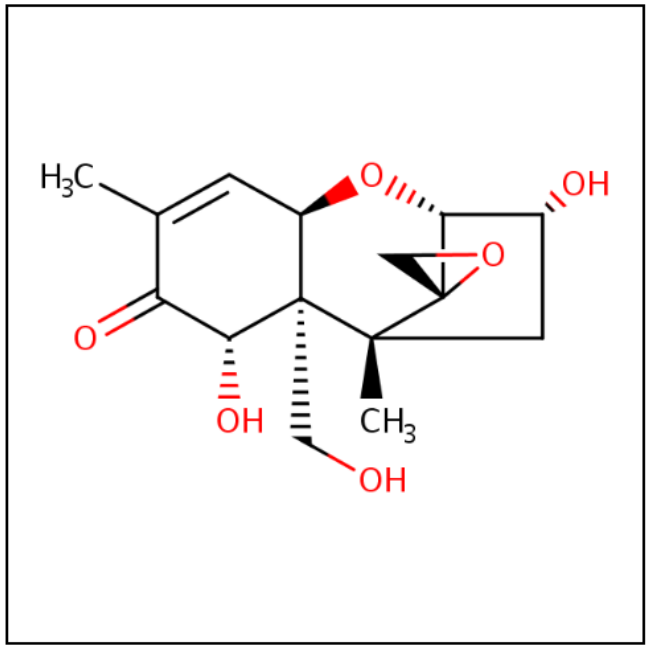


Vomitoxin
CASRN:51481-10-8
DTXSID:DTXSID3020382



Fumonisin B1
CASRN:116355-83-0
DTXSID:DTXSID6020644

Vomitoxin



Structural Identifiers

IUPAC Name: (3alpha,7alpha)-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

InChI String: InChI=1S/C15H20O6/c1-7-3-9-14(5-16,11(19)10(7)18)13(2)4-8(17)12(2

2H3/t8-,9-,11-,12-,13-,14-,15+/m1/s1

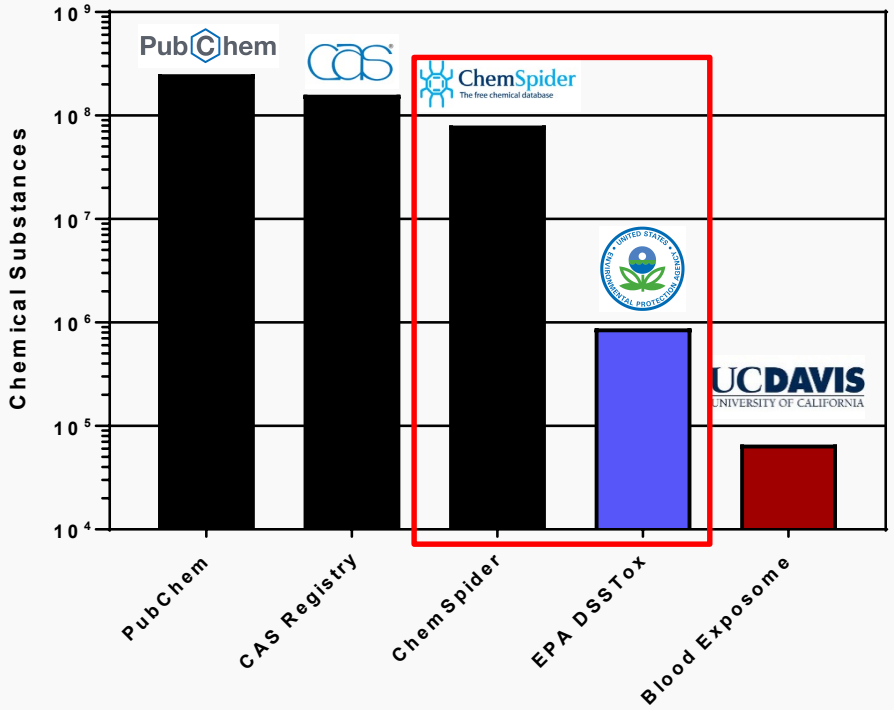
InChIKey: LINOMUASTDIRTM-QGRHZQQGSA-N

Search Google for: Structural Skeleton | Full Structure

Copy All

BIG databases are GREAT!

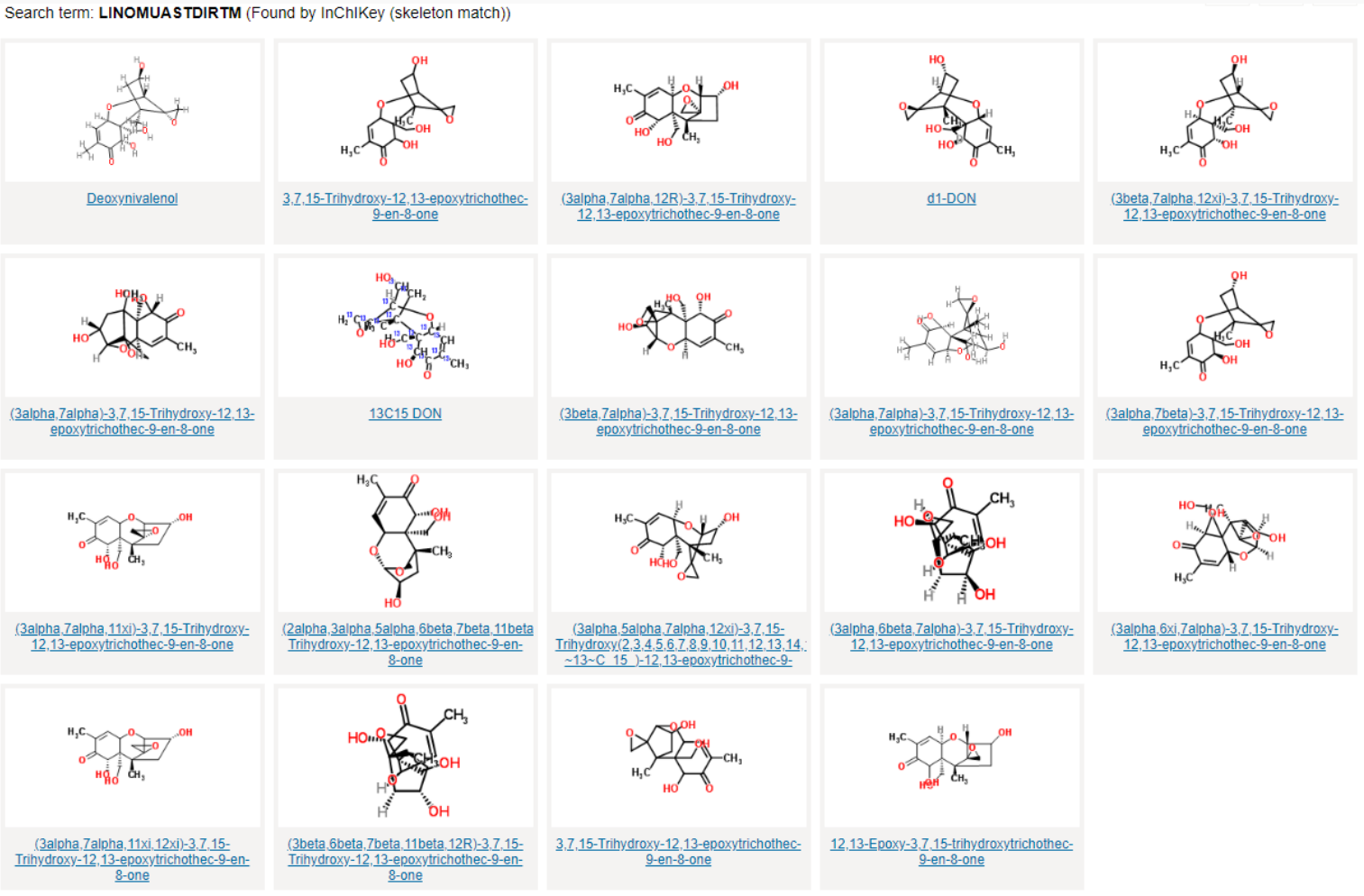
- Thanks to all of the public database efforts
- So much benefit from what's been done
- There are hundreds of them at this point…



Vomitoxin - ChemSpider

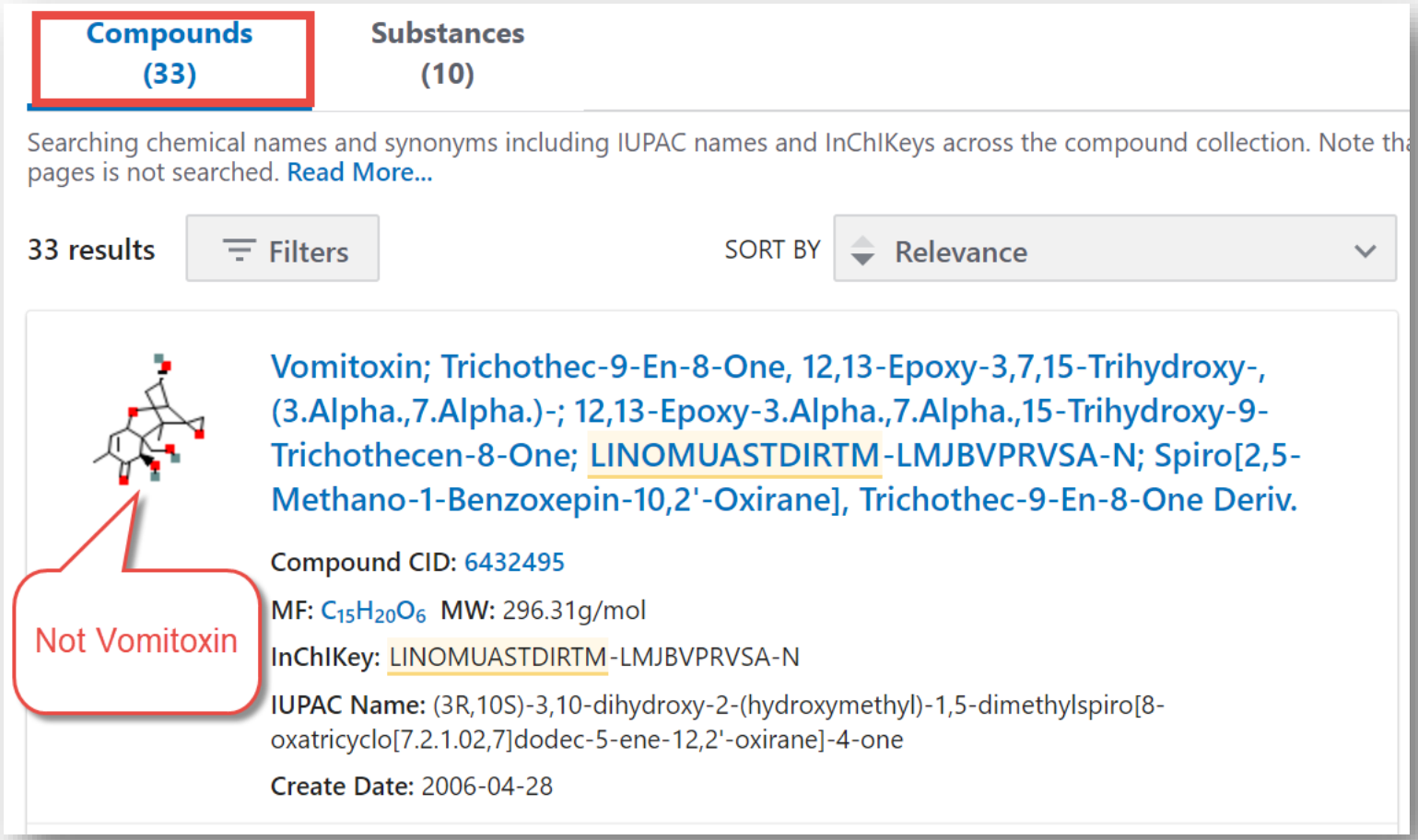
- 19 “Vomitoxins” – 3 isotopically labeled

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



Vomitoxin – PubChem

- 33 unique InChI Keys



PubChem – “virtual chemistry”

- Other databases grow quickly…a lot of “virtual chemistry” and “make on demand” compounds. 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

- The Dashboard database grows slowly (next release is +20k chemicals in 6 months)

ChemSpider – lots of virtuals???

- 52 million chemicals from one vendor

Data Sources

Data Source	Count	Date Created	Last Updated
Aurora Fine Chemicals	51885566	13/04/2009	09/01/2020
Chemspace	14283313	30/11/2016	04/12/2018
AKos	12326374	15/04/2008	09/10/2017
Mcule	9299739	21/01/2014	26/10/2018
Molport	8200357	09/02/2010	09/01/2020
Enamine	3056649	15/04/2008	15/10/2019

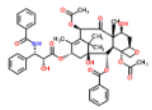
Taxol: 79 Results

EPA United States Environmental Protection Agency

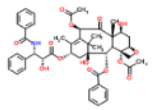
Found 79 results

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

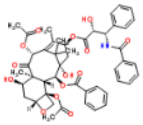
First | < | > | Last



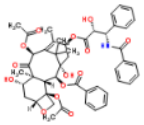
paclitaxel



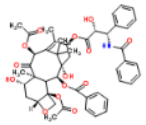
(2alpha,3xi,5beta,7beta,10beta,13alpha)-4 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



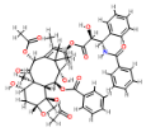
(2alpha,5beta,7alpha,10beta,13alpha)-4,1 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



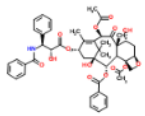
(5beta,7beta,10alpha,13alpha)-4,10-bis(acetyloxy)-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



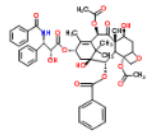
(2alpha,3xi,5beta,7beta,10alpha,13alpha)-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



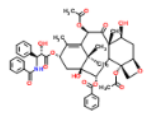
NSC-LSC-1



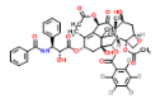
(2alpha,3xi,5beta,7alpha,10beta,13alpha)-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



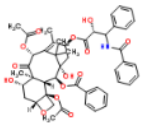
(2alpha,3xi,7beta,10beta,13alpha)-4,10-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



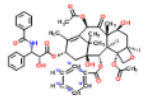
(2alpha,3xi,5beta,7beta,10beta,13alpha)-4 Diacetoxy-13-{[(2S,3S)-3-(benzoylamino)-2-hydroxy-3-



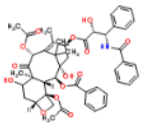
(2alpha,3beta,5beta,7beta,10beta,13alpha Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



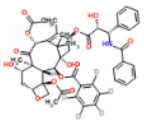
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2R)-3-(benzoylamino)-2-hydroxy-3-phenylpropanoyl]oxy}-1,7-



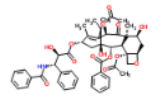
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



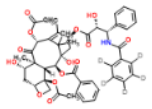
(2alpha,5beta,10beta,13alpha)-4,10-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



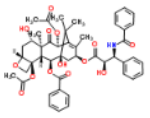
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



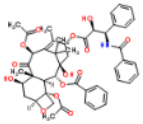
(2alpha,5beta,7beta,8alpha,10beta,13alph Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



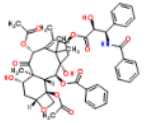
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-1,7-dihydroxy-13-{[(2R,3S)-2-hydroxy-3-phenyl-3-



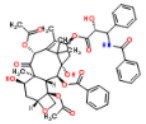
(1b,2a,5b,7b,10b,13a)-4,10-bis(acetyloxy)-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



(1beta,2beta,3beta,4alpha,5alpha,7alpha,1 Diacetoxy-13-{[(2S,3R)-3-(benzoylamino)-2-hydroxy-3-



(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2S,3R)-3-(benzoylamino)-2-hydroxy-3-



(2alpha,5beta,7alpha,10beta,13alpha)-4,1 Diacetoxy-13-{[(2R)-3-(benzoylamino)-2-hydroxy-3-phenylpropanoyl]oxy}-1,7-

Data Quality is important

- Data quality in free web-based databases!

Drug Discovery Today

Volume 17, Issues 13–14, July 2012, Pages 685-701

ELSEVIER

Review

Keynote

Towards a gold standard:
quality in public domain
databases and approaches
the

Antony J.

Show

https://d

Drug Discovery Today

Volume 16, Issues 17–18, September 2011, Pages 747-750

ELSEVIER

Editorial

es

and content

Machines first, humans second: on the importance of algorithmic interpretation of open chemistry data

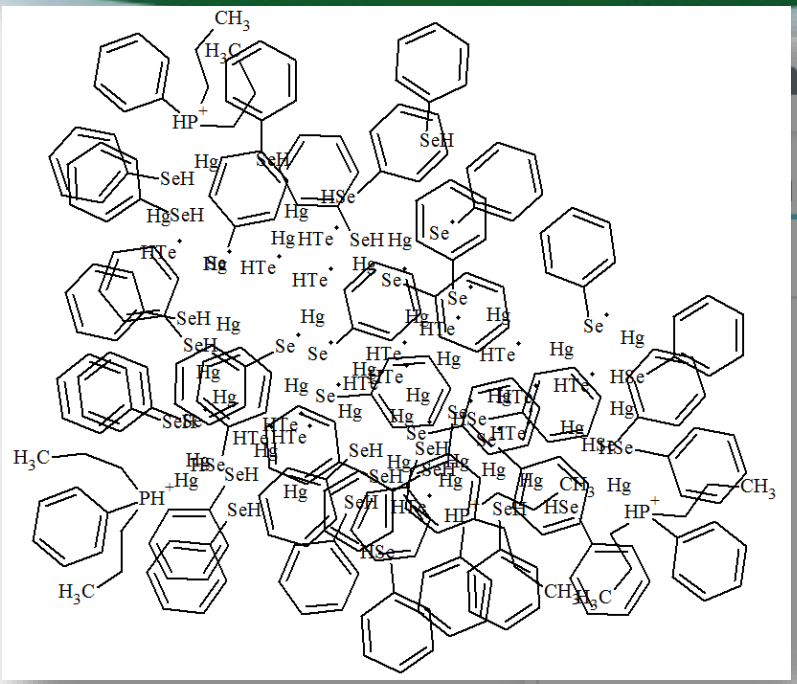
Alex M Clark, Antony J Williams and Sean Ekins

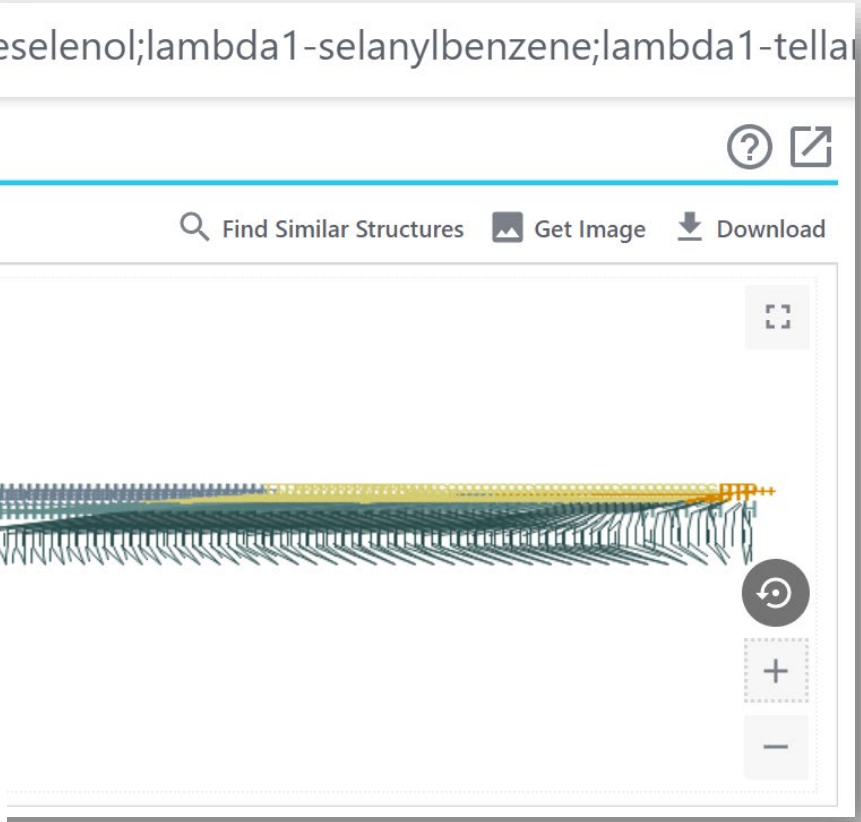
Journal of Cheminformatics 2015 **7**:9

https://doi.org/10.1186/s13321-015-0057-7 | © Clark et al.; licensee Springer. 2015

Received: 24 November 2014 | **Accepted:** 23 February 2015 | **Published:** 22 March 2015

We're still cleaning data too





Record Information

Citation: U.S. Environmental Protection Agency. Chemistry Dashboard. https://comptox.epa.gov/dashboard/DTXSID10581532 (accessed January 24, 2020), PUBCHEM_16048566

Data Quality:

Level 1: Expert curated, highest confidence in accuracy and consistency of unique chemical identifiers
Level 2: Expert curated, unique chemical identifiers using multiple sources
Level 3: Programmatically curated from high quality EPA source, unique chemical identifiers have no conflicts in ChemID and PubChem
Level 4: Programmatically curated from ChemID, unique chemical identifiers have no conflicts in PubChem
Level 5: Programmatically curated from ACToR or PubChem, unique chemical identifiers with low confidence, single public source

Tire Crumb Rubber (298)

EPA United States Environmental Protection Agency

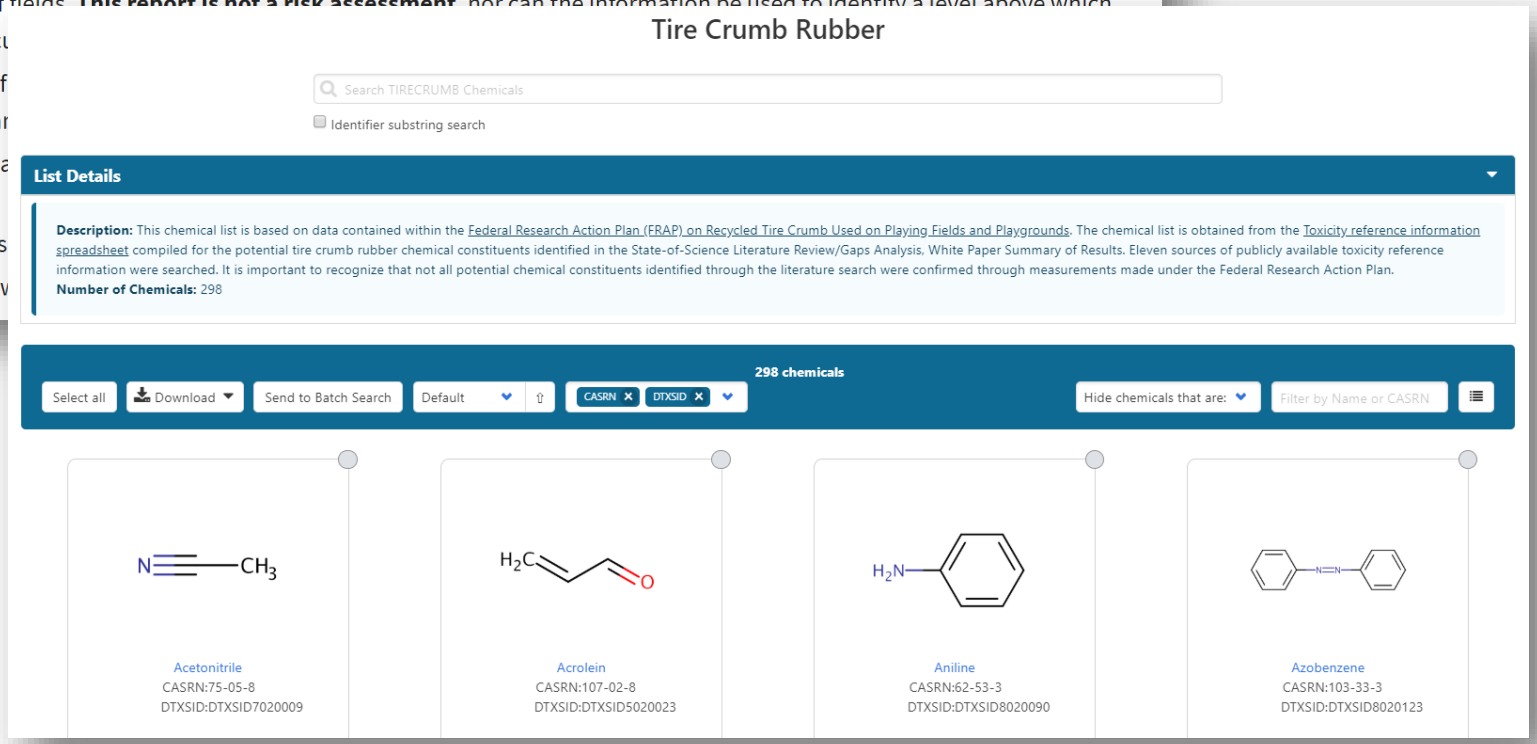
Related Topics: Safer Chemicals Research

CONTACT US SHARE

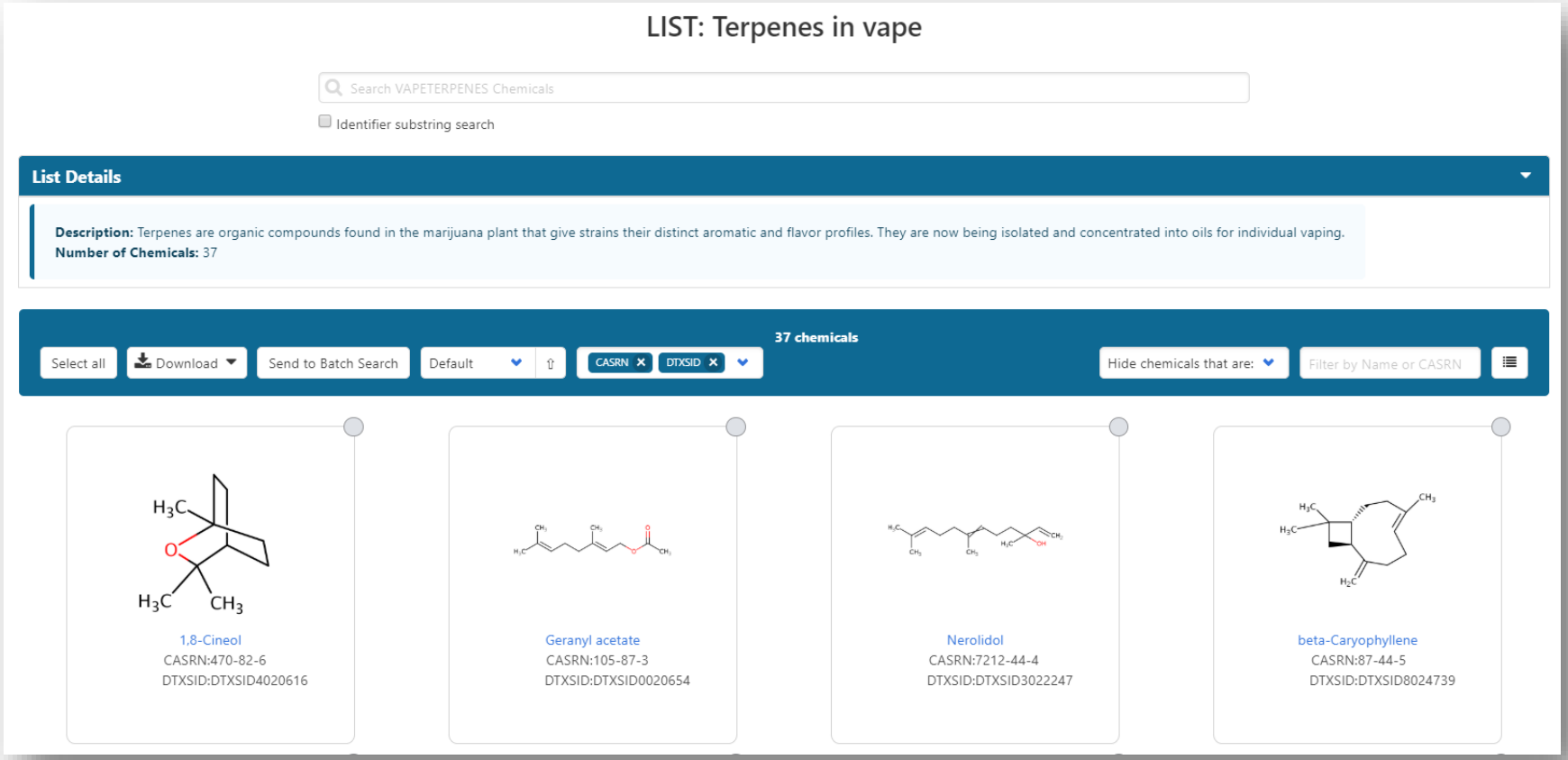
July 2019 Report: Tire Crumb Rubber Characterization

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 occu
- In general, the findings f human exposure appear
- Only Part 1 is being relea assessment.
- Part 1 of this report pres
- The scope of this study w



Terpenes in Vape (37)



Hydraulic Fracturing (1640)

EPA United States Environmental Protection Agency

Contact Us

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

- 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

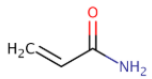
*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances (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

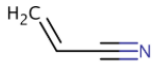
1640 chemicals

Select all | Download | Send to Batch Search | Default | CASRN | DTXSID | Hide chemicals that are: | Filter by Name or CASRN

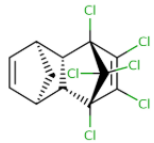
Acrolein
CASRN:107-02-8
DTXSID:DTXSID5020023



Acrylamide
CASRN:79-06-1
DTXSID:DTXSID5020027

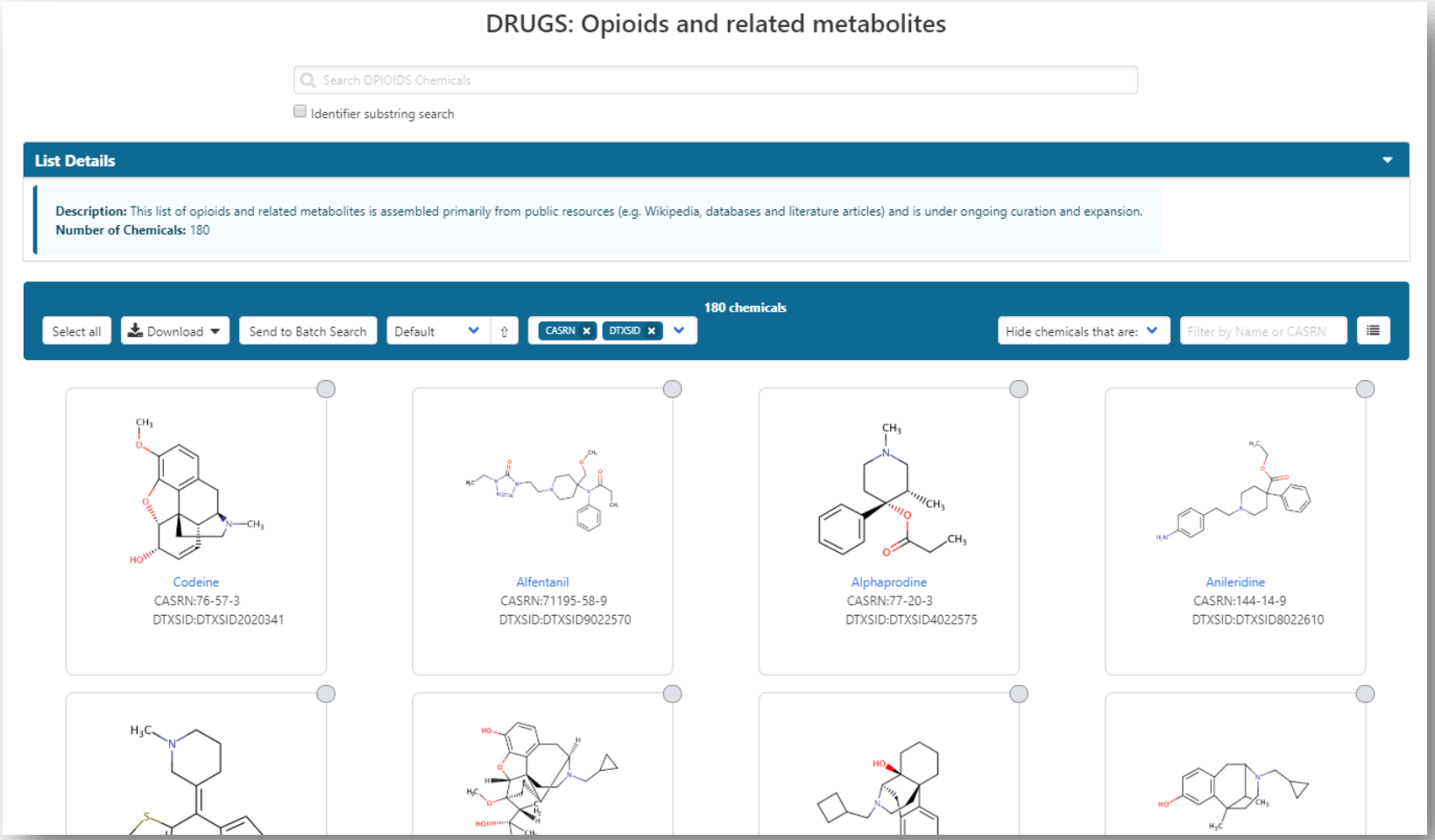


Acrylonitrile
CASRN:107-13-1
DTXSID:DTXSID5020029

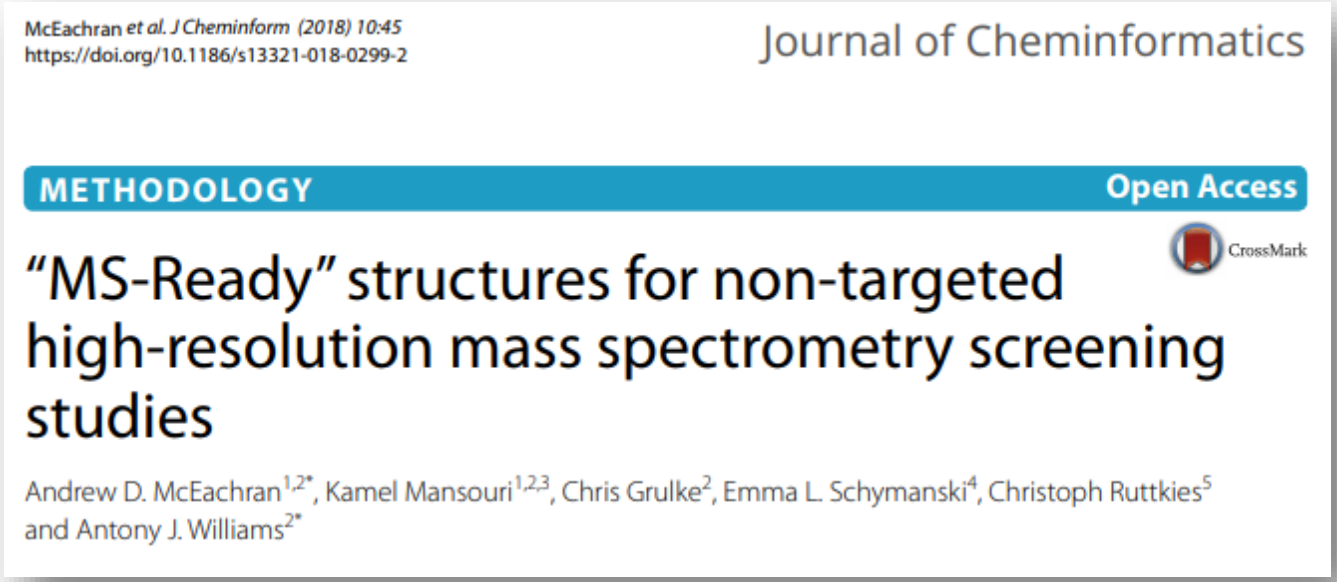


Aldrin
CASRN:309-00-2
DTXSID:DTXSID8020040

Opioids and Metabolites (160)



"MS-ready" structures



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

Journal of Cheminformatics

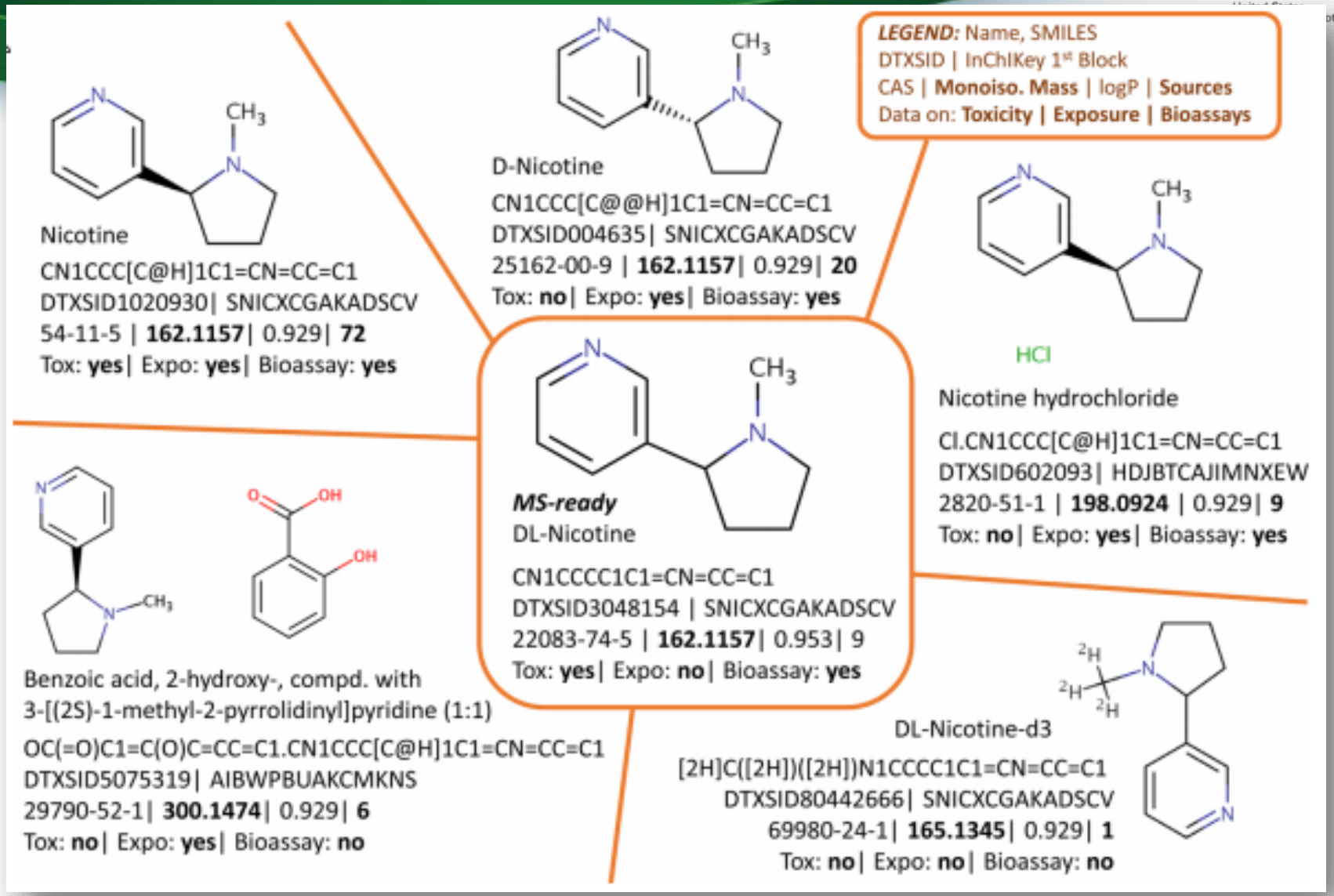
METHODOLOGY Open Access

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

Andrew D. McEachran[1,2*], Kamel Mansouri[1,2,3], Chris Grulke[2], Emma L. Schymanski[4], Christoph Ruttkies[5] and Antony J. Williams[2*]

Overview of MS-Ready Structures

- All structure-based chemical substances are algorithmically processed to
  - Split multicomponent chemicals into individual structures
  - Desalt and neutralize individual structures
  - Remove stereochemical bonds from all chemicals
- MS-Ready structures are then mapped to original substances to provide a path between chemicals detected by mass spectrometry to original substances

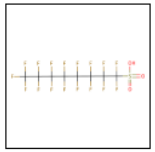


Environmental Science & Technology
Viewpoint
pubs.acs.org/est

Open Science for Identifying "Known Unknown" Chemicals

Emma L. Schymanski*,† and Antony J. Williams*,‡

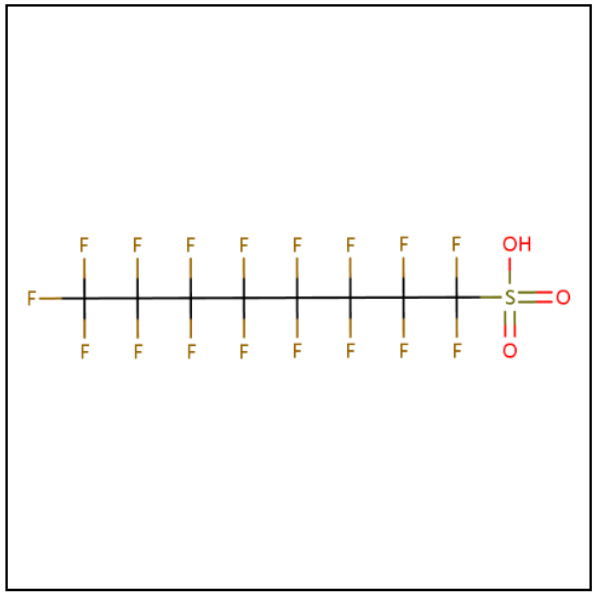
MS-Ready Mappings from Details Page



Perfluorooctanesulfonic acid

1763-23-1 | DTXSID3031864

Searched by Synonym from Valid Source.



Wikipedia

Perfluorooctanesulfonic acid (conjugate base **perfluorooctanesulfonate**) (**PFOS**) is an anthropogenic fluorosurfactant and global pollutant. PFOS was the key ingredient in Scotchgard, a fabric protector made by 3M, and numerous stain repellents. It was added to Annex B of the Stockholm Convention on Persistent Organic Pollutants in May 2009. PFOS can be synthesized in industrial production or result from the degradation of precursors. PFOS levels that have been detected in wildlife

...

Read more

Quality Control Notes

Intrinsic Properties

Structural Identifiers

Linked Substances

Same Connectivity: 4 records (based on first layer of InChI)

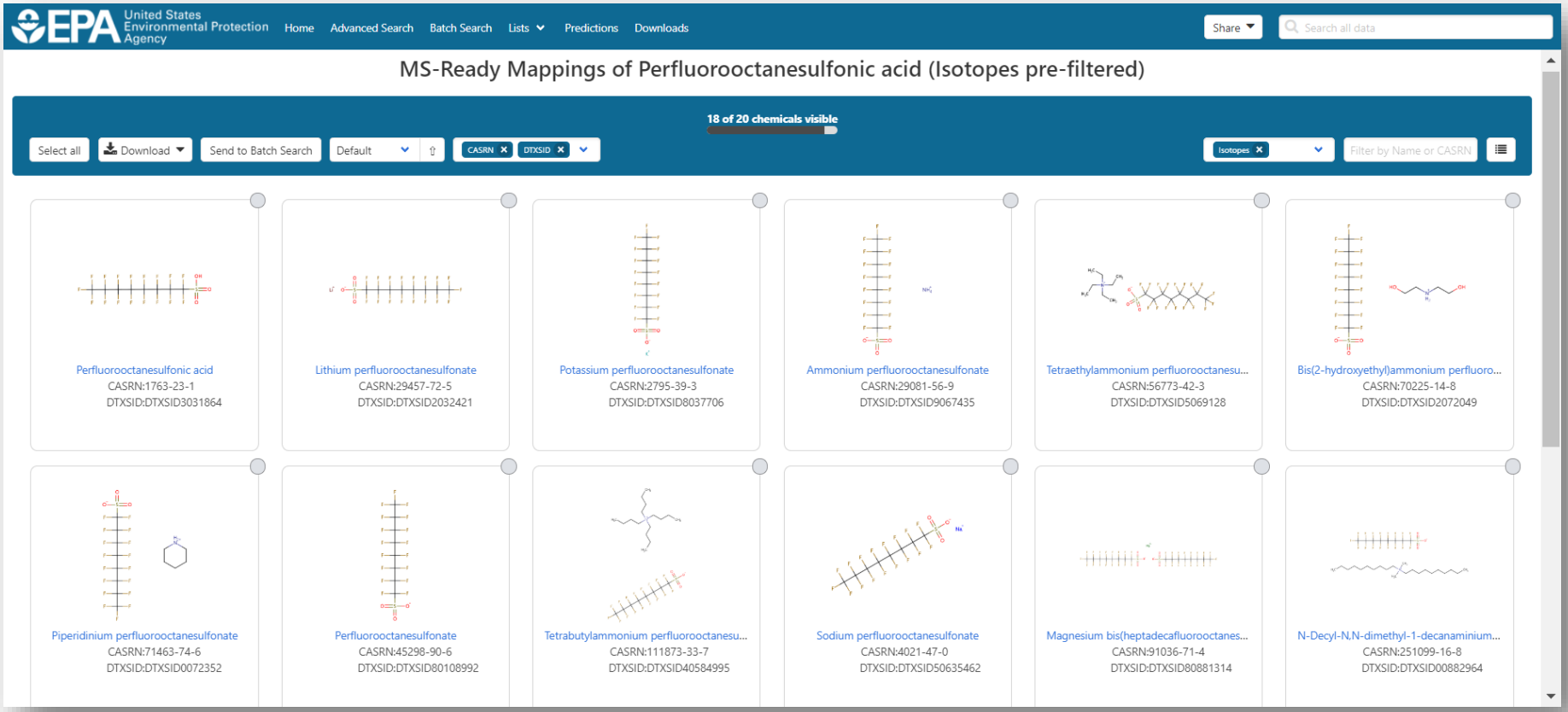
Mixtures, Components and Neutralized Forms: 9 records (based on QSAR ready mappings and with the compound as a component of a mixture)

MS-Ready Mappings: DTXCID1011864: 18 records;

Similar Compounds: 83 records (based on Tanimoto coefficient >0.8)

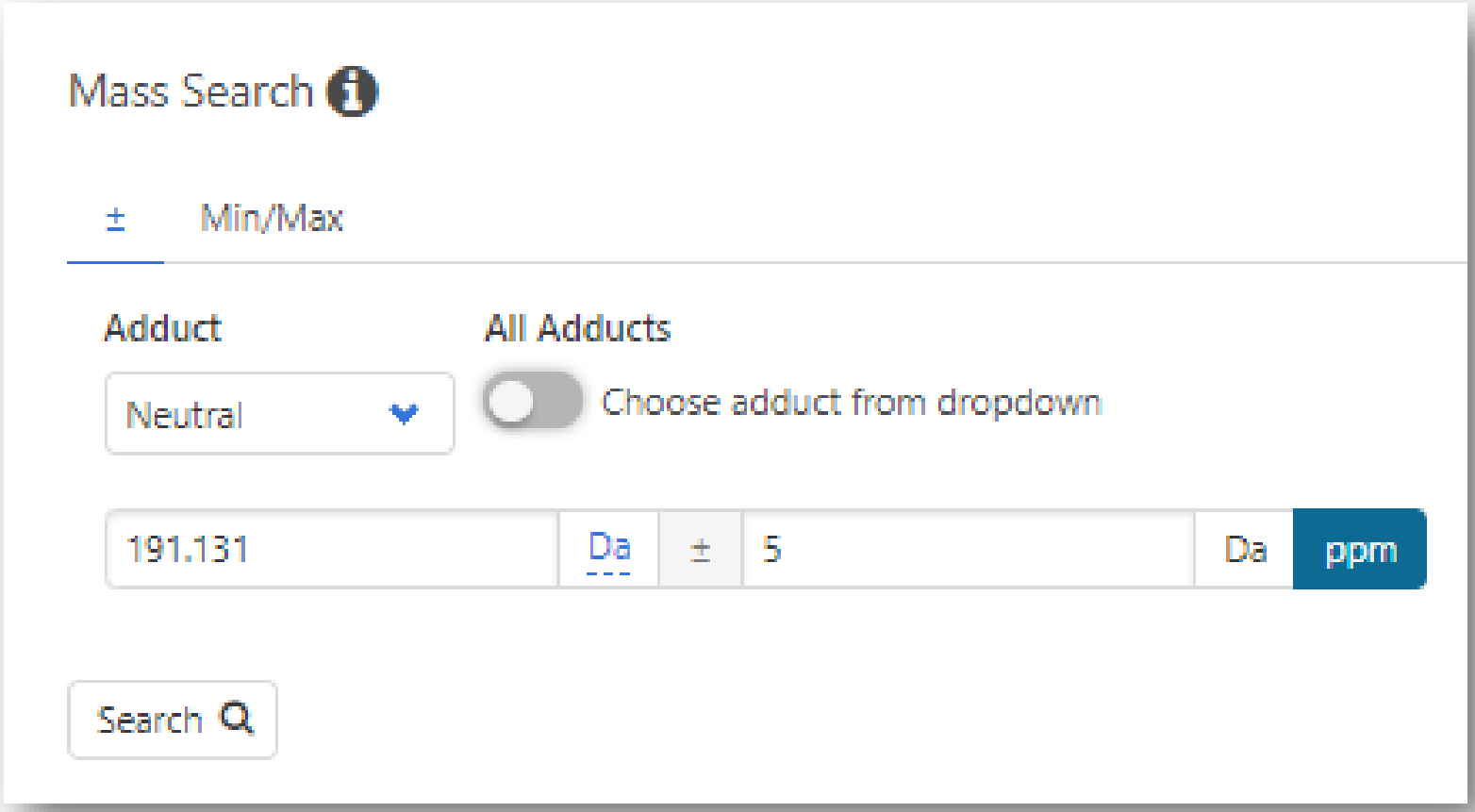
ed_search/index

MS-Ready Mappings Set of 20 substances for “PFOS”

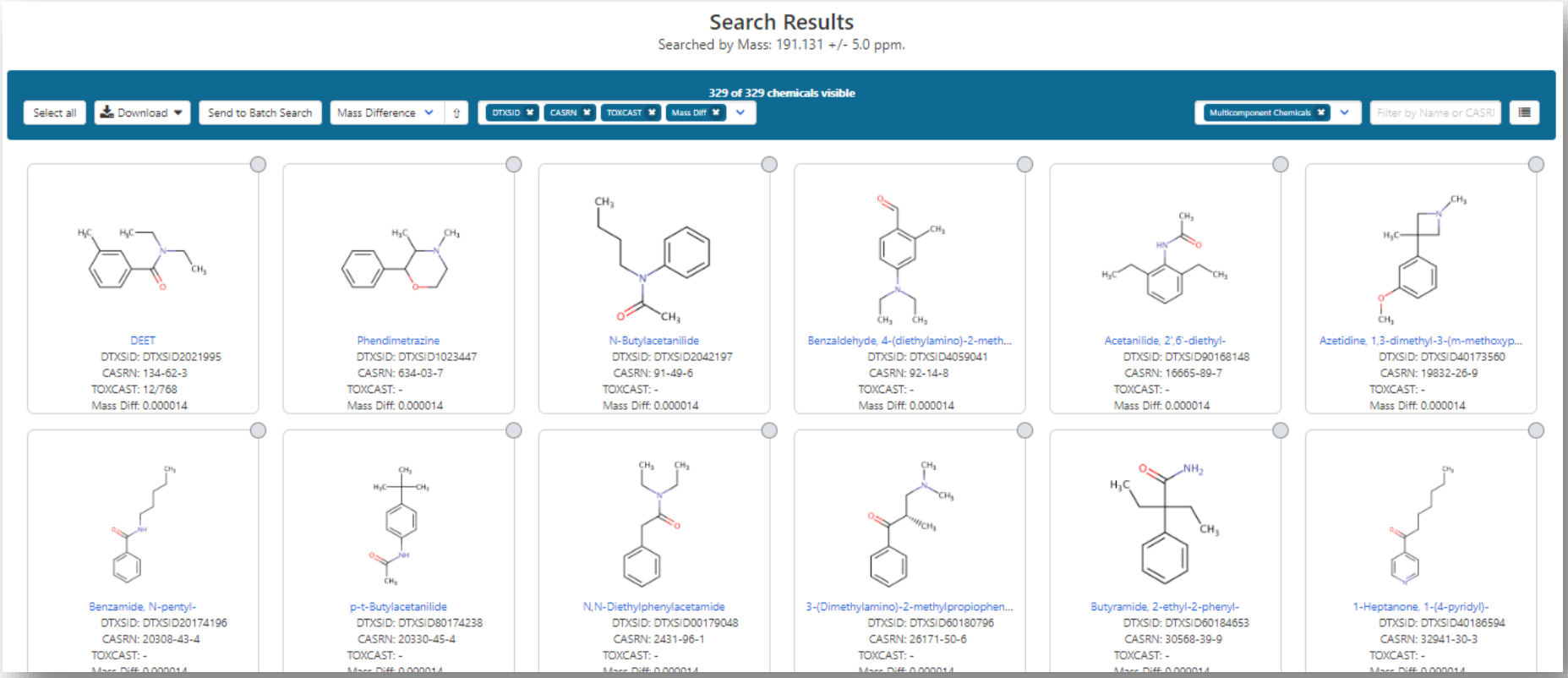


Mass and Formula Searching

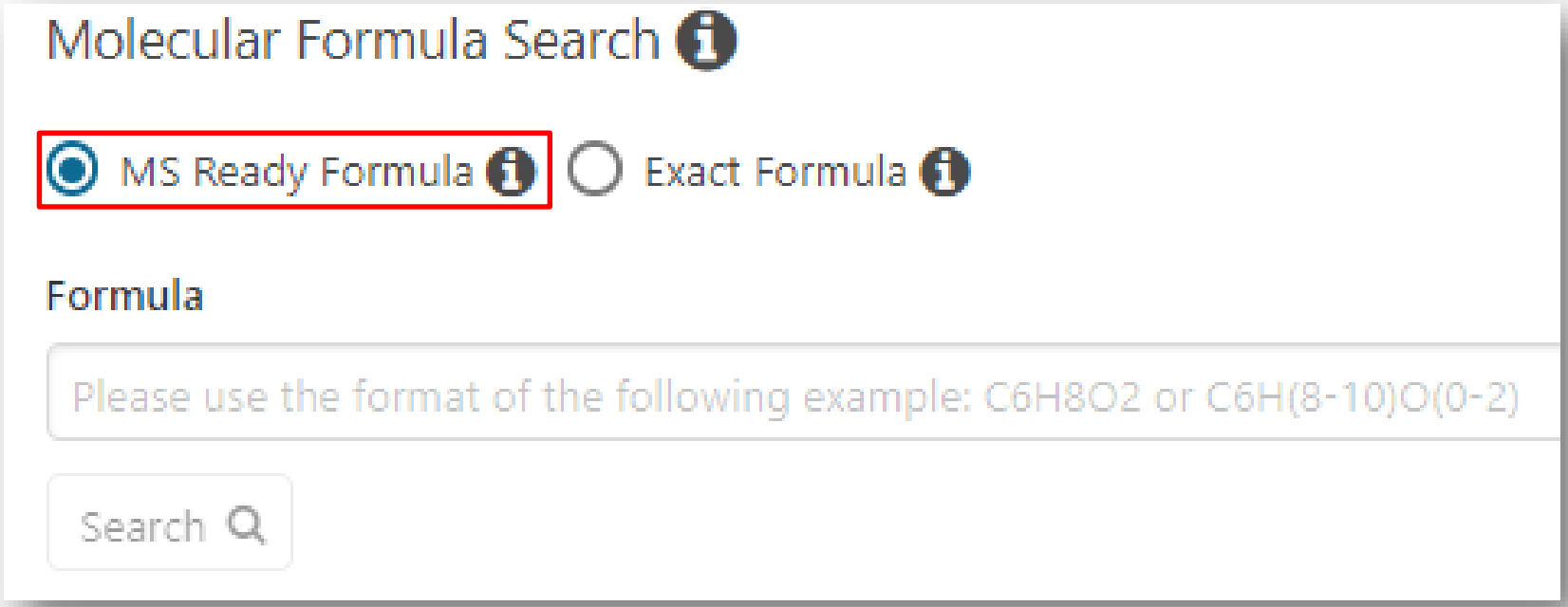
Advanced Searches
Mass Search



Advanced Searches
Mass Search

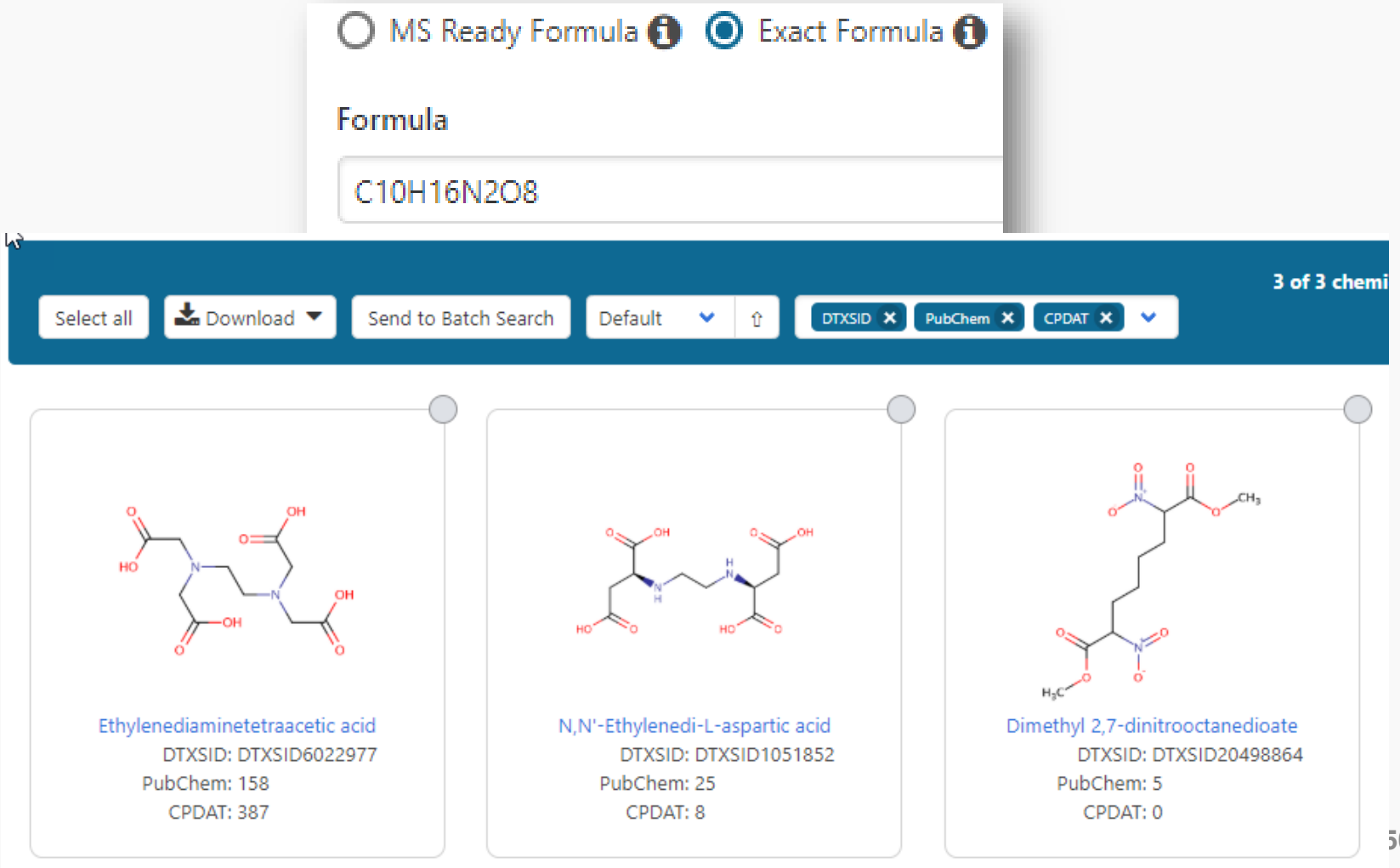


MS-Ready Structures for **Formula Search**



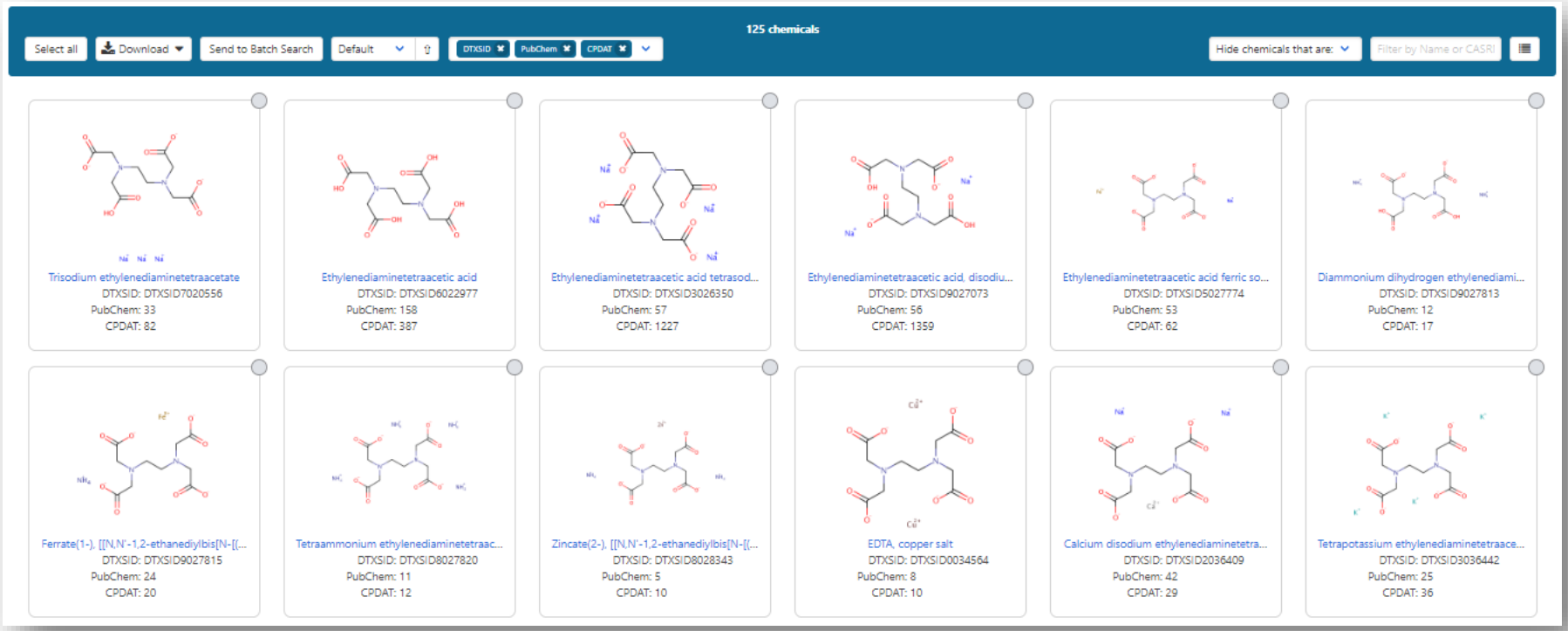
MS-Ready Mappings

- **EXACT Formula**: C10H16N2O8: 3 Hits



MS-Ready Mappings

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



MS-Ready Mappings

- 125 chemicals returned in total
  - 8 of the 125 are **single component** chemicals
  - 3 of the 8 are **isotope-labeled**
  - 3 are **neutral compounds** and 2 are **charged**

- Multiple components, stereo, isotopes and charge all collapsed and mapped through MS-Ready

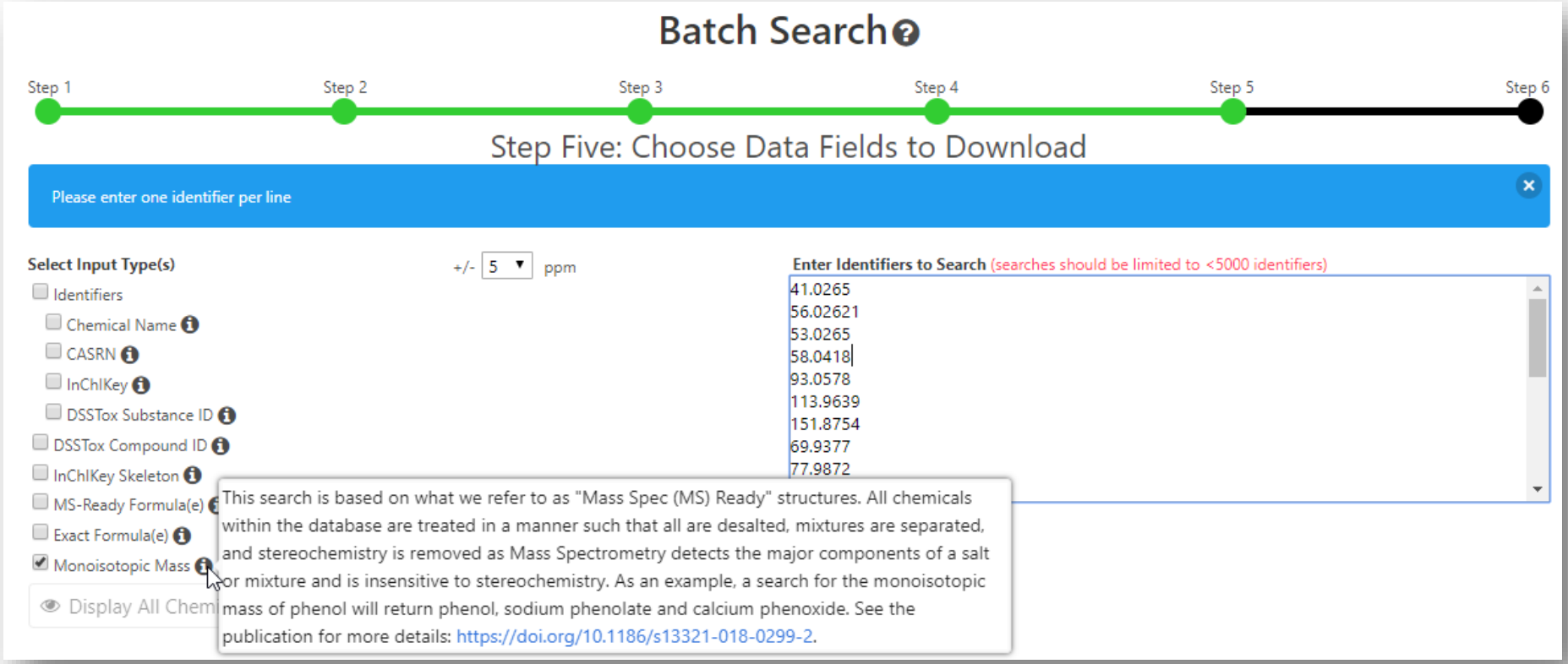
Batch Searching mass and formula

Batch Searching

- Singleton searches are useful but we work with **thousands** of masses and formulae!

- Typical questions
  - What is the list of chemicals for the formula $C_xH_yO_z$
  - What is the list of chemicals for a mass +/- error
  - Can I get chemical lists in Excel files? In SDF files?
  - Can I include properties in the download file?

Batch Searching Formula/Mass



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

EPA United States Environmental Protection Agency

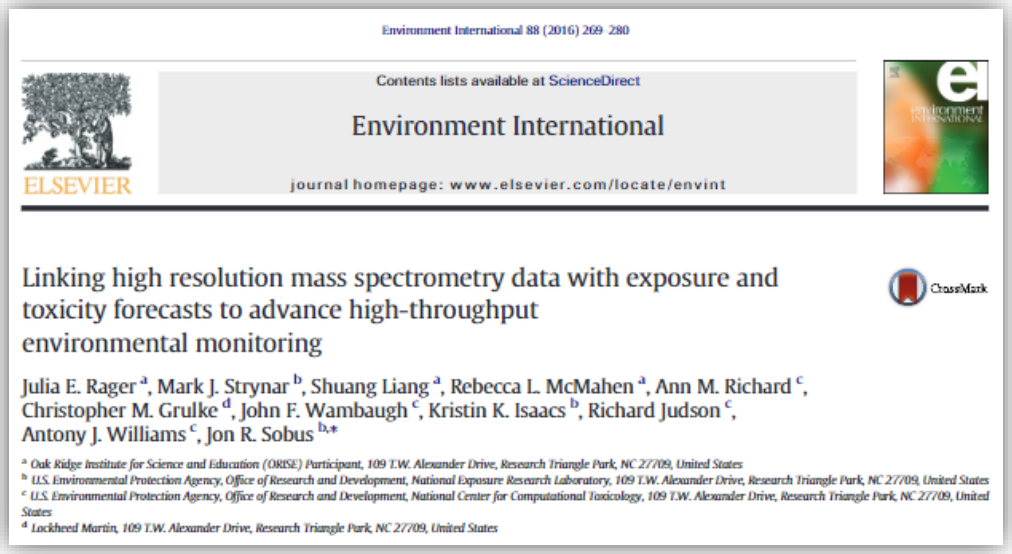
	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	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	DTXSID6026567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	60

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
Hydromorpl	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
Propoxyphe	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*



Environment International 88 (2016) 269–280

Contents lists available at ScienceDirect

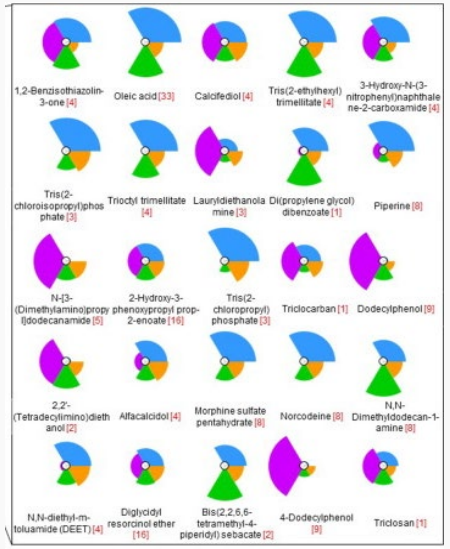
Environment International

journal homepage: www.elsevier.com/locate/envint

Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring

Julia E. Rager [a], Mark J. Strynar [b], Shuang Liang [a], Rebecca L. McMahen [a], Ann M. Richard [c], Christopher M. Grulke [d], John F. Wambaugh [c], Kristin K. Isaacs [b], Richard Judson [c], Antony J. Williams [c], Jon R. Sobus [b,*]

[a] Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[b] U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[c] U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[d] Lockheed Martin, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States



EPA United States Environmental Protection Agency

Candidate ranking using metadata

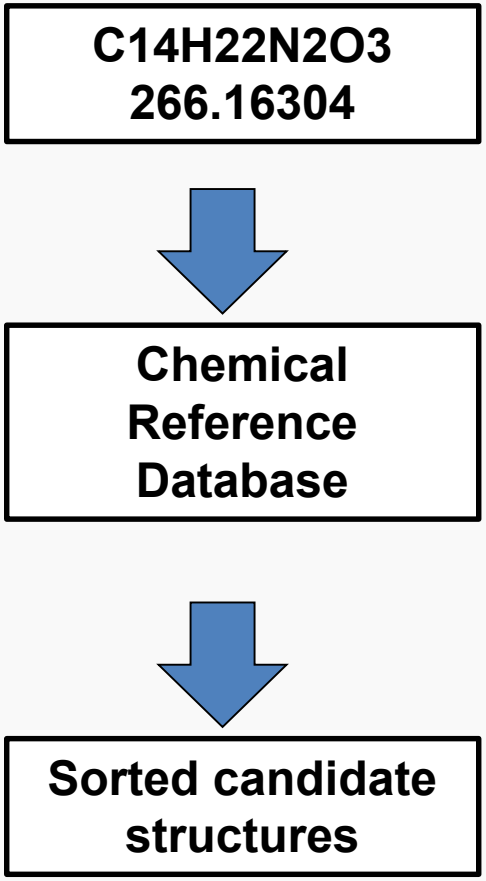
J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

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

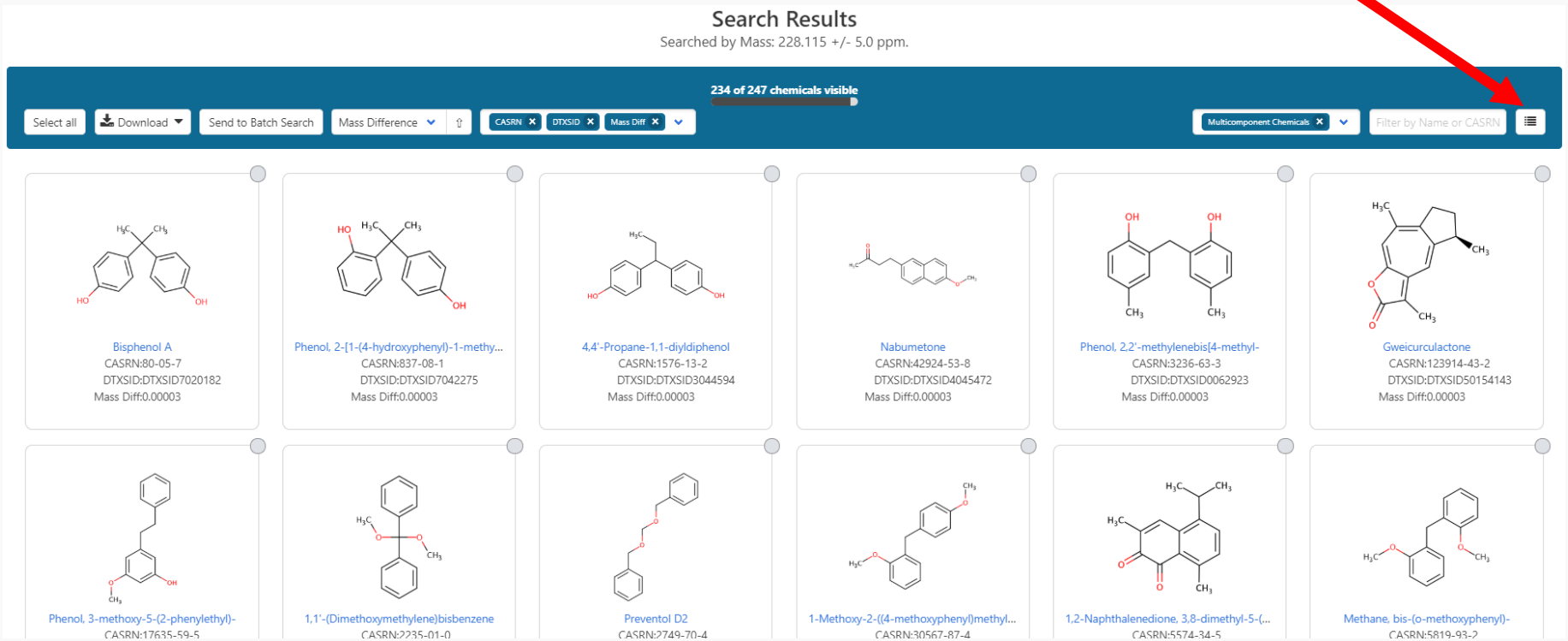


Data Streams for Ranking

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

Search 228.115 +/- 5.0 ppm
234 single component chemicals

EPA United States Environmental Protection Agency



Search 228.115 +/- 5.0 ppm
234 single component chemicals

CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts
80-05-7	Level 1	326	170	161	3850
42924-53-8	Level 2	14	45	138	342
87619-52-1	Level 5	0	2		0
87607-32-7	Level 5	0	2		0

The original ChemSpider work

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			

Is a bigger database better?

- ChemSpider was 26 million chemicals for the original work
- Much BIGGER today
- Is bigger better??
- Are there other metadata to use for ranking?

Comparing Search Performance

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. McEachran[1] · Jon R. Sobus[2] · Antony J. Williams[3]

- 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?

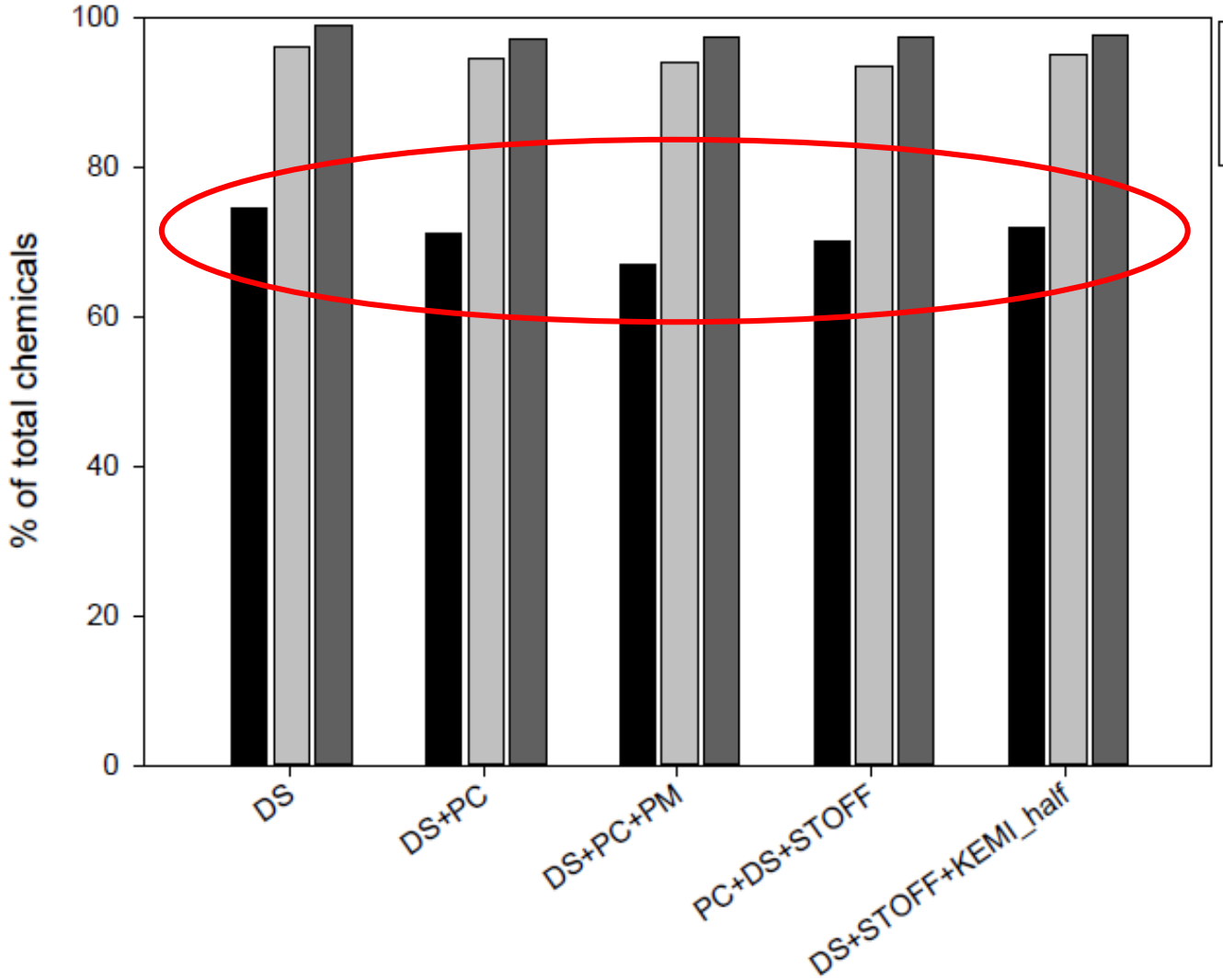
	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2[a]	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

[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

Identification ranks for **1783 chemicals** using multiple data streams

Data Sources alone rank ~75% of the chemicals as Top Hit



DS: Data Sources
PC: PubChem
PM: PubMed
STOFF: DB
KEMI: DB

"UVCB" Chemicals

UVCB Chemicals

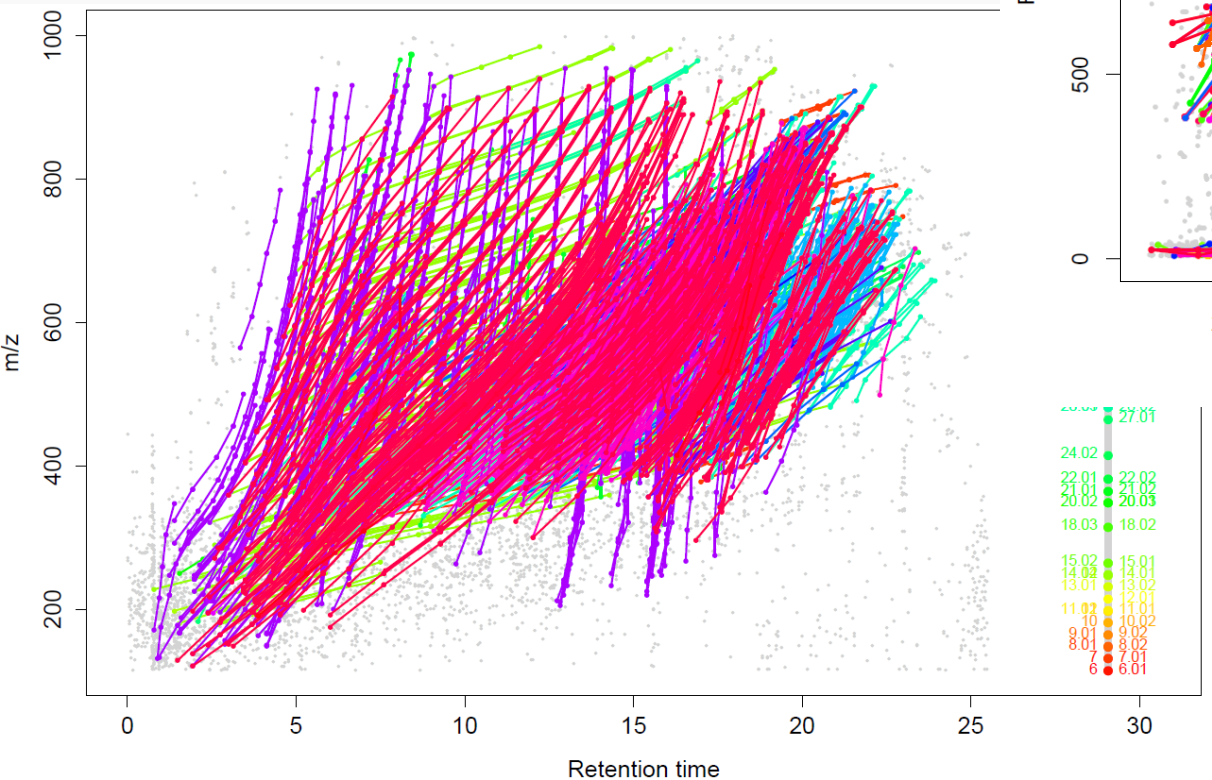
EPA United States Environmental Protection Agency

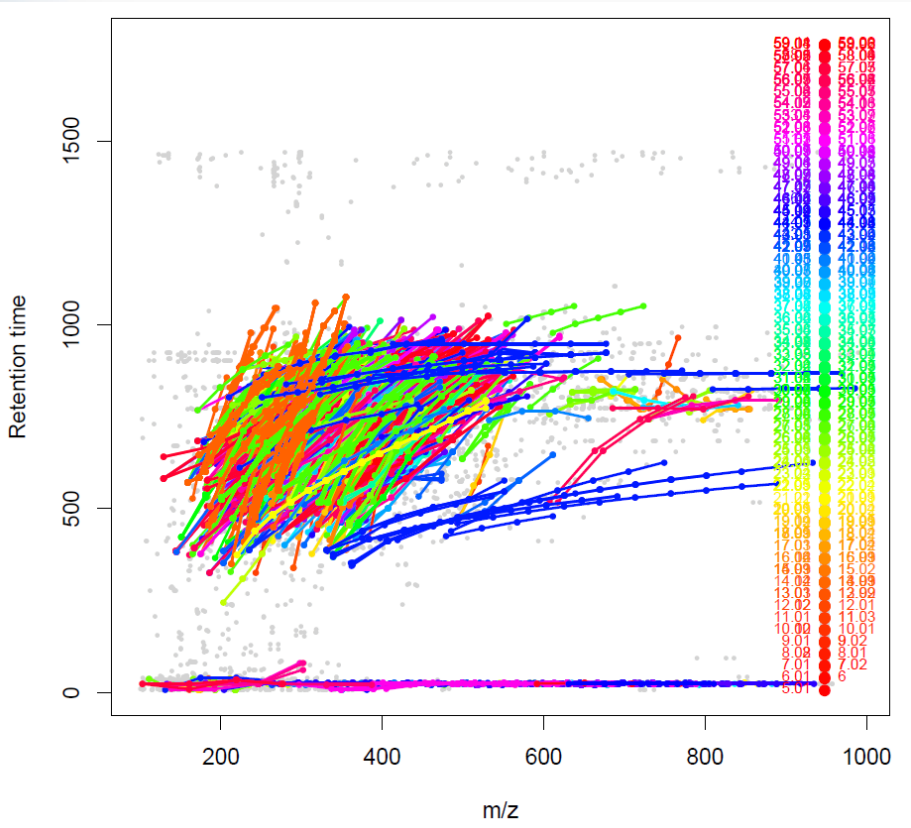
Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB Substance) on the TSCA Inventory

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.

UVCBs challenge in non-target analysis

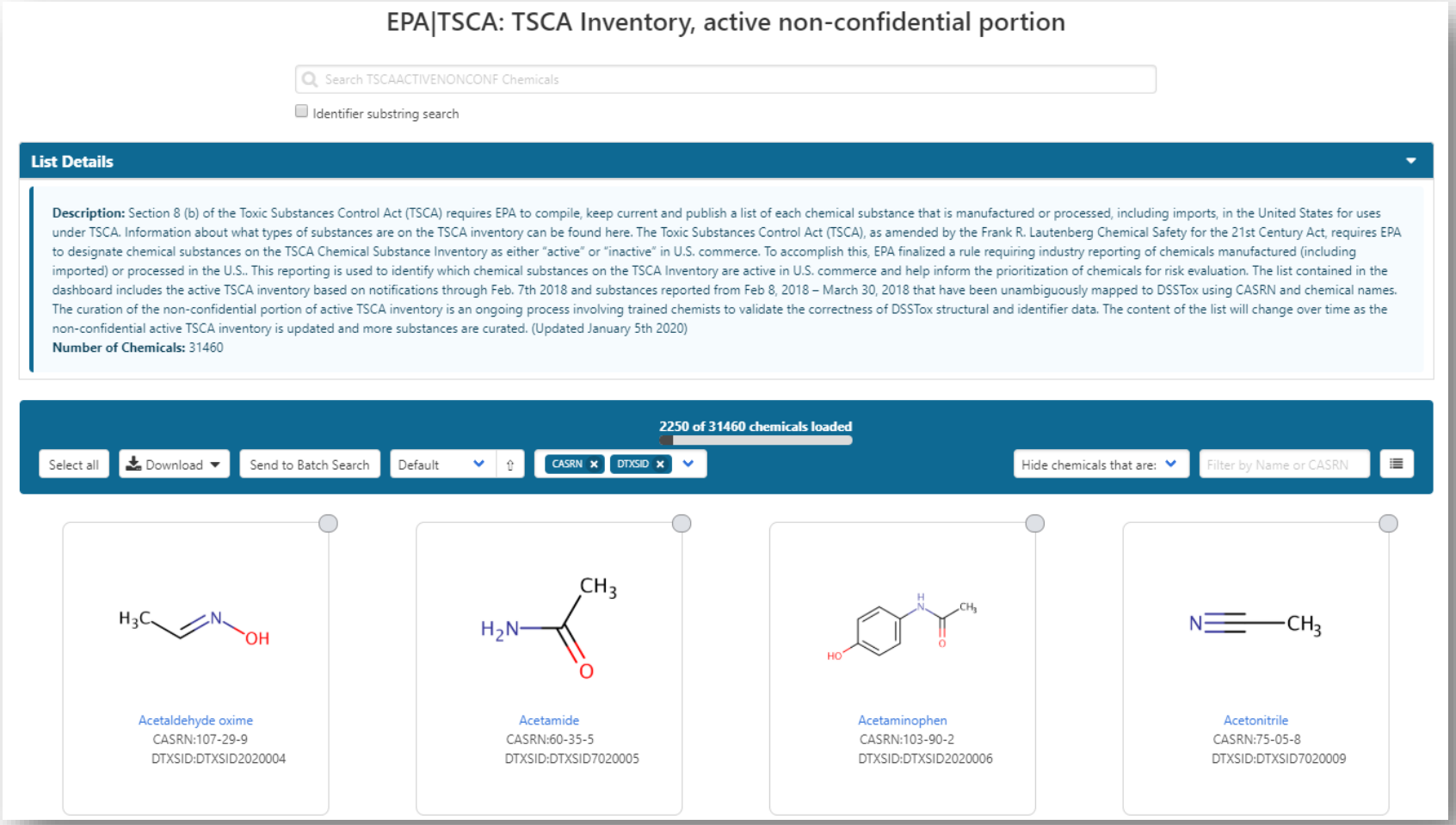
- Complex mixtures (UVCBs) are a ***huge*** and ***very challenging*** part of the unknowns in many environmental samples





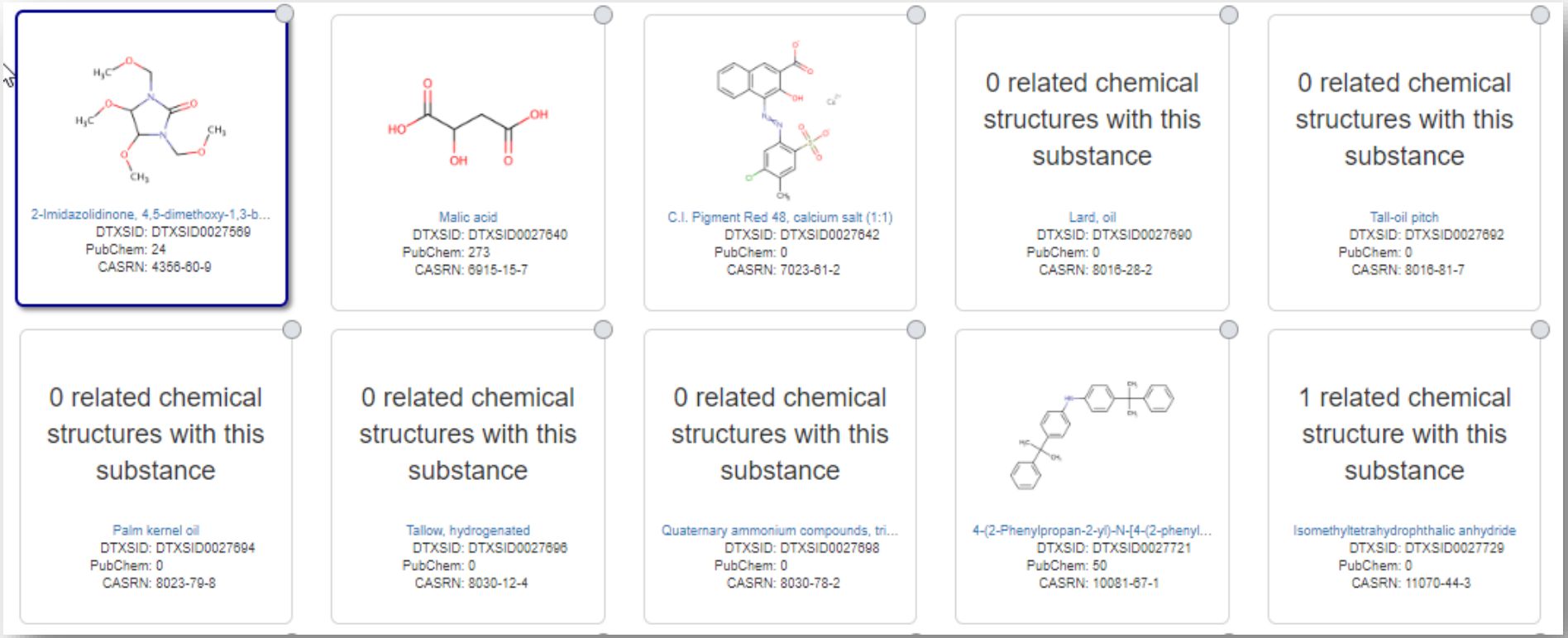
Homologue screening plots from Swiss Wastewater (Schymanski *et al* 2014, left) and Novi Sad (right)

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



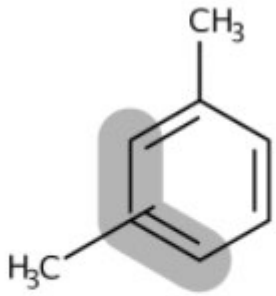
Many Chemicals are “Complex”
>14000 chemicals are UVCBs

EPA United States Environmental Protection Agency

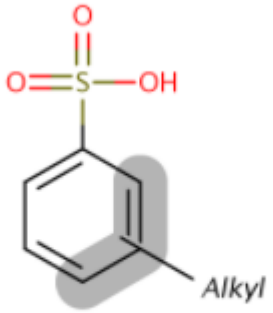


“Markush Structures”

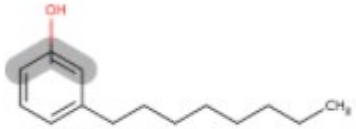
https://en.wikipedia.org/wiki/Markush_structure



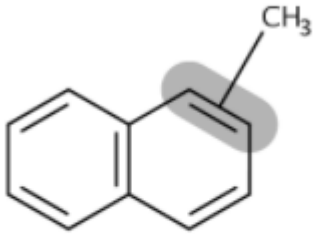
Xylenes
1330-20-7



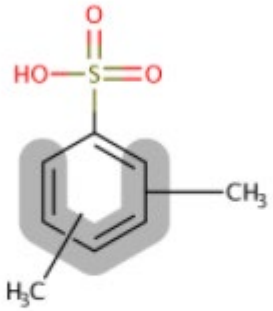
(C10-C16) Alkylbenzenesulfonic acid
68584-22-5



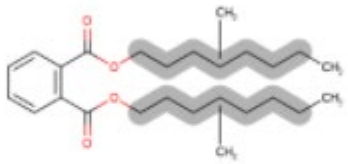
n-Nonylphenol
25154-52-3



Methylnaphthalene
1321-94-4

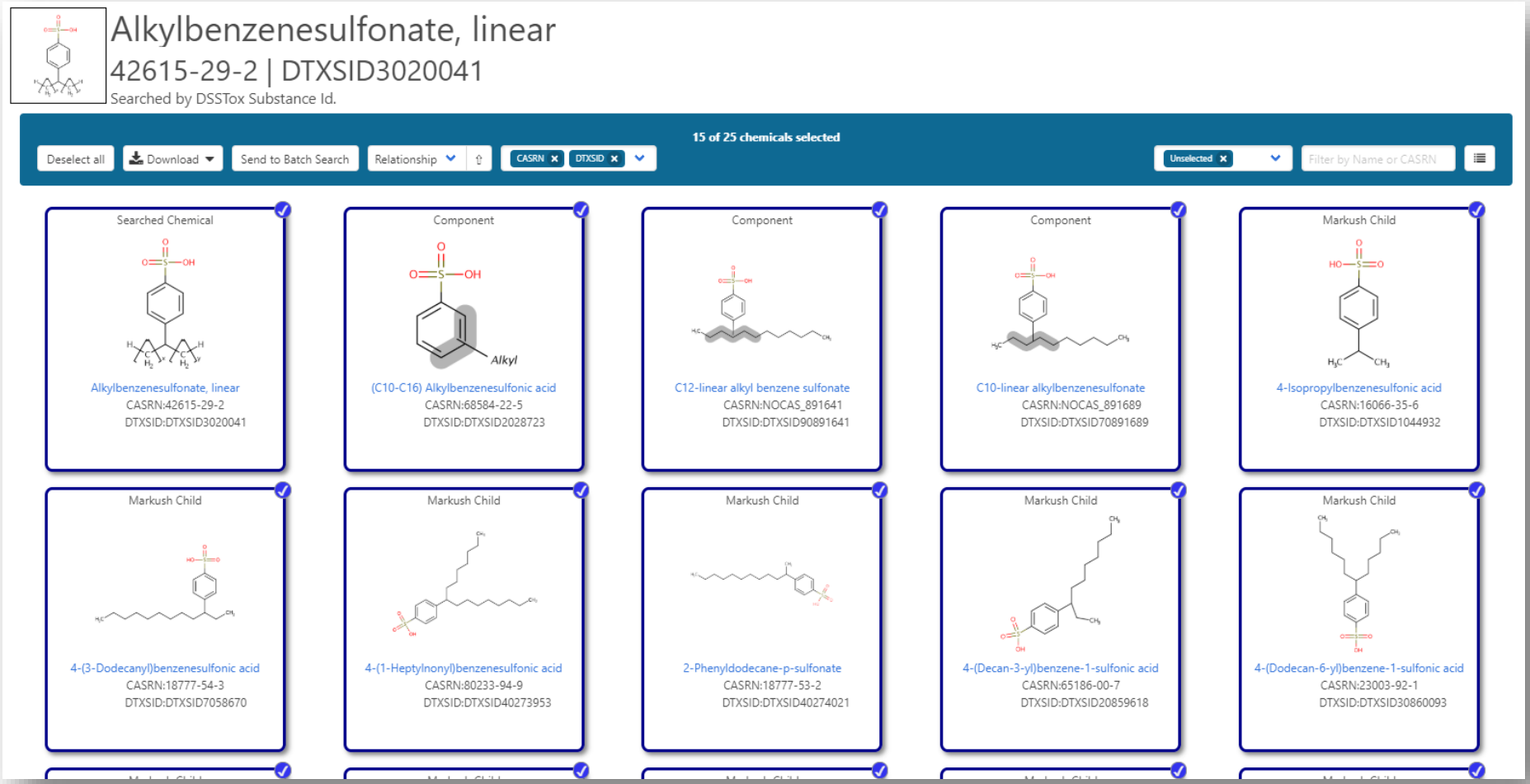


Sodium xylenesulfonate
1300-72-7



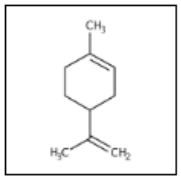
Diisononyl phthalate
28553-12-0

How to represent complexity?



In the Dashboard Abstract Sifter

Literature Searching



Limonene

138-86-3 | DTXSID2029612

Searched by Approved Name.

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

Select a Query Term | Retrieve Articles

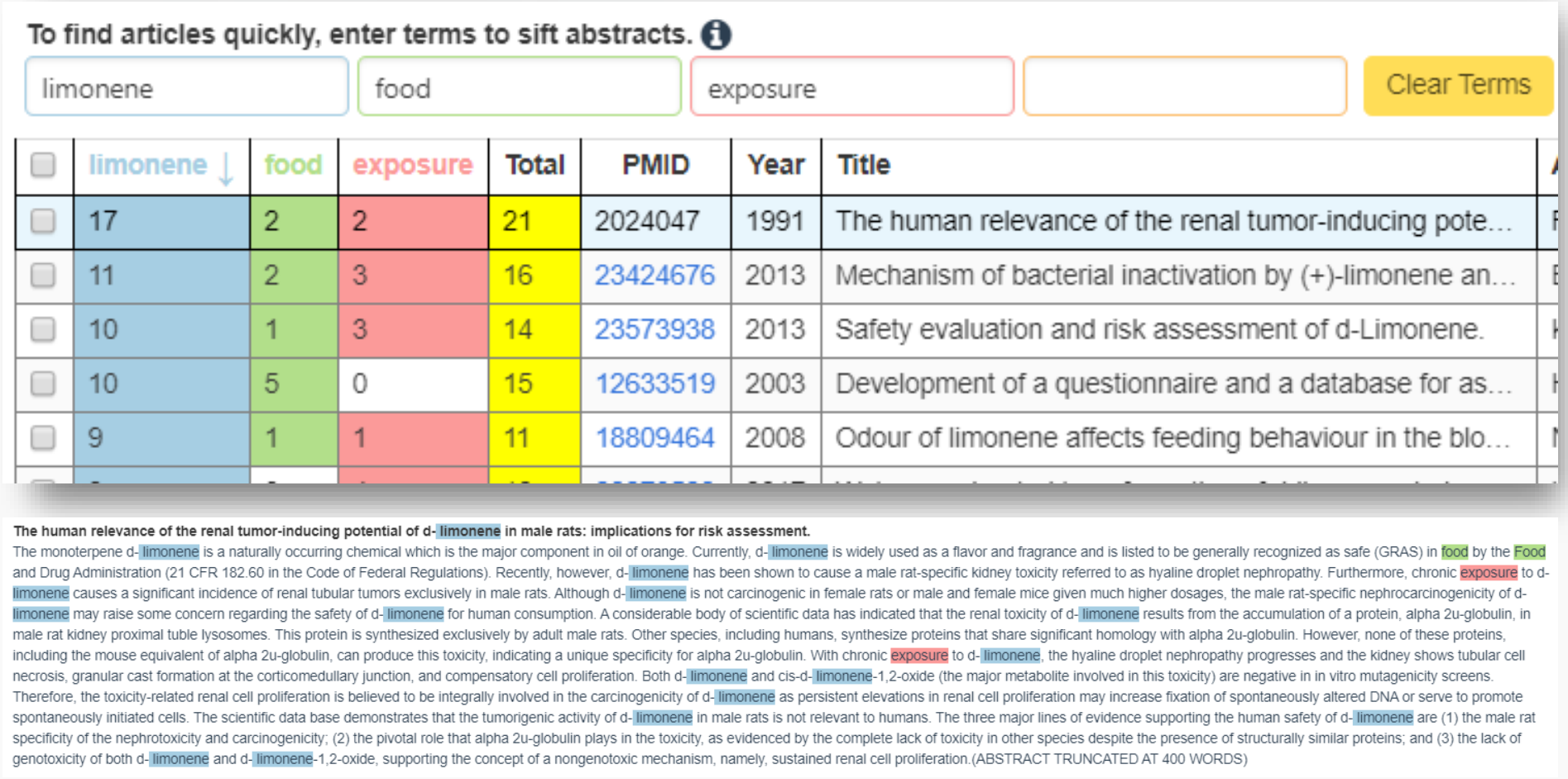
- Select a Query Term
- Hazard
- Fate and Transport
- Metabolism/PK/PD
- Chemical Properties
- Exposure
- Mixtures
- Male Reproduction
- Androgen Disruption
- Female Reproduction
- GeneTox
- Cancer
- Clinical Trials
- Embryo and embryonic development
- Child (infant through adolescent)
- Dust and Exposure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

Optionally, edit the query before retrieving.

("138-86-3" OR "Limonene") AND (food AND (exposure OR near-field OR far-field OR nhanes OR Environmental Monitoring OR Environmental Exposure OR exposome))

70 of 70 articles loaded...

Literature Searching



To find articles quickly, enter terms to sift abstracts.

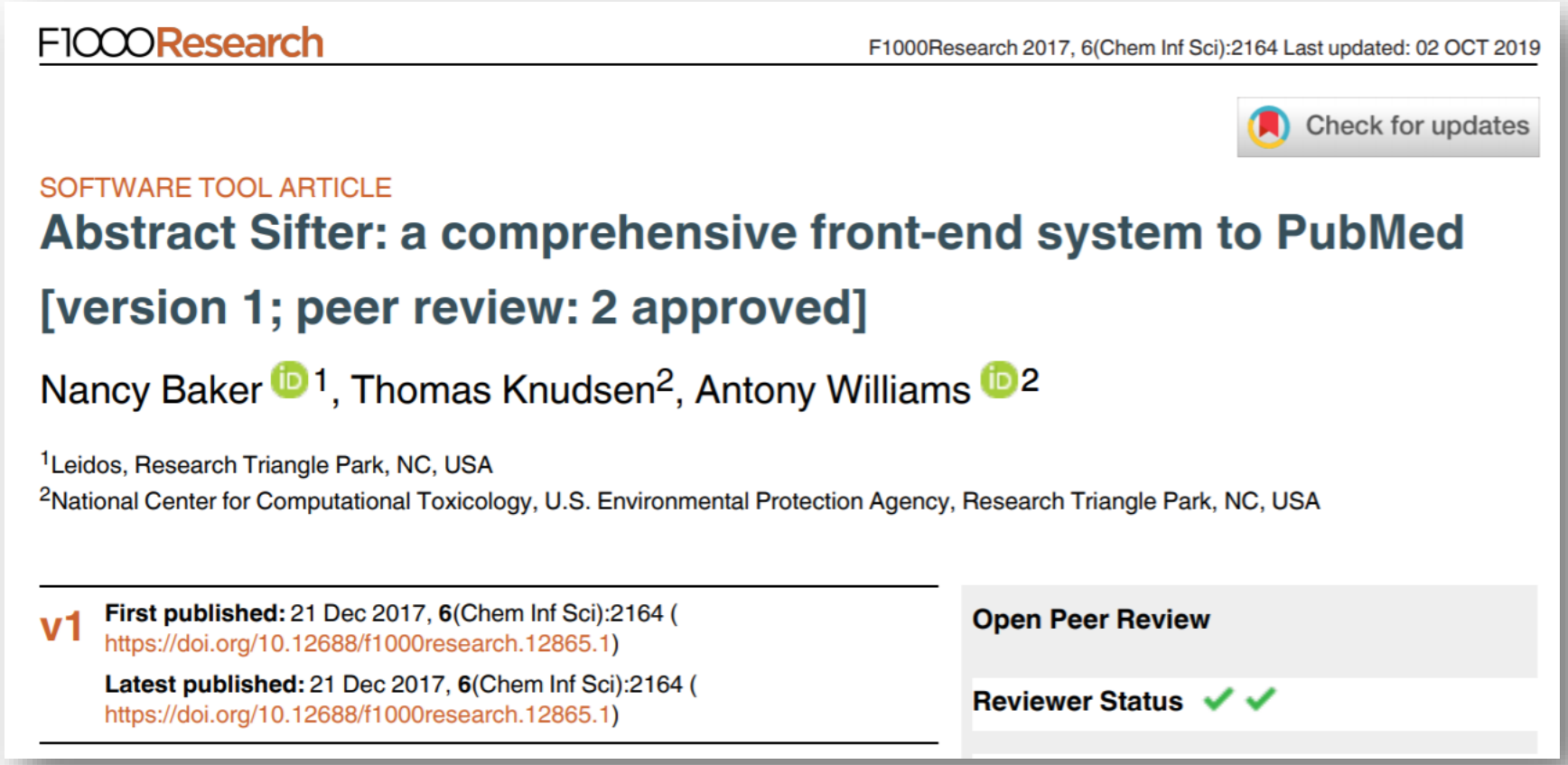
limonene | food | exposure | Clear Terms

limonene	food	exposure	Total	PMID	Year	Title
17	2	2	21	2024047	1991	The human relevance of the renal tumor-inducing pote…
11	2	3	16	23424676	2013	Mechanism of bacterial inactivation by (+)-limonene an…
10	1	3	14	23573938	2013	Safety evaluation and risk assessment of d-Limonene.
10	5	0	15	12633519	2003	Development of a questionnaire and a database for as…
9	1	1	11	18809464	2008	Odour of limonene affects feeding behaviour in the blo…

The human relevance of the renal tumor-inducing potential of d-limonene in male rats: implications for risk assessment.

The monoterpene d-limonene is a naturally occurring chemical which is the major component in oil of orange. Currently, d-limonene is widely used as a flavor and fragrance and is listed to be generally recognized as safe (GRAS) in food by the Food and Drug Administration (21 CFR 182.60 in the Code of Federal Regulations). Recently, however, d-limonene has been shown to cause a male rat-specific kidney toxicity referred to as hyaline droplet nephropathy. Furthermore, chronic exposure to d-limonene causes a significant incidence of renal tubular tumors exclusively in male rats. Although d-limonene is not carcinogenic in female rats or male and female mice given much higher dosages, the male rat-specific nephrocarcinogenicity of d-limonene may raise some concern regarding the safety of d-limonene for human consumption. A considerable body of scientific data has indicated that the renal toxicity of d-limonene results from the accumulation of a protein, alpha 2u-globulin, in male rat kidney proximal tuble lysosomes. This protein is synthesized exclusively by adult male rats. Other species, including humans, synthesize proteins that share significant homology with alpha 2u-globulin. However, none of these proteins, including the mouse equivalent of alpha 2u-globulin, can produce this toxicity, indicating a unique specificity for alpha 2u-globulin. With chronic exposure to d-limonene, the hyaline droplet nephropathy progresses and the kidney shows tubular cell necrosis, granular cast formation at the corticomedullary junction, and compensatory cell proliferation. Both d-limonene and cis-d-limonene-1,2-oxide (the major metabolite involved in this toxicity) are negative in in vitro mutagenicity screens. Therefore, the toxicity-related renal cell proliferation is believed to be integrally involved in the carcinogenicity of d-limonene as persistent elevations in renal cell proliferation may increase fixation of spontaneously altered DNA or serve to promote spontaneously initiated cells. The scientific data base demonstrates that the tumorigenic activity of d-limonene in male rats is not relevant to humans. The three major lines of evidence supporting the human safety of d-limonene are (1) the male rat specificity of the nephrotoxicity and carcinogenicity; (2) the pivotal role that alpha 2u-globulin plays in the toxicity, as evidenced by the complete lack of toxicity in other species despite the presence of structurally similar proteins; and (3) the lack of genotoxicity of both d-limonene and d-limonene-1,2-oxide, supporting the concept of a nongenotoxic mechanism, namely, sustained renal cell proliferation.(ABSTRACT TRUNCATED AT 400 WORDS)

Abstract Sifter for Excel



F1000Research F1000Research 2017, 6(Chem Inf Sci):2164 Last updated: 02 OCT 2019

Check for updates

SOFTWARE TOOL ARTICLE

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

Nancy Baker [1], Thomas Knudsen[2], Antony Williams [2]

[1]Leidos, Research Triangle Park, NC, USA
[2]National Center for Computational Toxicology, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

v1 **First published:** 21 Dec 2017, **6**(Chem Inf Sci):2164 (https://doi.org/10.12688/f1000research.12865.1)
Latest published: 21 Dec 2017, **6**(Chem Inf Sci):2164 (https://doi.org/10.12688/f1000research.12865.1)

Open Peer Review

Reviewer Status ✓ ✓

EPA United States Environmental Protection Agency

Work in Progress

List Registration Activities

- Registering and curating numerous lists
  - NIST library of chemicals –clean up especially around stereochemical representation
  - United States Geological Survey chemicals in water
  - Scientific Working Group for the Analysis of Seized Drugs
  - Synthetic Cannabinoids
  - Blood Exposome Database

Blood Exposome Curation

- Blood exposome data collection from Barupal and Fiehn. Great work and we reviewing.

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

- Aggregating large datasets is CHALLENGING
- Comparing with our “Abstract Sifter” approach
- We will iterate into a dashboard form..

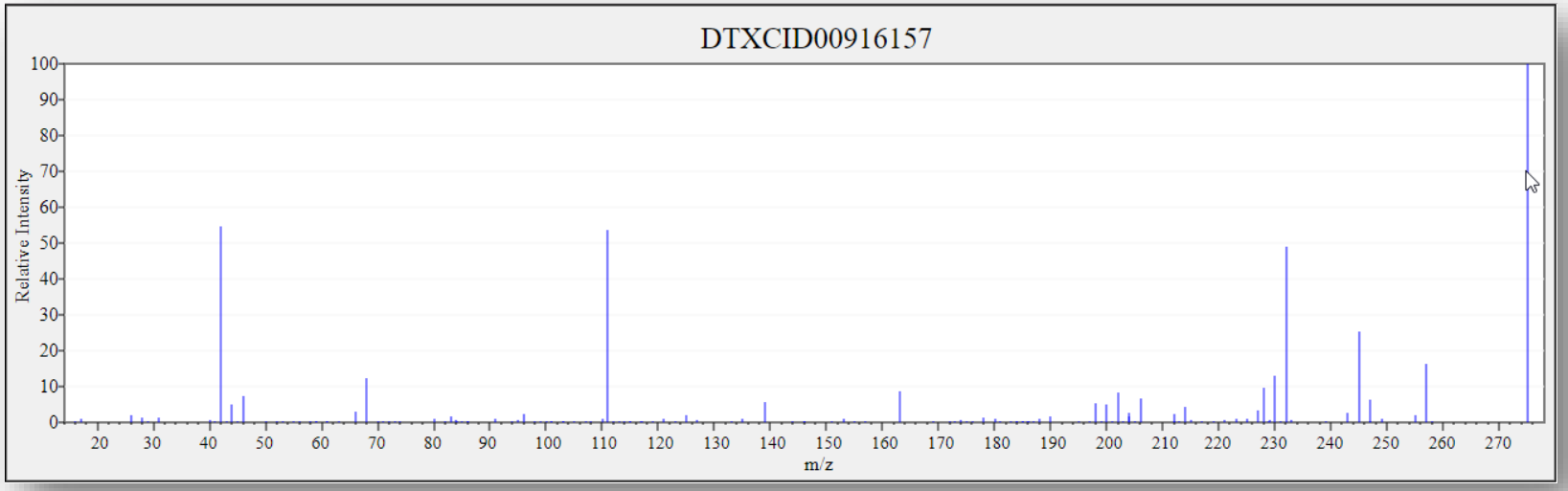
Prototype Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Structure/substructure/similarity search
- Access to API and web services

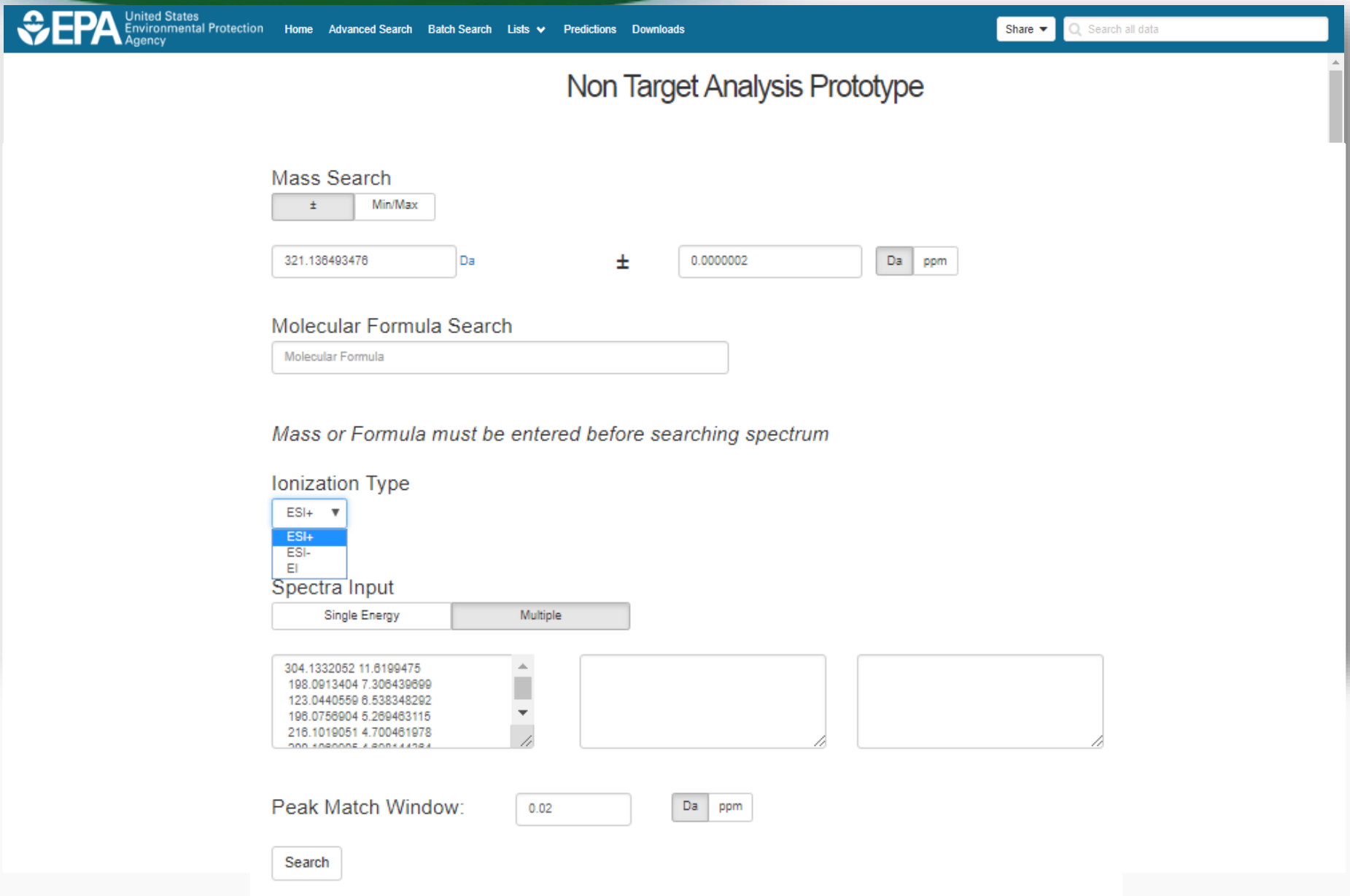
Predicted Mass Spectra

http://cfmid.wishartlab.com/

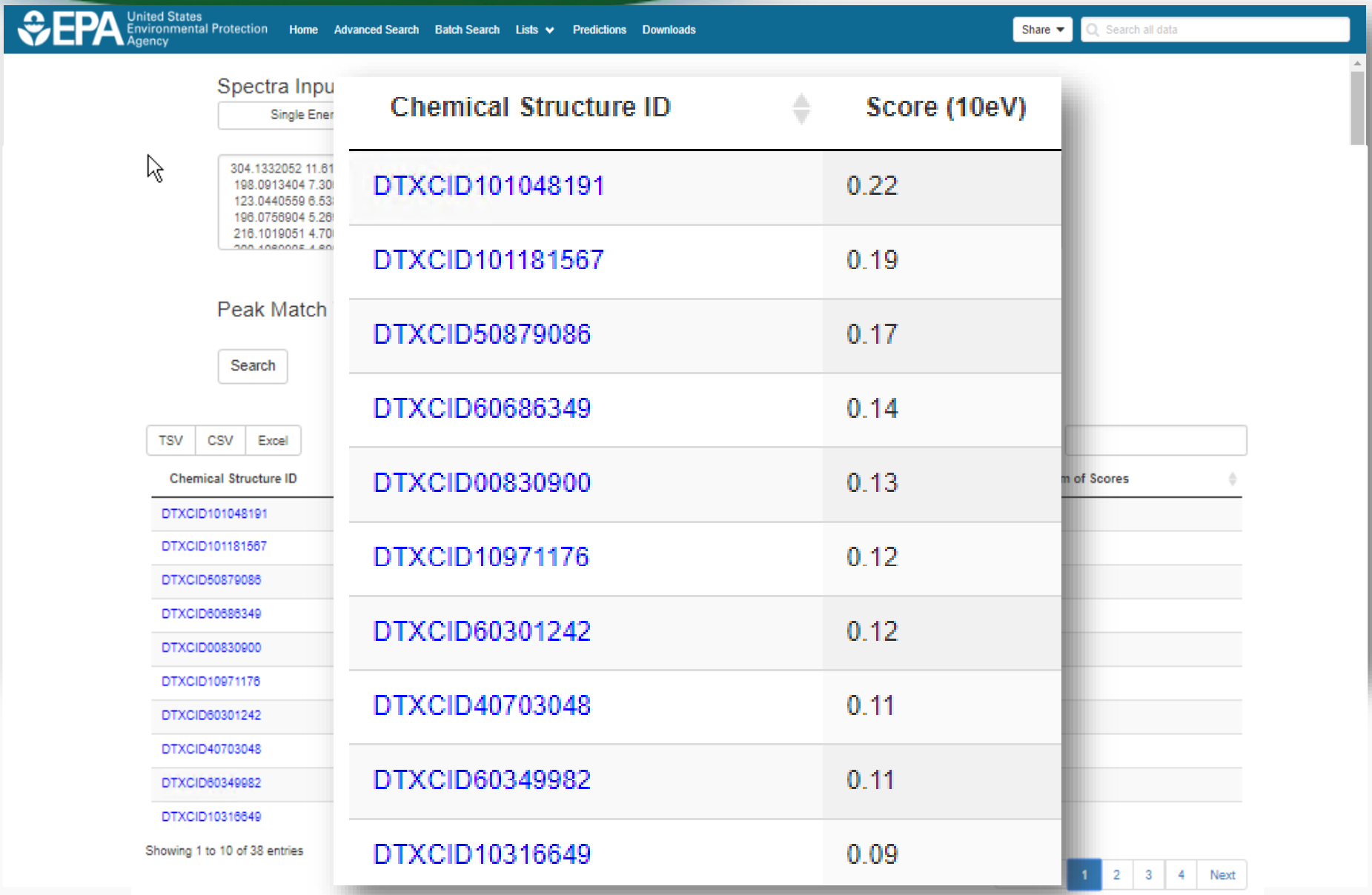
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra

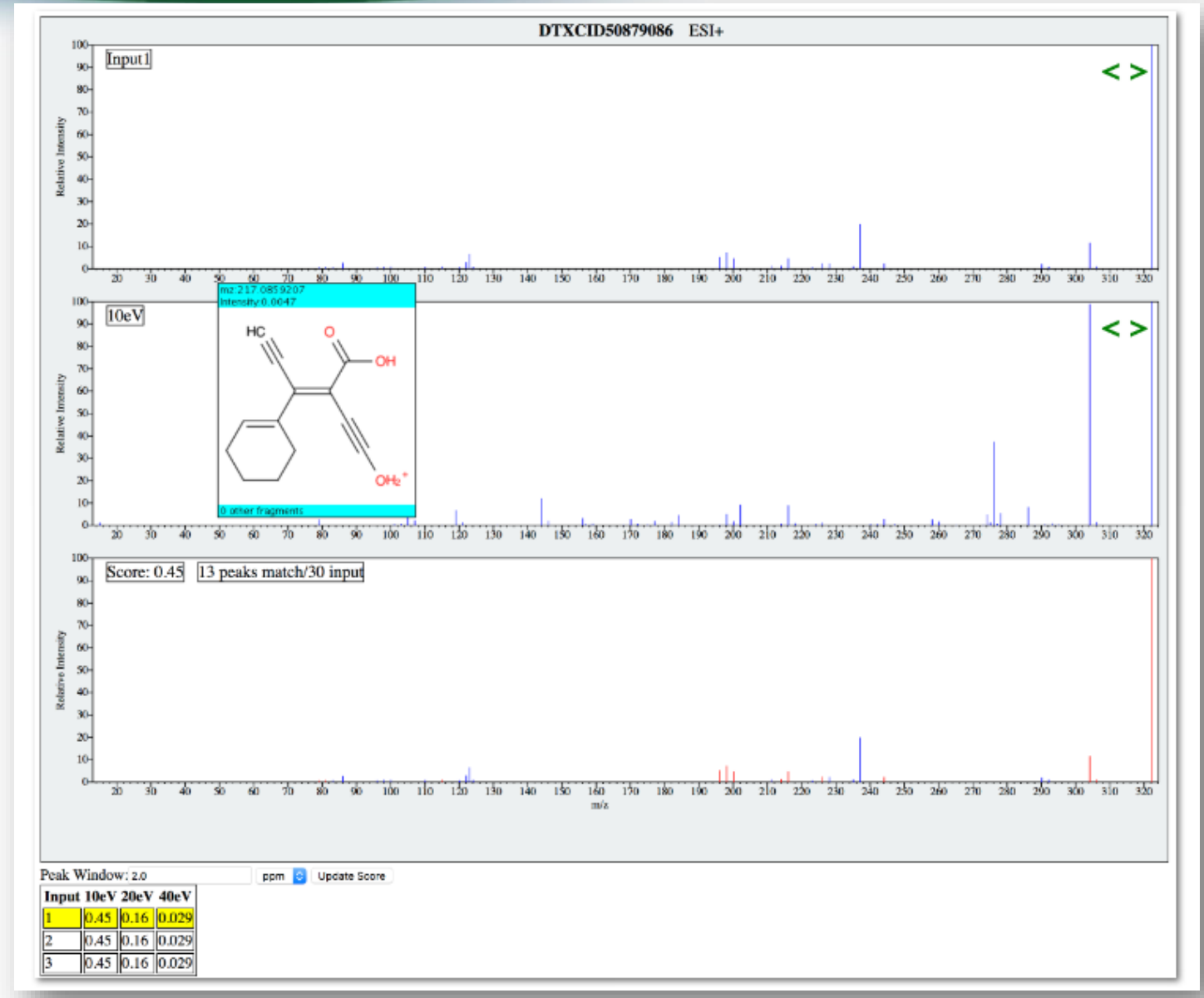


Search Expt. vs. Predicted Spectra



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



Predicted Data Already Public
Publication and Data Files

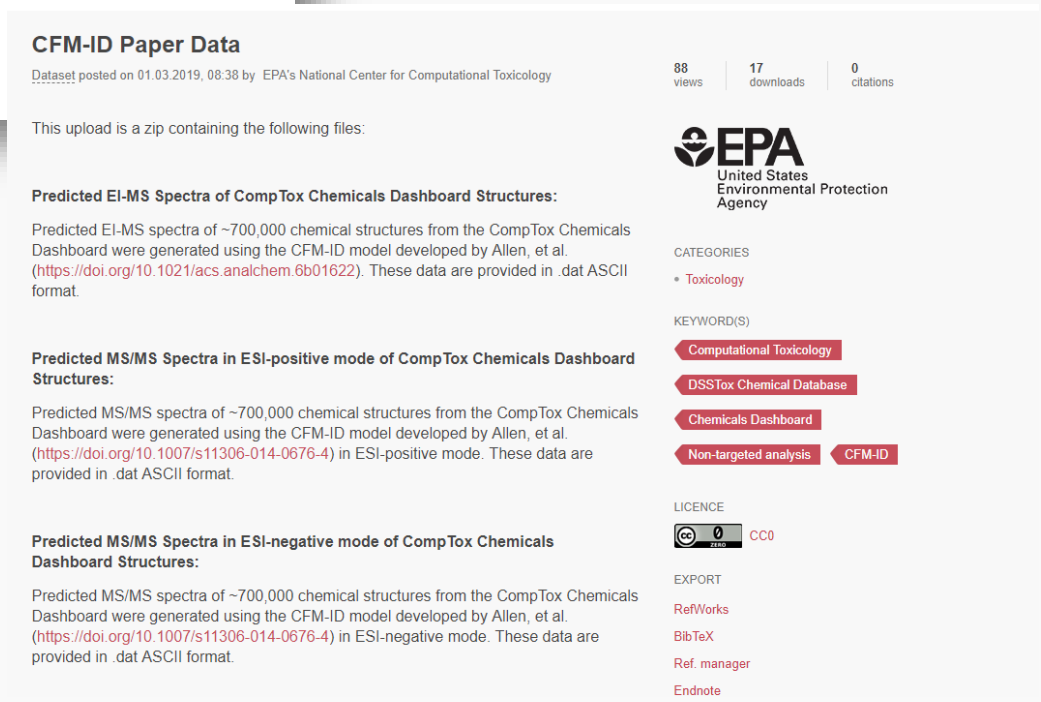
EPA United States Environmental Protection Agency

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.

EPA United States Environmental Protection Agency

CATEGORIES
- Toxicology

KEYWORD(S)
Computational Toxicology | DSSTox Chemical Database | Chemicals Dashboard | Non-targeted analysis | CFM-ID

LICENCE
CC0

EXPORT
RefWorks
BibTeX
Ref. manager
Endnote

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

Published: Chao et al

Analytical and Bioanalytical Chemistry

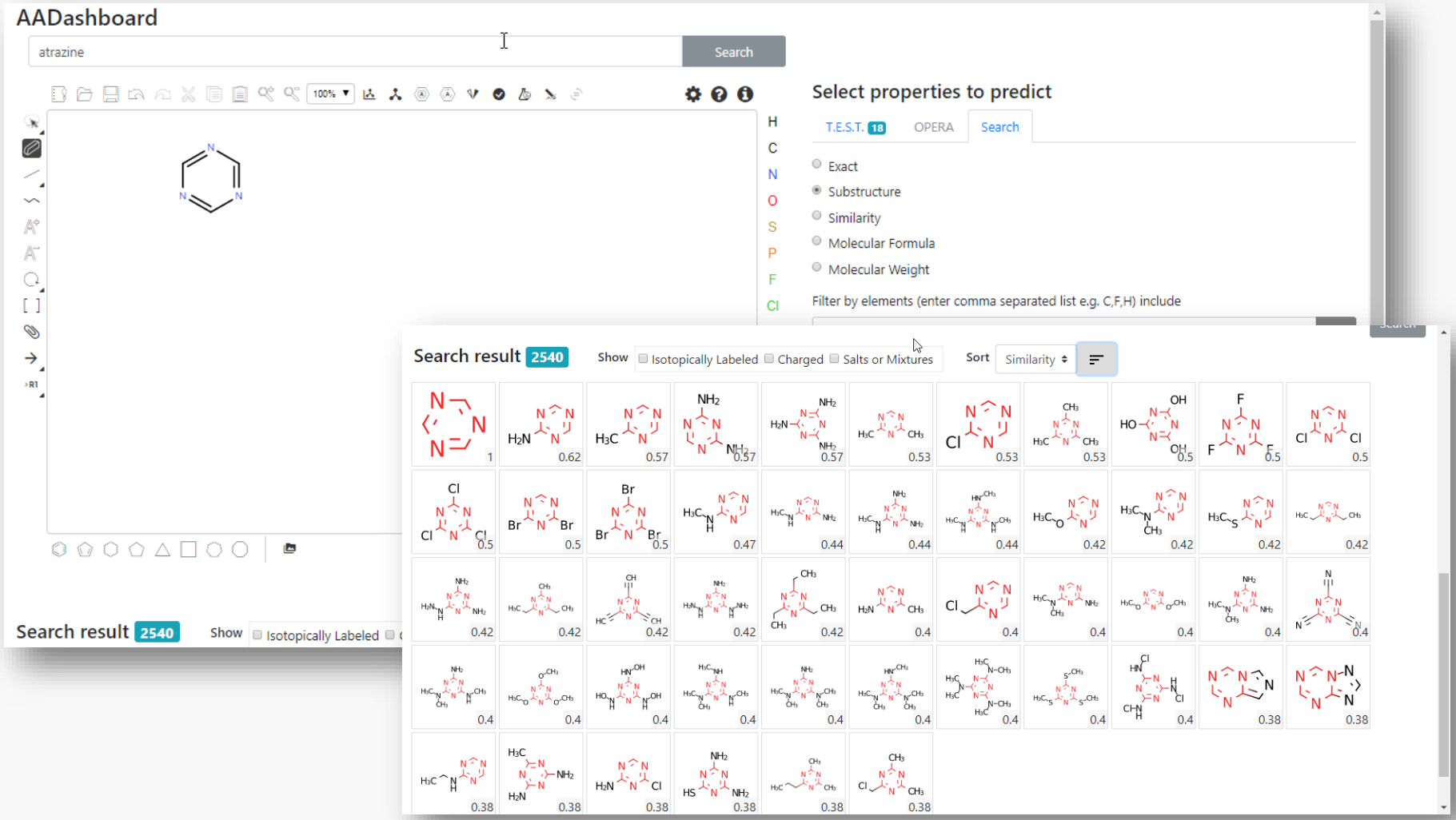
RESEARCH PAPER

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

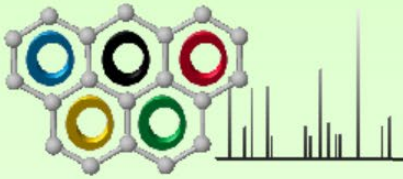
Alex Chao [1,2] · Hussein Al-Ghoul [1,2] · Andrew D. McEachran [1,3] · Ilya Balabin [4] · Tom Transue [4] · Tommy Cathey [4] · Jarod N. Grossman [2,3] · Randolph Singh [1,5] · Elin M. Ulrich [2] · Antony J. Williams [6] · Jon R. Sobus [2]

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

Prototype Development



CASMI 2012-2017 revisited



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.

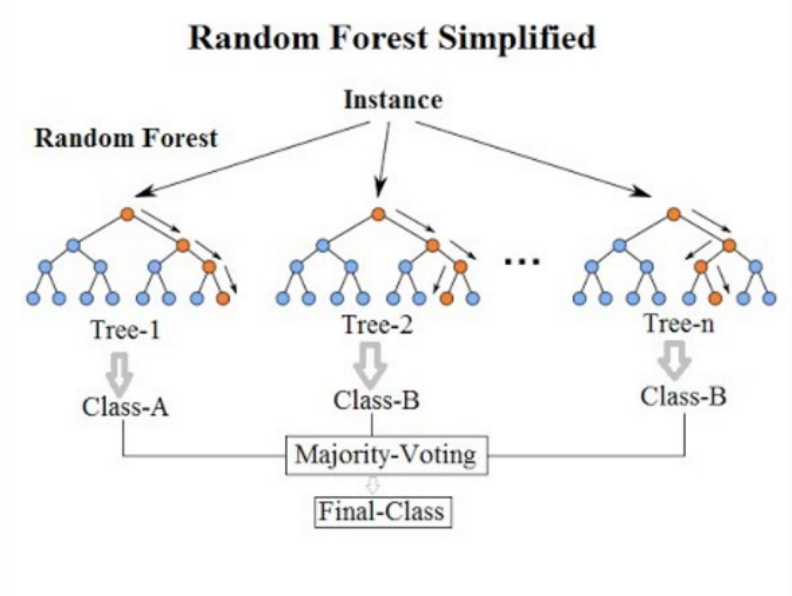
This year the contest will test the applicability of MS and MS/MS on natural products chemistry identifications. With 45 (Category 1) and up to 243 (Categories 2&3) natural products challenges - including a few tricky ones - there's something for everyone!

- Application of metadata candidate ranking and CFM-ID to all five years of CASMI data

Method Amenability Prediction
Charlie Lowe

Why?

- Chromatography-mass spectrometry can be LC or GC
- Which phase is more appropriate for which chemicals?



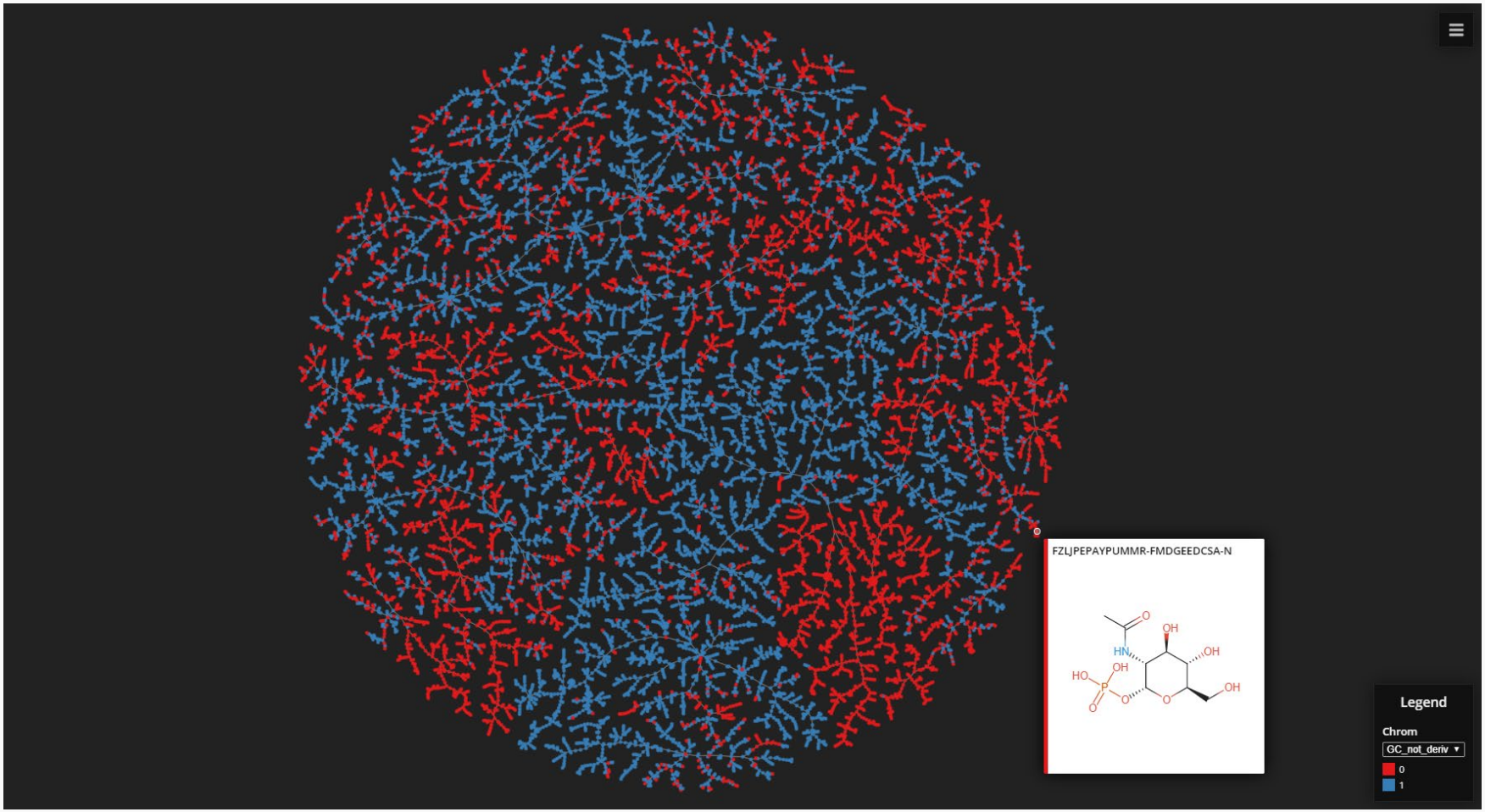
Ongoing Work

- Data sources to date
  - **Massbank of North America**
    - 9,275 chemicals for non-derivatized GC
    - 846 chemicals for derivatized GC
    - 816 chemicals for APCI+
    - 454 chemicals for APCI-
    - 4,907 chemicals for ESI+
    - 3,430 chemicals for ESI-

  - **EPA Non-targeted Analysis Collaborative Trial (ENTACT)**
    - 886 chemicals for non-derivatized GC
    - 44 chemicals for derivatized GC
    - 774 chemicals for APCI+
    - 431 chemicals for APCI-
    - 1,113 chemicals for ESI+
    - 648 chemicals for ESI-

TMAP Visualization of MoNA GC Data

EPA United States Environmental Protection Agency



Future Work: Add database of Collision Cross Section Prediction

EPA United States Environmental Protection Agency

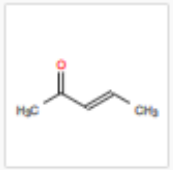
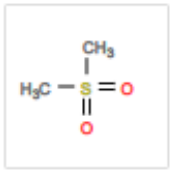
PNNL Collision Cross Section Database

Pacific Northwest NATIONAL LABORATORY

Search

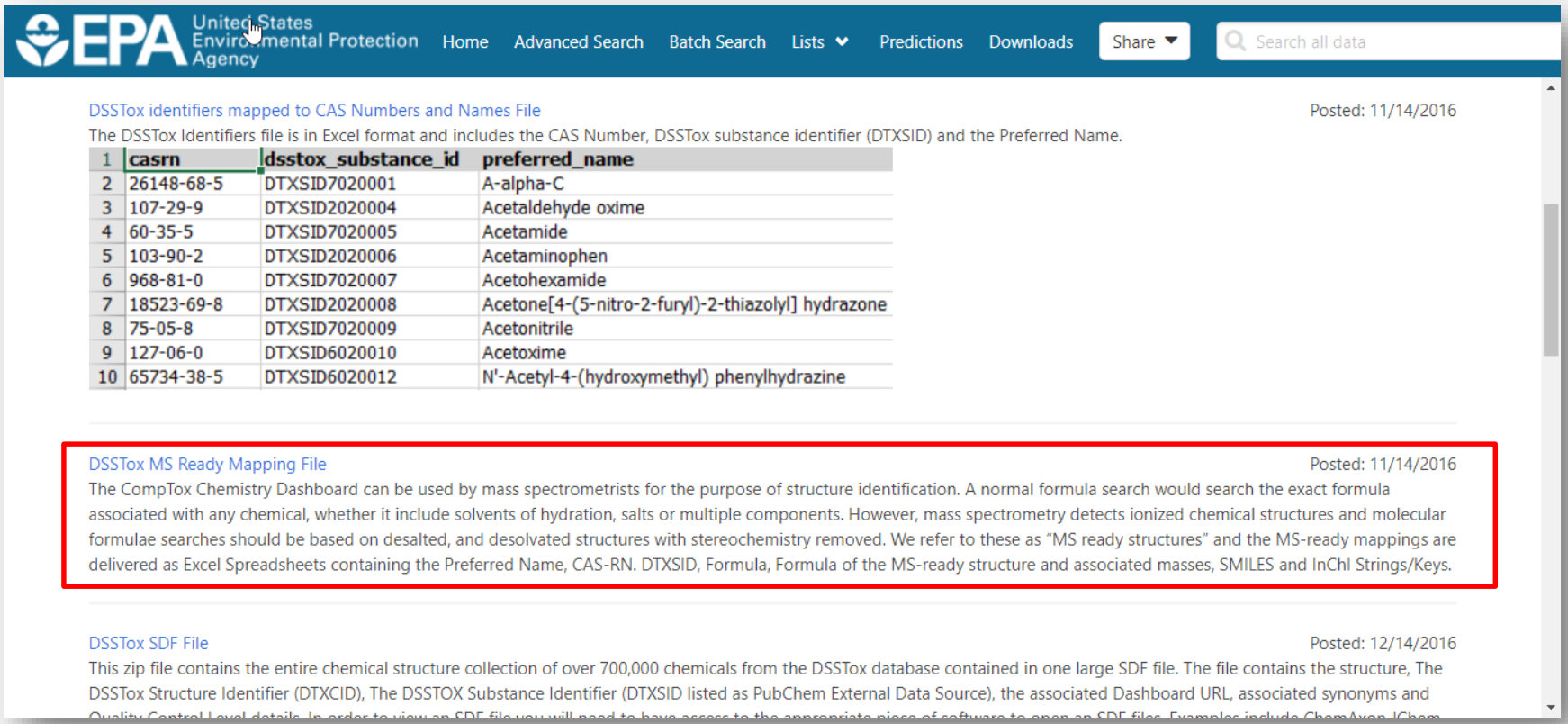
Showing 1 to 25 of 1000 rows 25 rows per page

‹ 1 2 3 4 5 ... 40 ›

Chemical	SMILES	InChI	Formula	Mass	CCS (Å^2)
 (3E)-pent-3-en-2-one	*	*	C_5H_8O	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
 Dimethyl sulfone	*	*	$C_2H_6O_2S$	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
 isothiocyanatocyclopropane	*	*	C_4H_5NS	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

API services and Open Data

- Web Services https://actorws.epa.gov/actorws/
- Data sets also available for download..



	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

Web Services
https://actorws.epa.gov/actorws/

- Data in UI, JSON and XML format
- Our services are free of course..

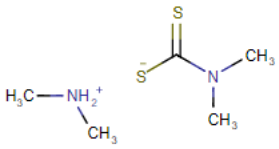
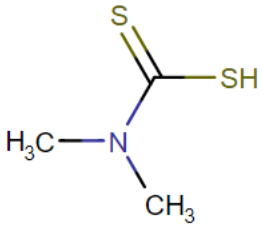
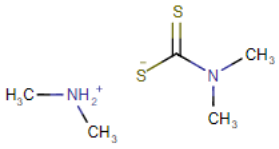
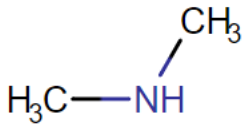
https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=80-05-7
https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=80-05-7
https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=80-05-7

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

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

InChIKey to DTXCIDs

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

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

Data and Services used by the Community

NORMAN Suspect List Exchange

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

EPA United States Environmental Protection Agency

NORMAN

Network of reference laboratories, research centres and related organisations for monitoring of emerging environmental substances

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 *et al.* (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

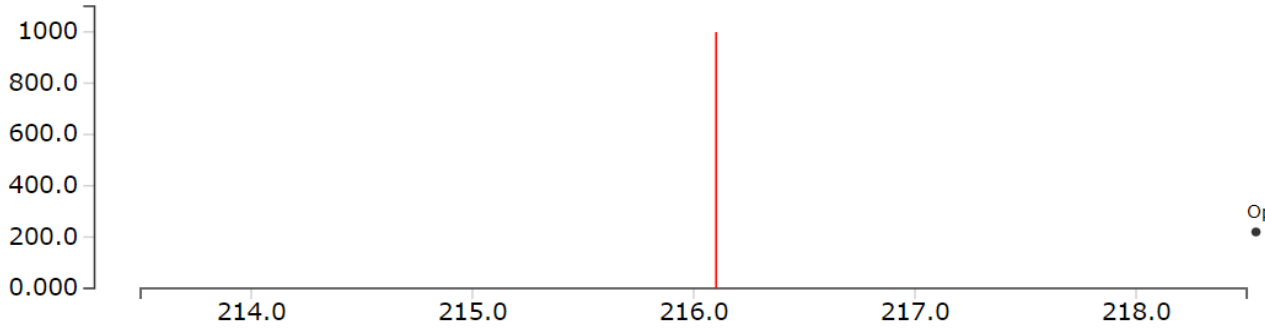
MassBank mapping to Dashboard Based on Web Service lookup

MassBank Record: EA028808

Home | Search | Record Index | Data Privacy | Imprint | MassBank ID: Go

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

Mass Spectrum



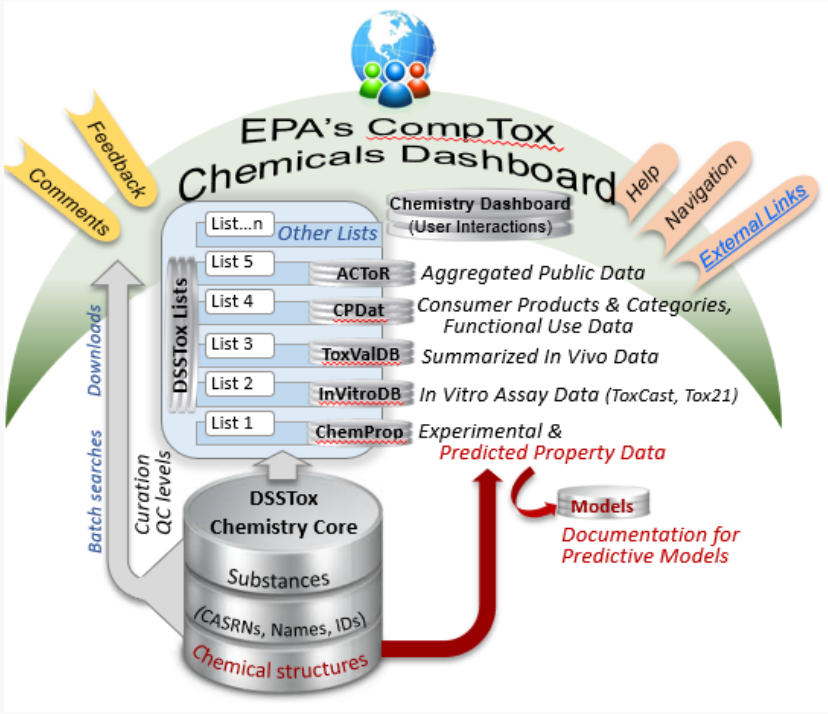
Options
• Labels

Chemical Structure

CH$NAME: Atrazine
CH$NAME: 6-chloro-N-ethyl-N'-isopropyl-1,3,5-triazine-2,4-diamine
CH$NAME: 6-chloranyl-N4-ethyl-N2-propan-2-yl-1,3,5-triazine-2,4-diamine
CH$COMPOUND_CLASS: N/A; Environmental Standard
CH$FORMULA: C8H14ClN5
CH$EXACT_MASS: 215.0932
CH$SMILES: c1(nc(nc(n1)Cl)NCC)NC(C)C
CH$IUPAC: InChI=1S/C8H14ClN5/c1-4-10-7-12-6(9)13-8(14-7)11-5(2)3/h5H,4H2,1-3H3,(H2,10,11,12,13,14)
CH$LINK: CAS 1912-24-9
CH$LINK: CHEBI 15930
CH$LINK: KEGG C06551
CH$LINK: PUBCHEM CID:2256
CH$LINK: INCHIKEY MXWJVTOOROXGIU-UHFFFAOYSA-N
CH$LINK: CHEMSPIDER 2169
CH$LINK: COMPTOX DTXSID9020112

Conclusion

- Dashboard access to data for ~875,000 chemicals (~895k in the Spring Release)
- MS-Ready data facilitates structure identification
- 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

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

TEAMS
IT Development Team
Curation Team

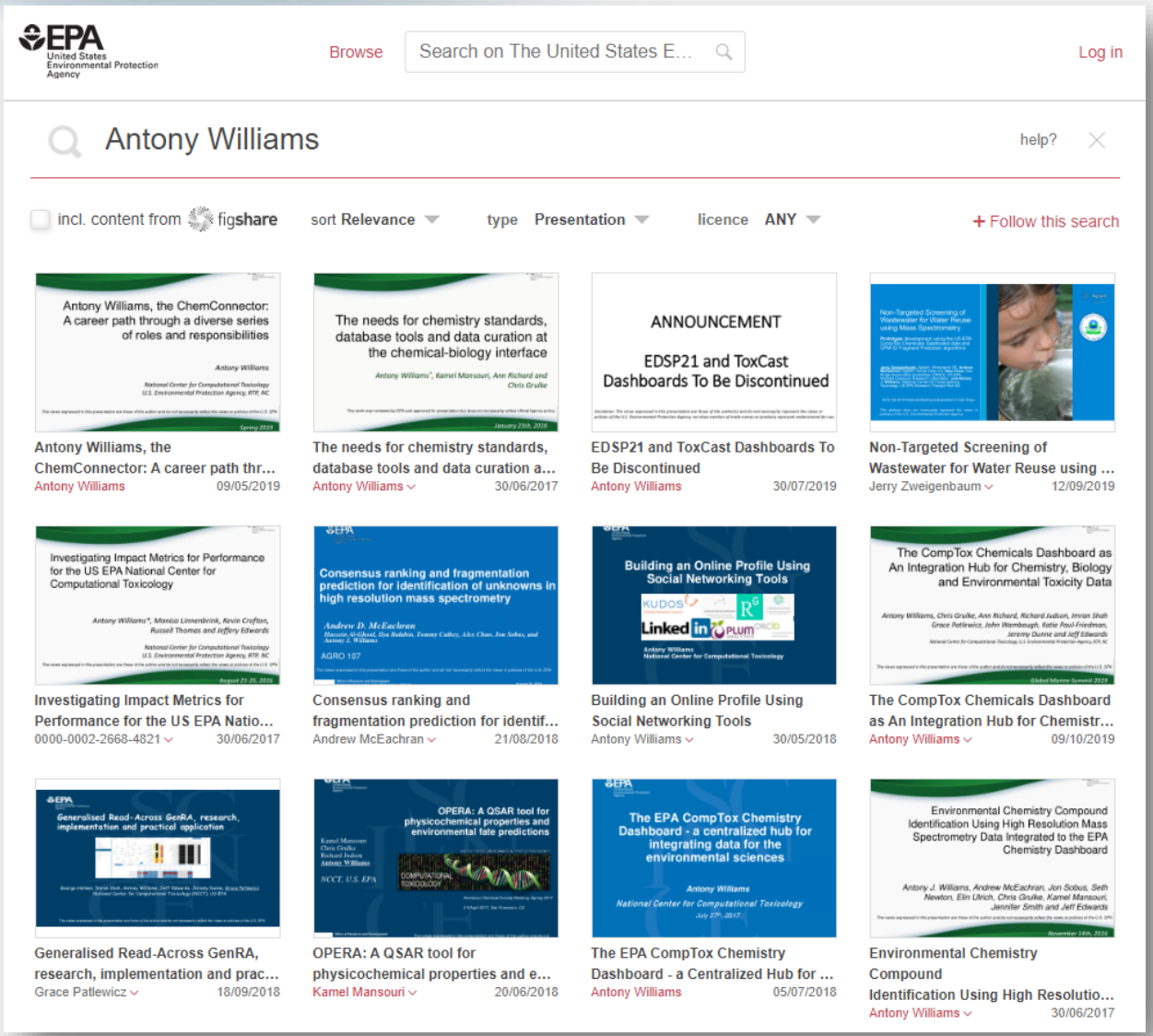
ILS
Kamel Mansouri

GDIT
Ilya Balabin
Tom Transue
Tommy Cathey

Collaborators
Emma Schymanski
NORMAN Network
Andrew McEachran

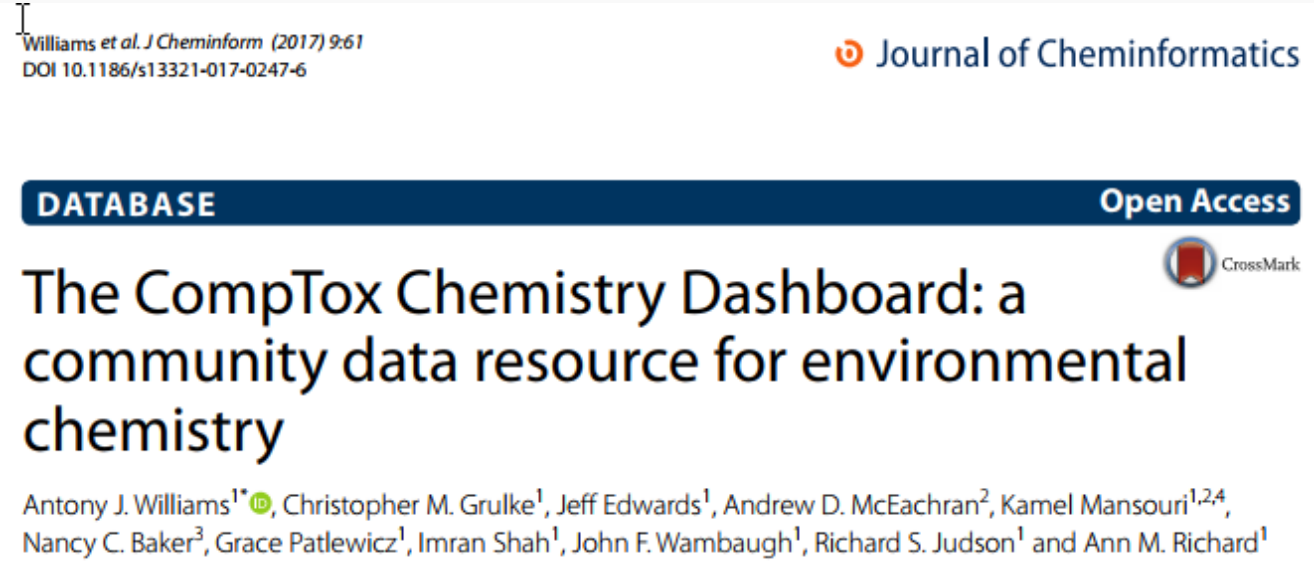
MANY presentations online

https://tinyurl.com/w5hqs55



Contact

Antony Williams
CCTE, 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. Grulke[1], Jeff Edwards[1], Andrew D. McEachran[2], Kamel Mansouri[1,2,4], Nancy C. Baker[3], Grace Patlewicz[1], Imran Shah[1], John F. Wambaugh[1], Richard S. Judson[1] and Ann M. Richard[1]

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