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

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

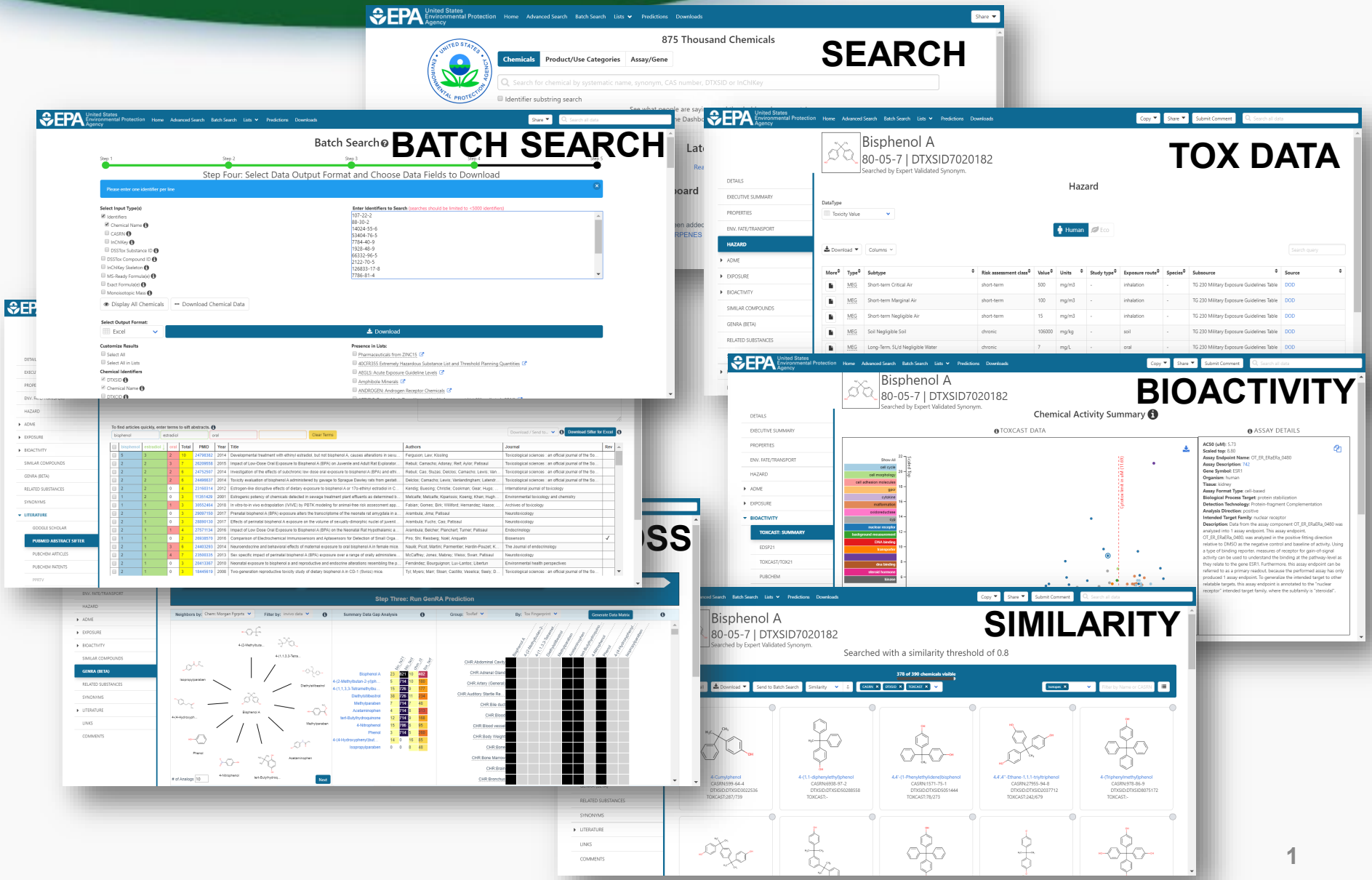
The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

ISES

September 2020

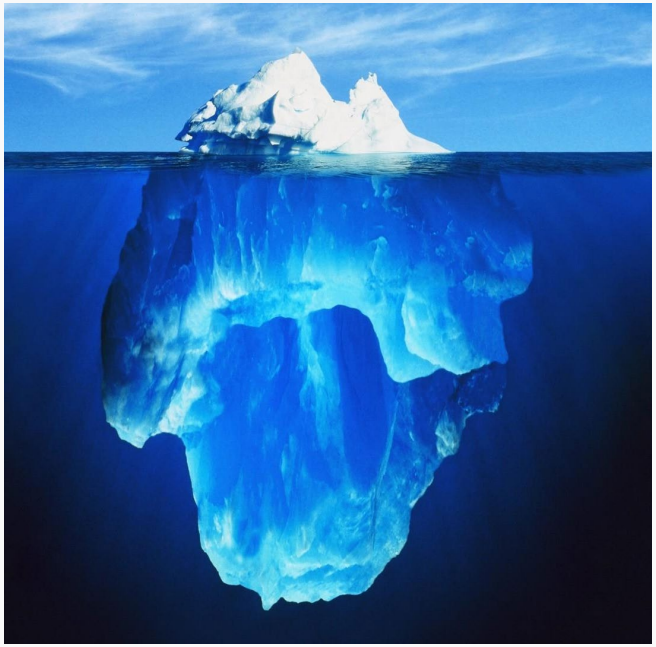
CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard



How many chemicals to consider?

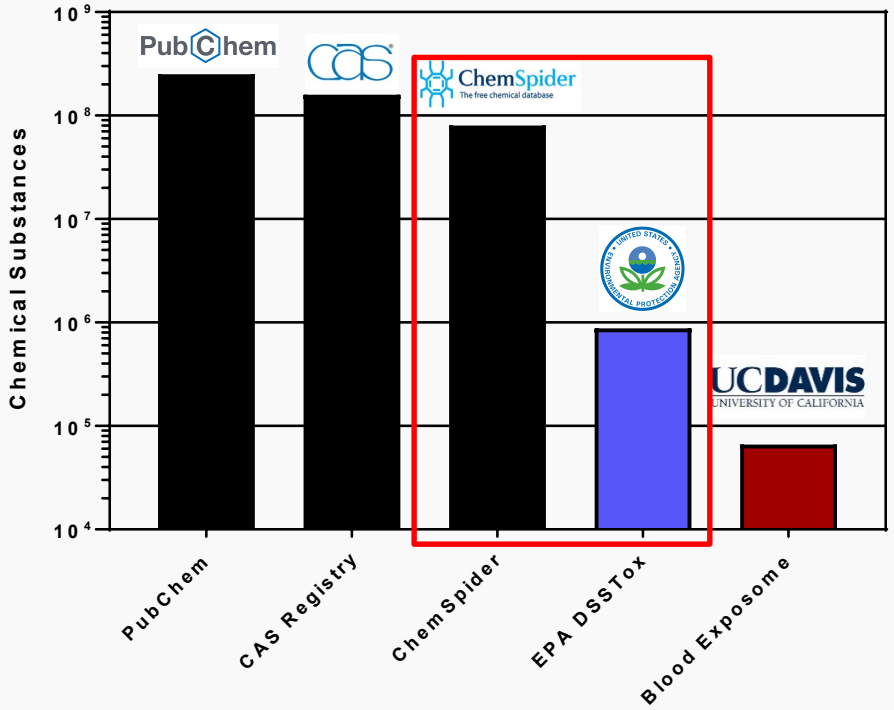
- Dashboard content is small but focused curated content



- **Dashboard** ~900,000
- ChemSpider ~90,000,000
- PubChem ~111,000,000
- CAS ~ 165,000,000

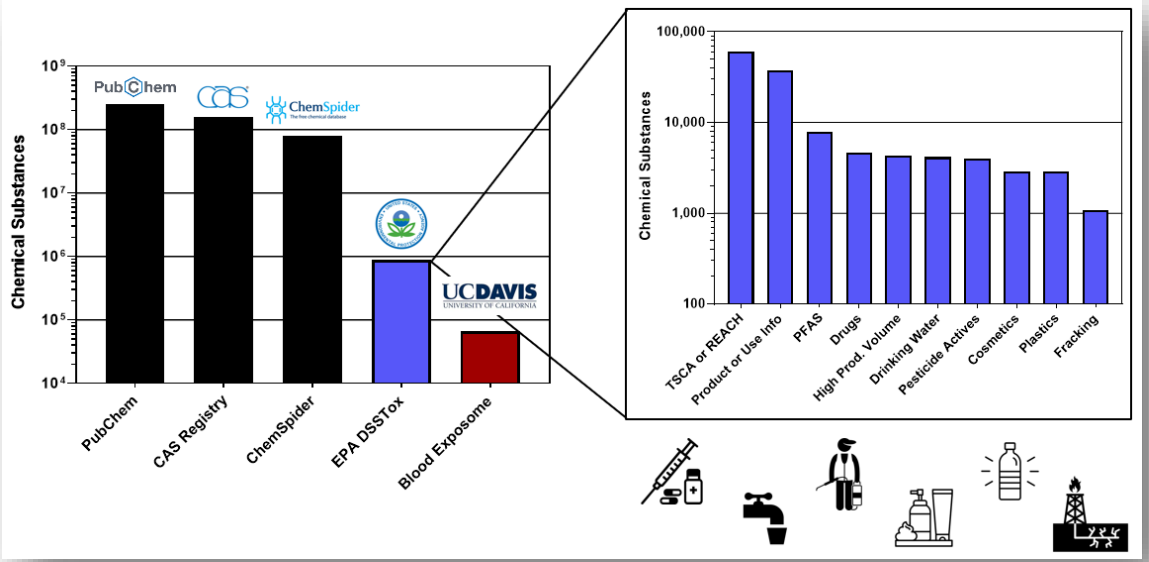
BIG databases are GREAT!

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



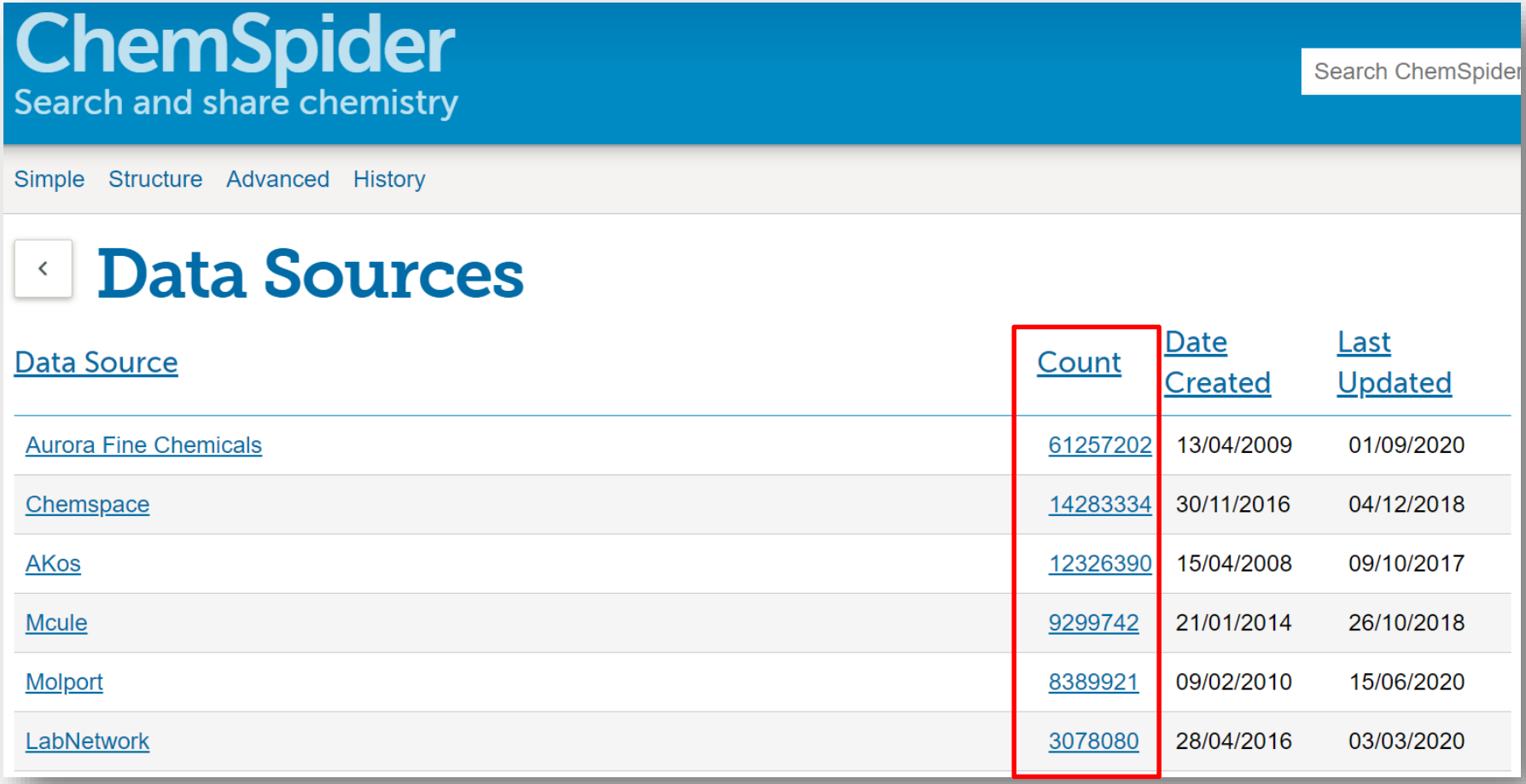
What chemicals constitute the Exposome?

- 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

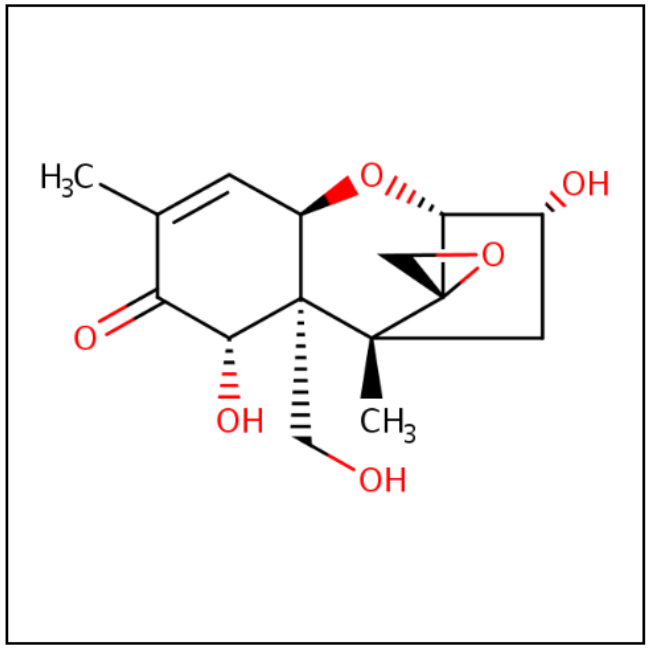
Search ChemSpider

Simple Structure Advanced History

Data Sources

Data Source	Count	Date Created	Last Updated
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: (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

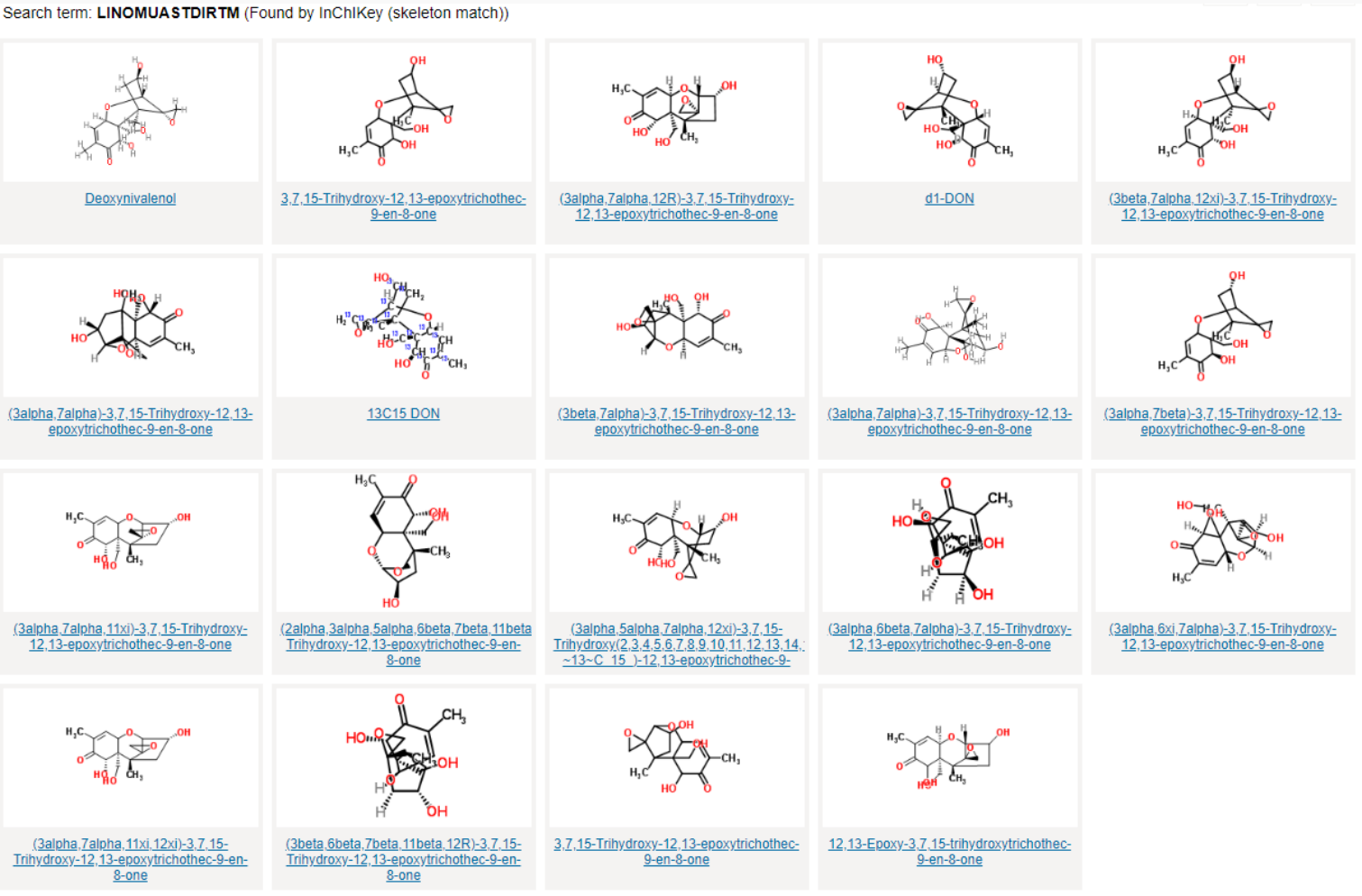
Search Google for: Structural Skeleton | Full Structure

Copy All

Vomitoxin - ChemSpider

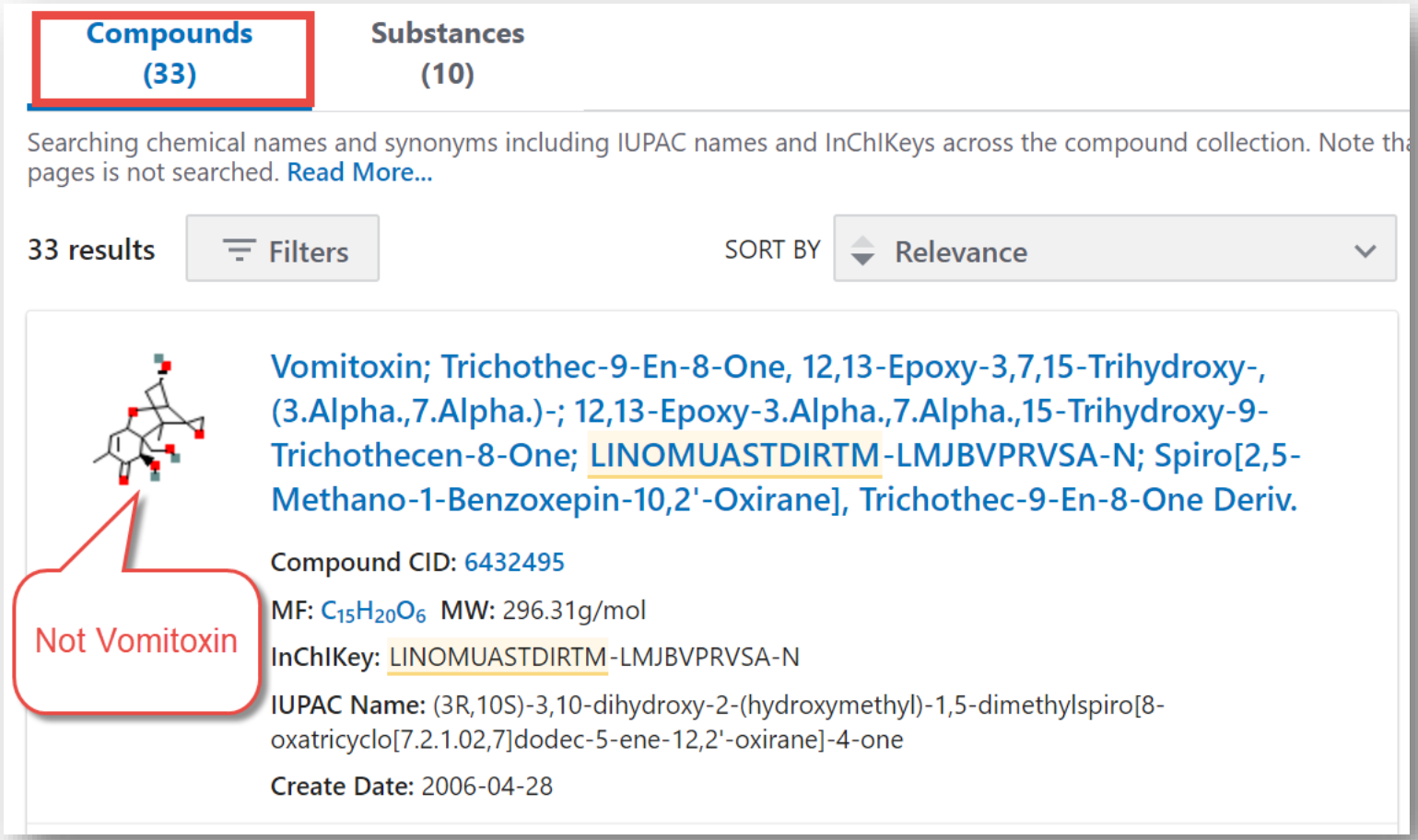
- 19 “Vomitoxins” – 3 isotopically labeled

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



Vomitoxin – PubChem

- 33 unique InChI Keys



PubChem – “virtual chemistry” Large Databases – collapse

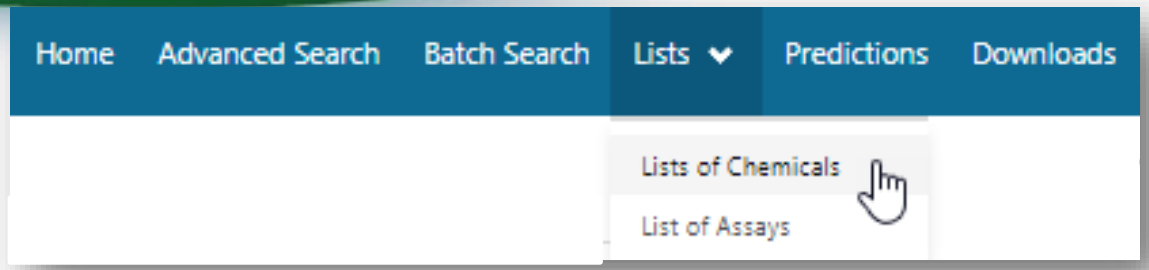
- Lots of “virtual chemistry” and “MODs”
- Vomitoxin has 7 ZINC stereoforms.

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

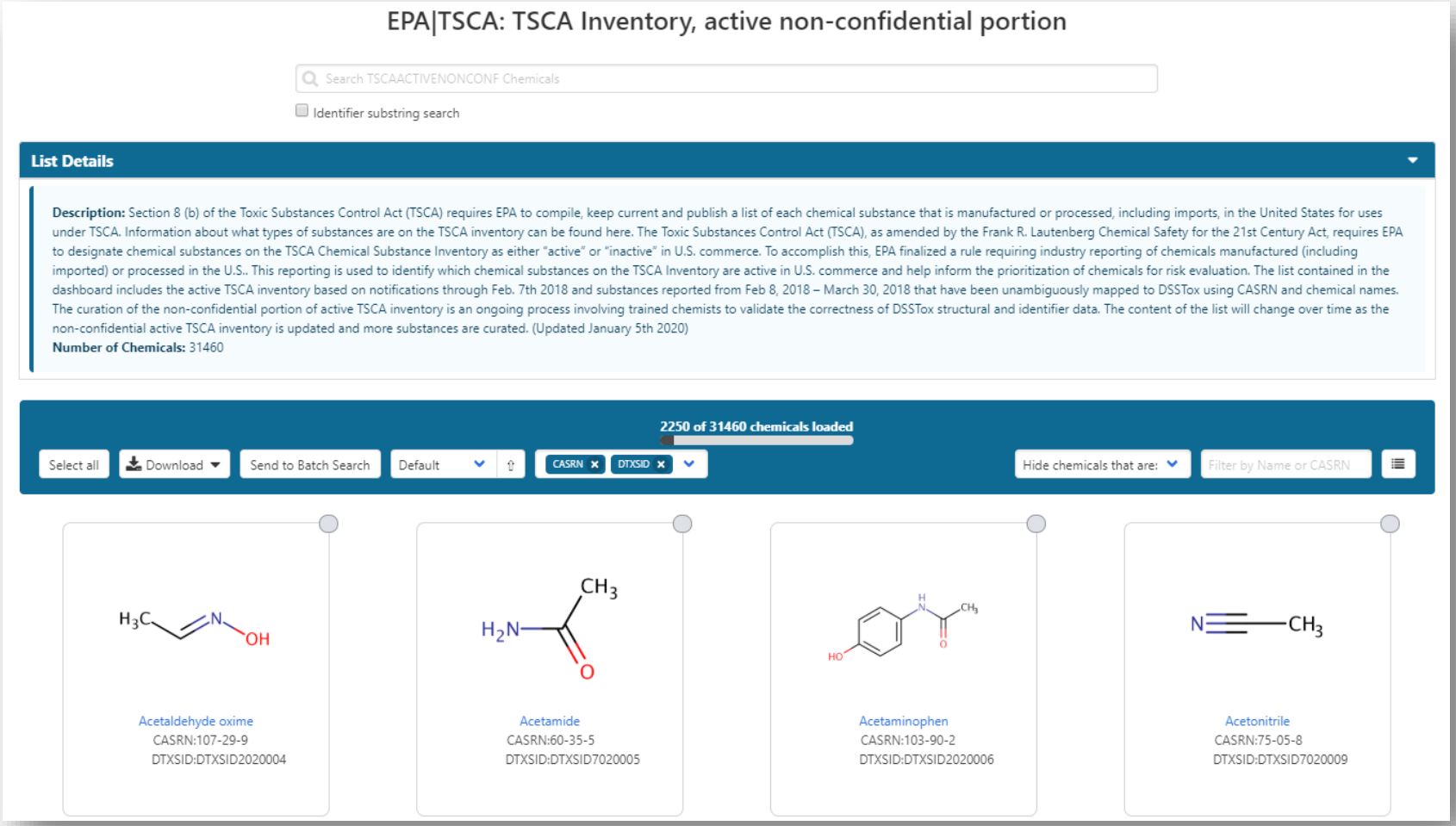
~275 Chemical Lists (and growing)



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)



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

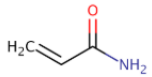
*Note that Appendix H chemical listings in Tables H-2 and H-4 were mapped to current DSSTox content, which has undergone additional curation since the publication of the original EPA HF Report (Dec 2016). In the few cases where a Chemical Name and CASRN from the original report map to distinct substances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

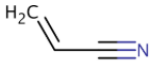
1640 chemicals

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

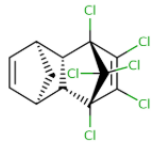
Acrolein
CASRN:107-02-8
DTXSID:DTXSID5020023



Acrylamide
CASRN:79-06-1
DTXSID:DTXSID5020027



Acrylonitrile
CASRN:107-13-1
DTXSID:DTXSID5020029



Aldrin
CASRN:309-00-2
DTXSID:DTXSID8020040

Tire Crumb Rubber (298)

EPA United States Environmental Protection Agency

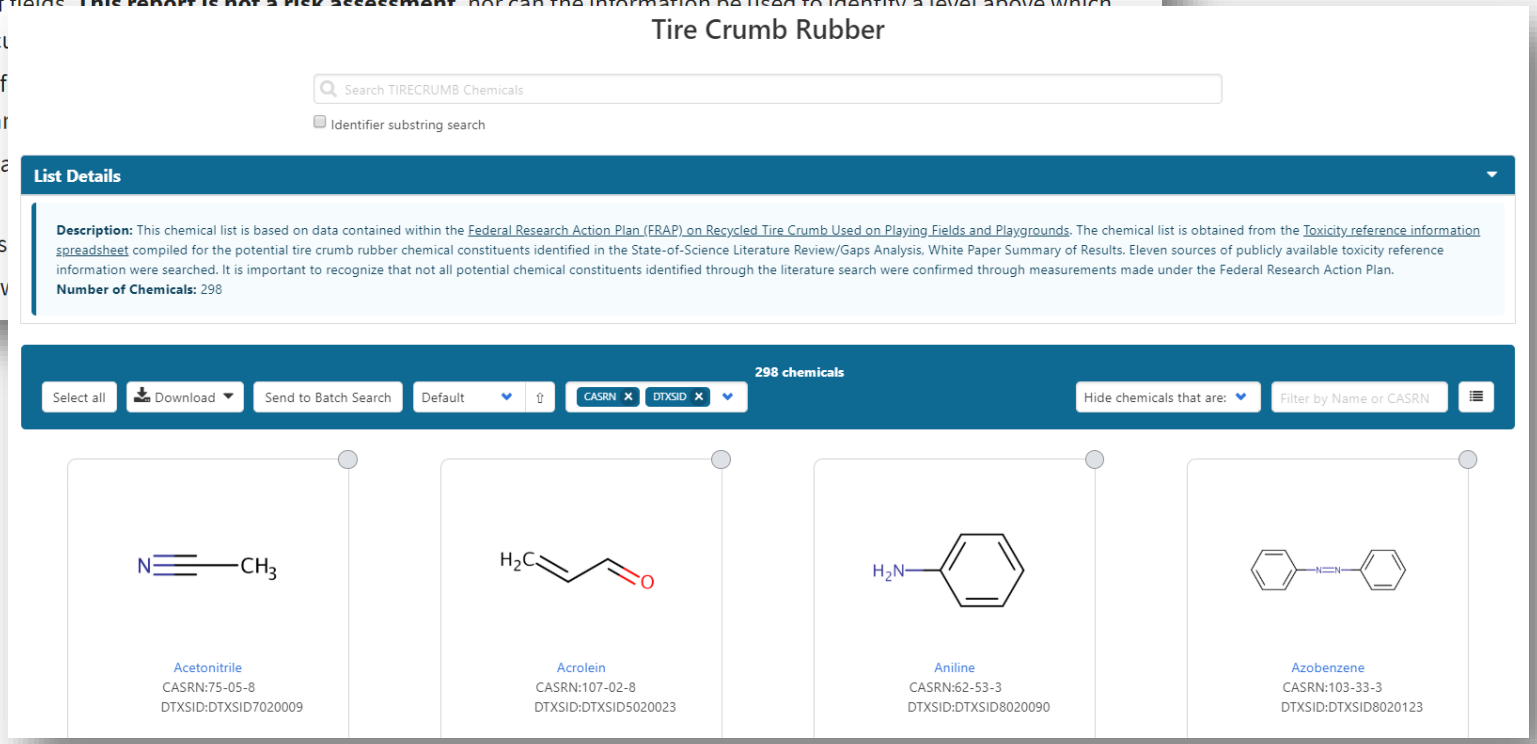
Related Topics: Safer Chemicals Research

CONTACT US SHARE

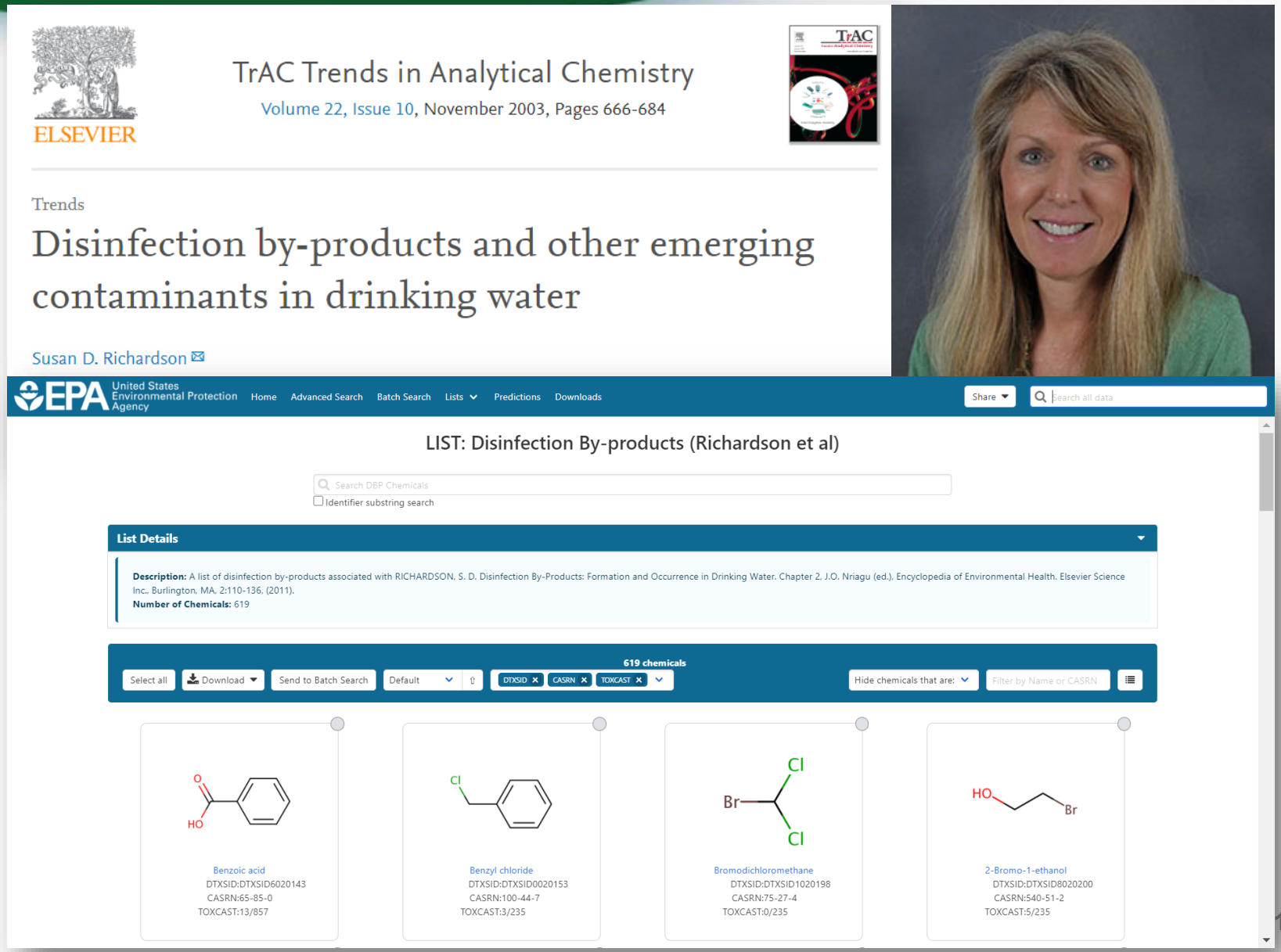
July 2019 Report: Tire Crumb Rubber Characterization

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



Disinfection By-Products



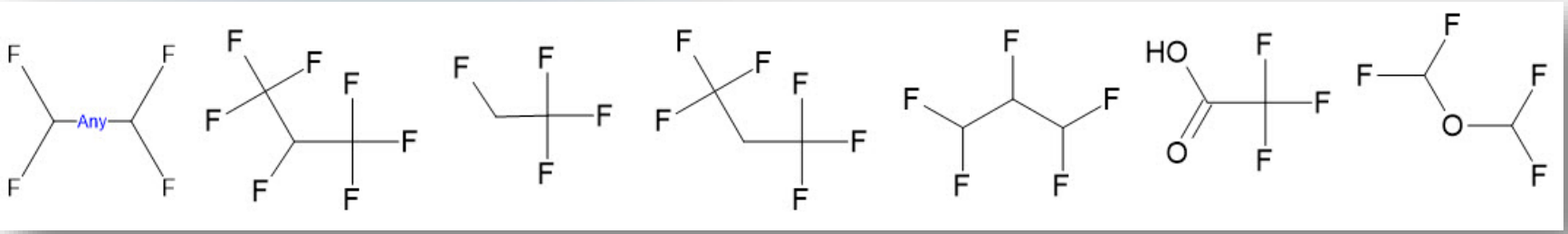
PFAS lists of Chemicals

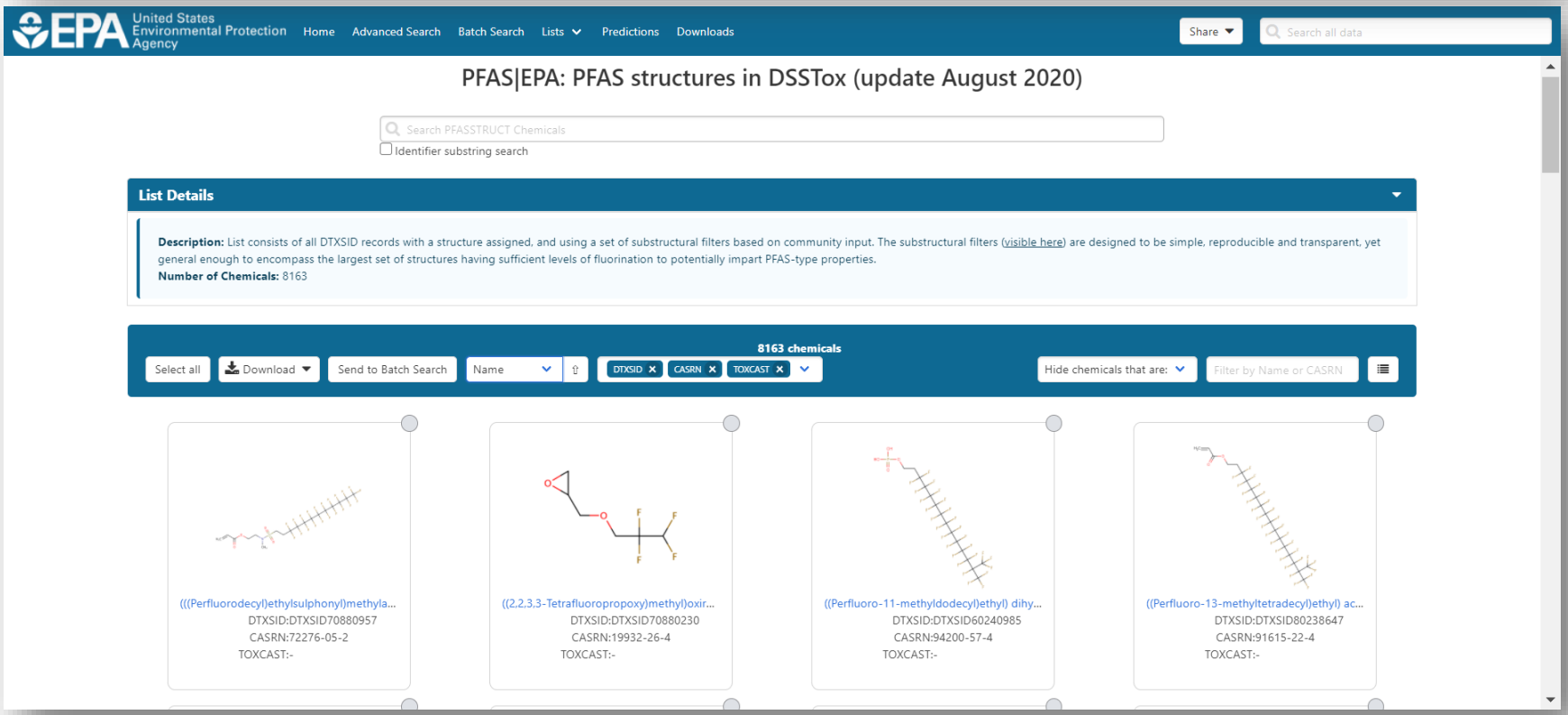
Select List

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)





Blood Exposome Curation in Process

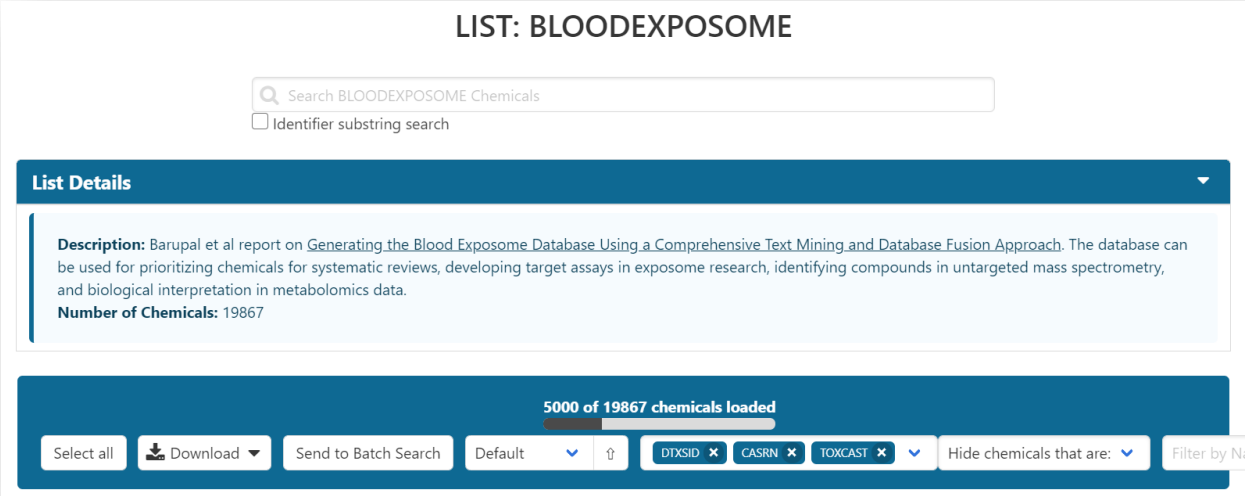
Vol. 127, No. 9 | Research

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.



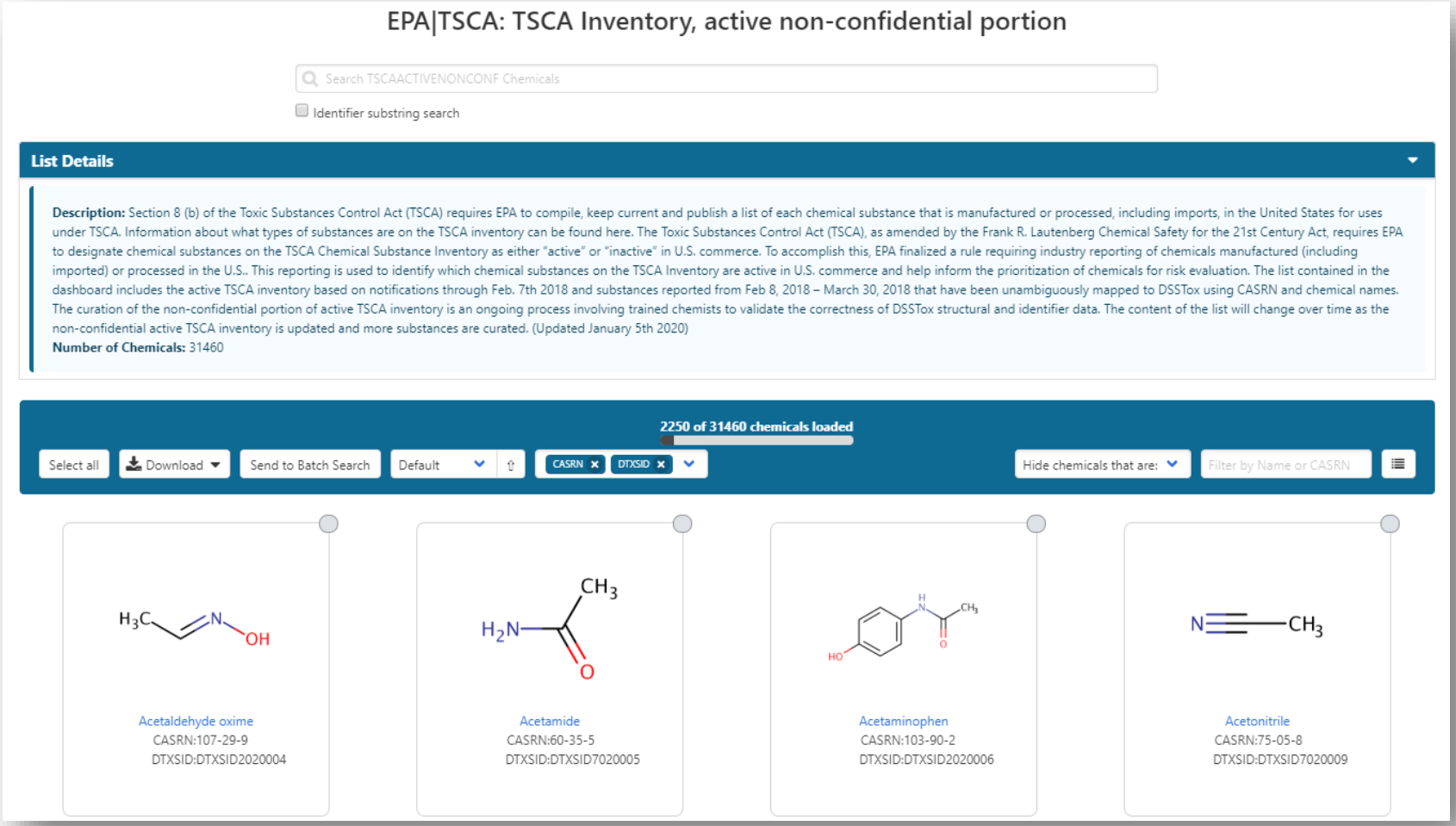
"UVCB" Chemicals

UVCB Chemicals

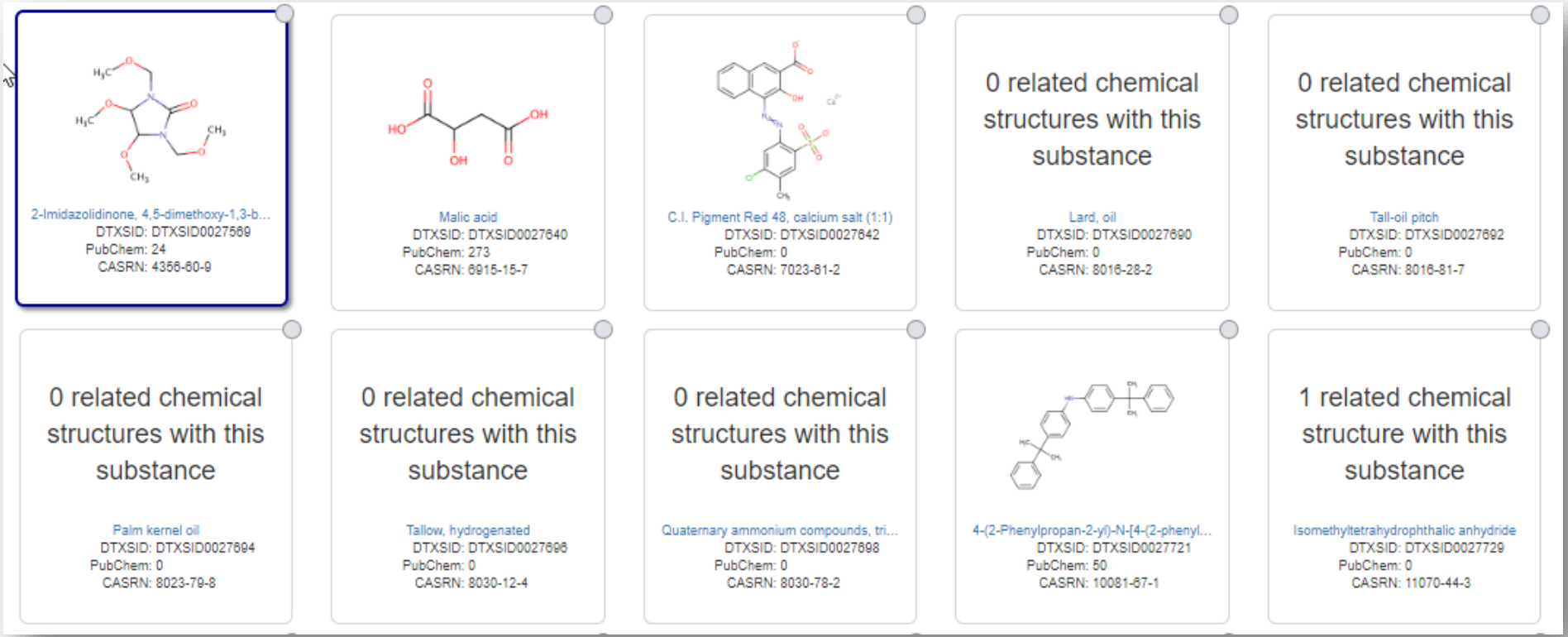
Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB Substance) on the TSCA Inventory

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.

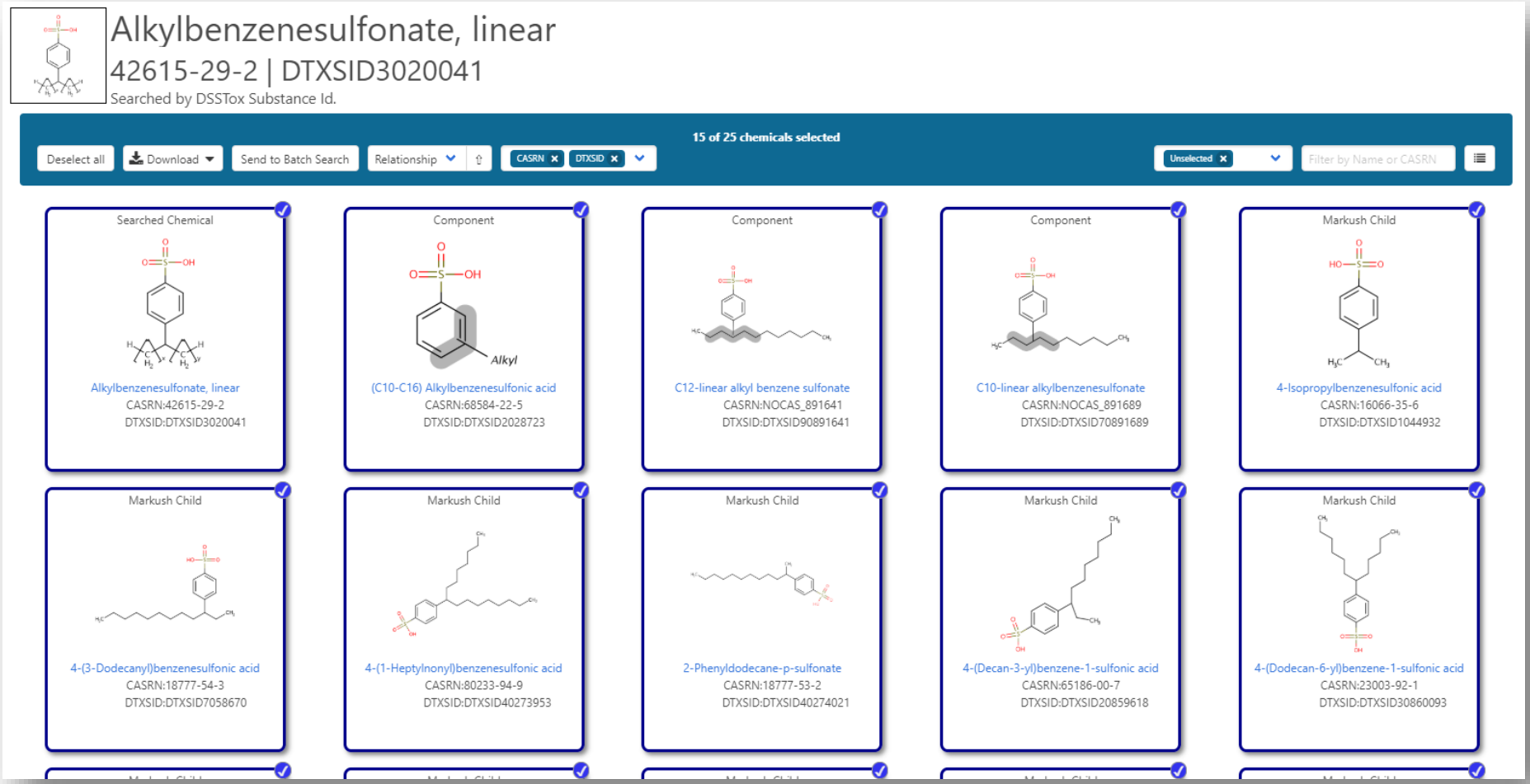
Public TSCA Inventory on Dashboard 31,460 Chemicals (1/24/2020)



Many Chemicals are “Complex”
>14000 chemicals are UVCBs

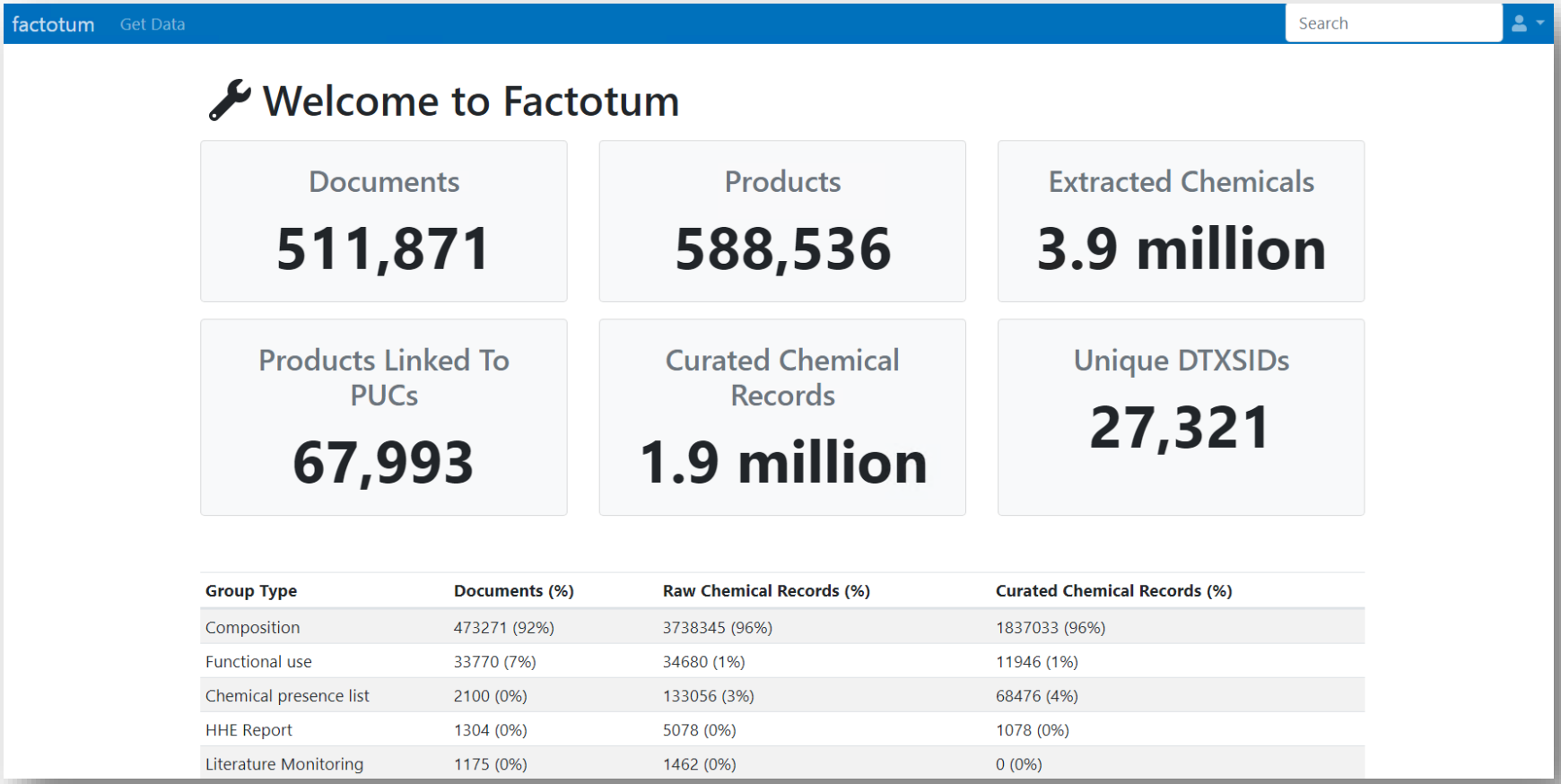


Representing complexity
Markush representations



Chemicals in Commerce

Factotum Database
Consumer Products Database



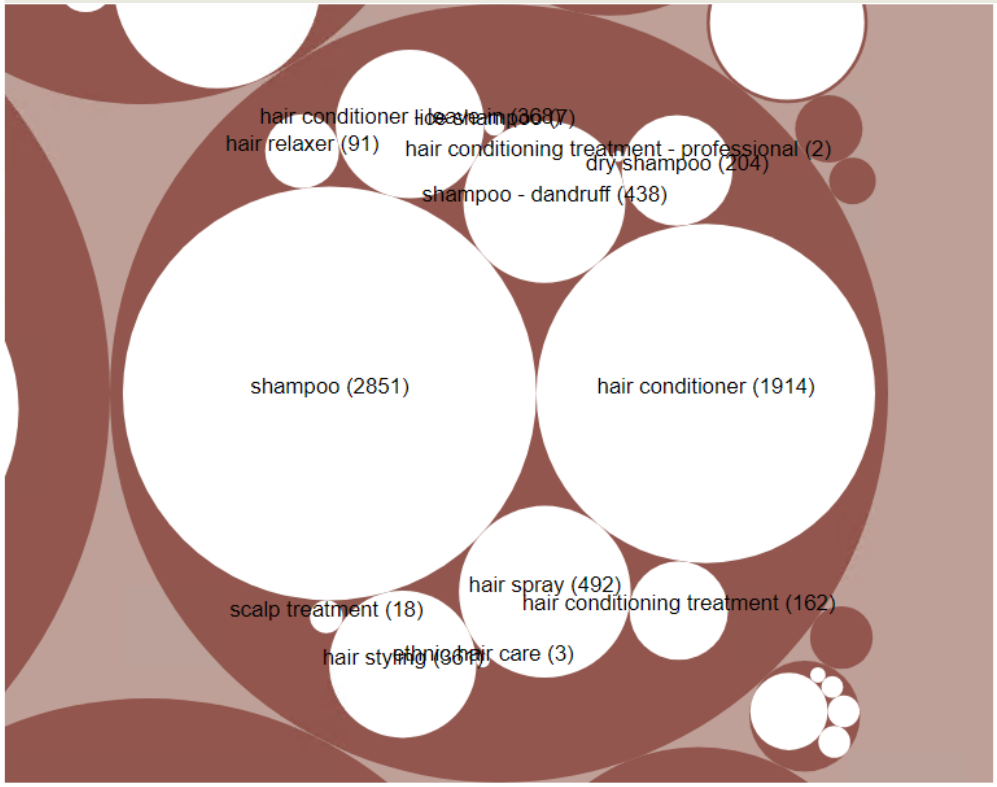
Group Type	Documents (%)	Raw Chemical Records (%)	Curated Chemical Records (%)
Composition	473271 (92%)	3738345 (96%)	1837033 (96%)
Functional use	33770 (7%)	34680 (1%)	11946 (1%)
Chemical presence list	2100 (0%)	133056 (3%)	68476 (4%)
HHE Report	1304 (0%)	5078 (0%)	1078 (0%)
Literature Monitoring	1175 (0%)	1462 (0%)	0 (0%)

Factotum Database
Consumer Products Database

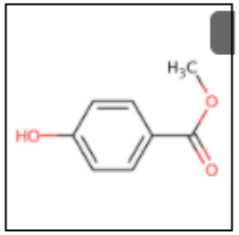
Formulation PUCs

General Category - Product Family - Product Type

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



Sources of Exposure to Chemicals



Methylparaben

99-76-3 | DTXSID4022529

Searched by DSSTox Substance Id.

Download | Columns | 10

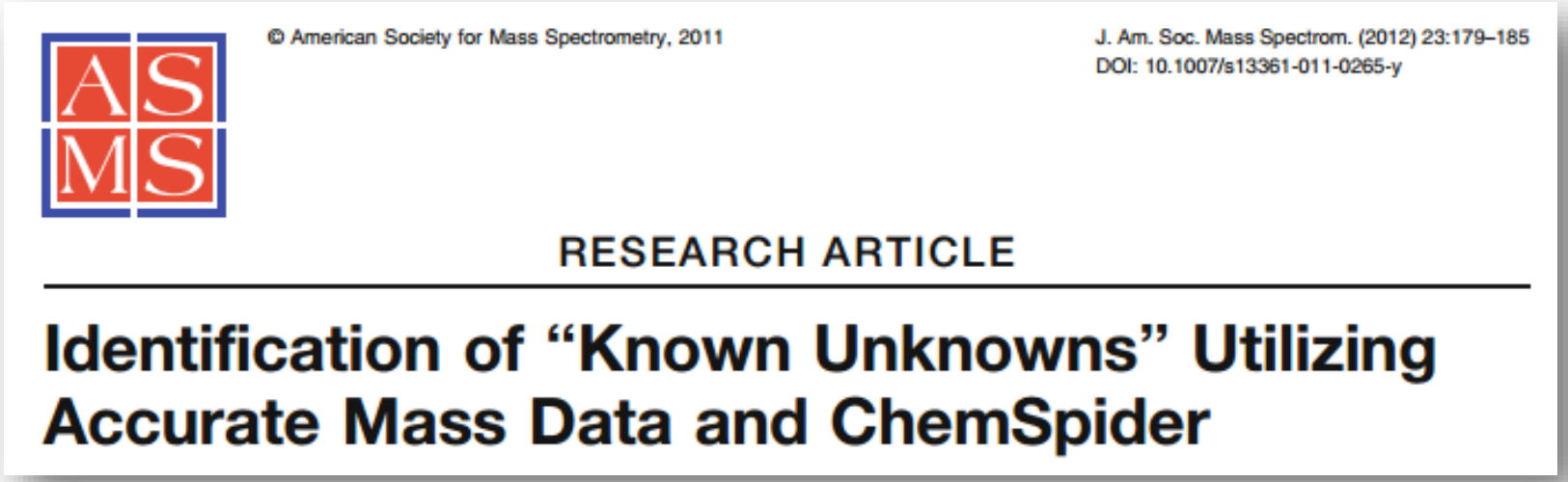
Product Name
rp super filter coat no. 418
rp super filter cost #421 bulk
sbs-30 waterless skin cleaner
sbs-40 medicated skin cream
schilling test kit
schilling test kt(cyanocobalamin cobalt co 57)_reference std
solution_ a e r pads
super filter coat 412
wondermask p peelable mask_ 2211 bulk
YOUTOPIA- PET SHAMPOO

Search query

...raction	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



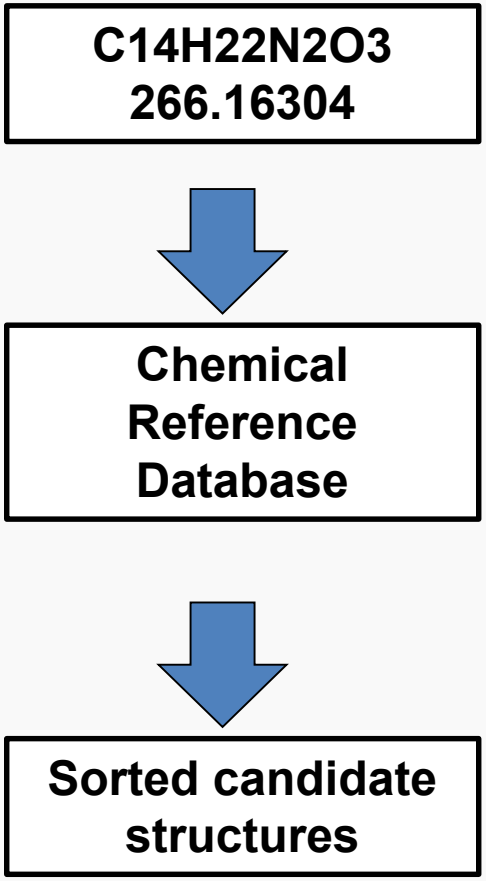
J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

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

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

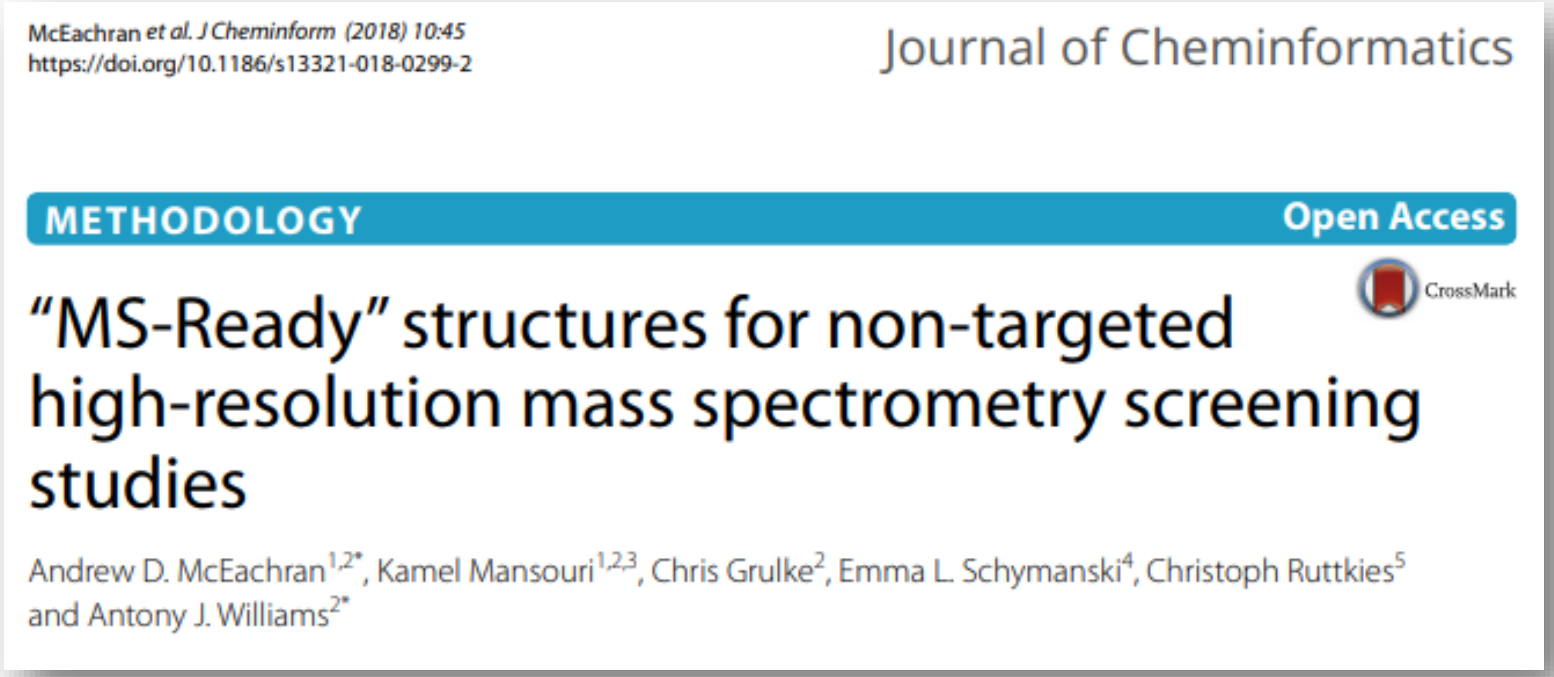
How did performance compare?

	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

"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 Grulke[2], Emma L. Schymanski[4], Christoph Ruttkies[5] and Antony J. Williams[2*]

“MS-Ready” Structures

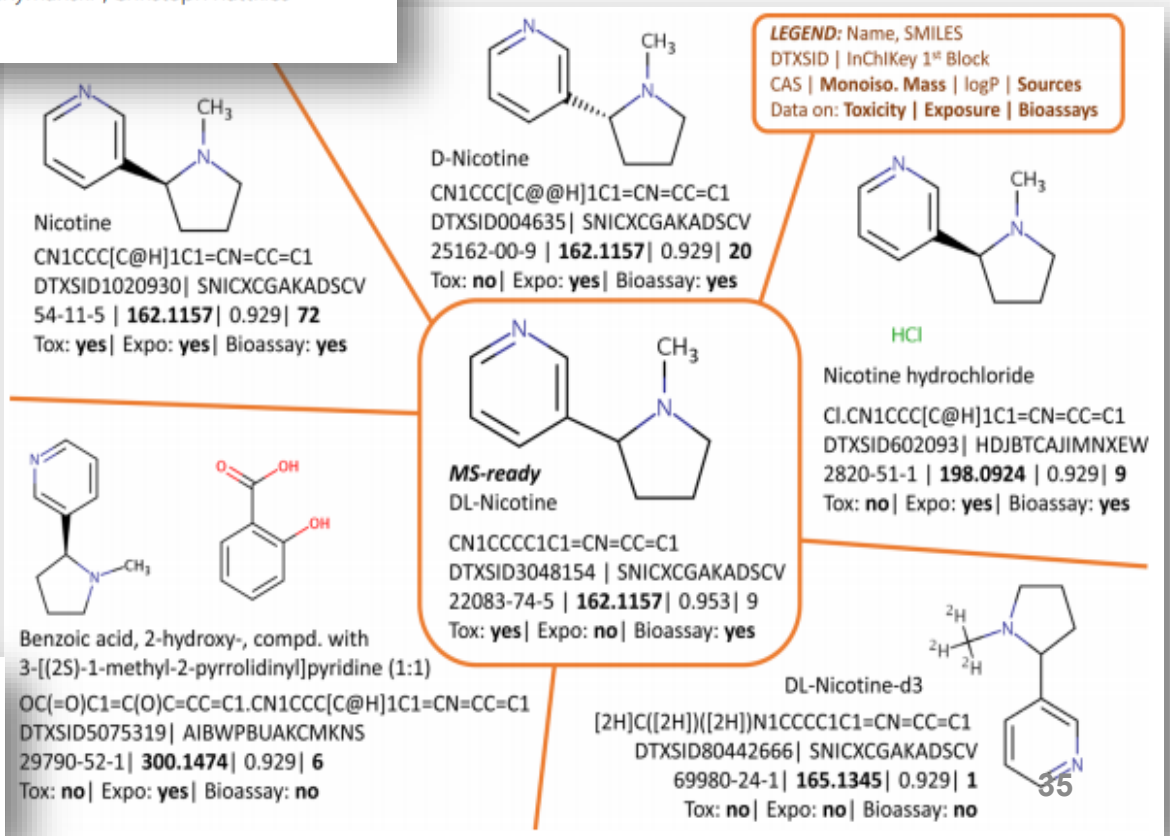
McEachran et al. J Cheminform (2018) 10:45
https://doi.org/10.1186/s13321-018-0299-2

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 Grulke[2], Emma L. Schymanski[4], Christoph Ruttkies[5] and Antony J. Williams[2*]



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

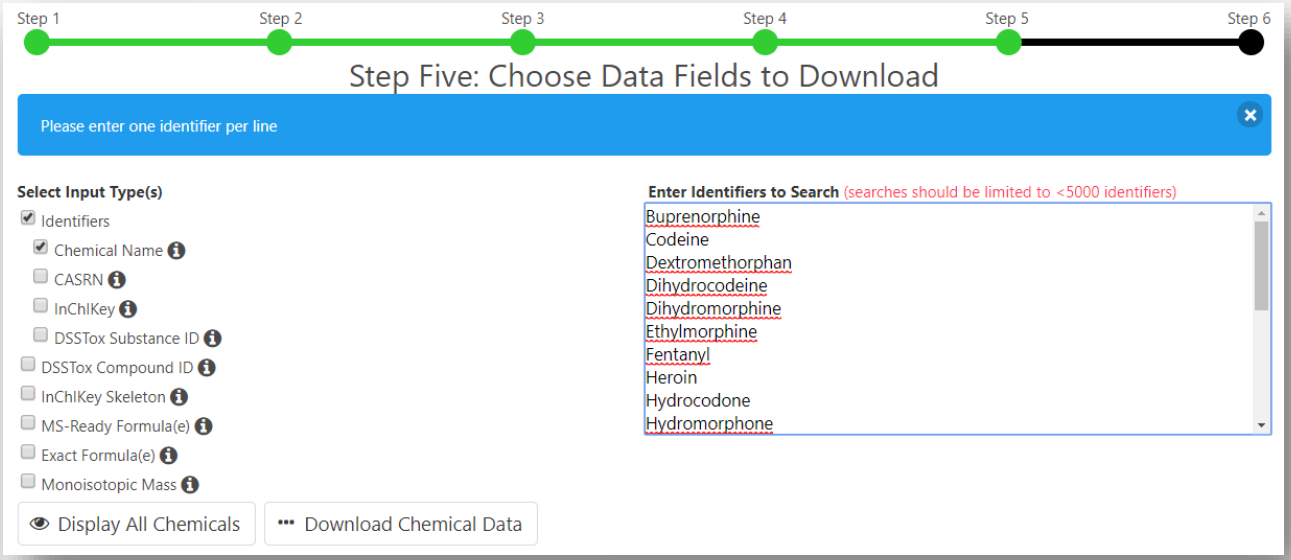
Open Science for Identifying “Known Unknown” Chemicals

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

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



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

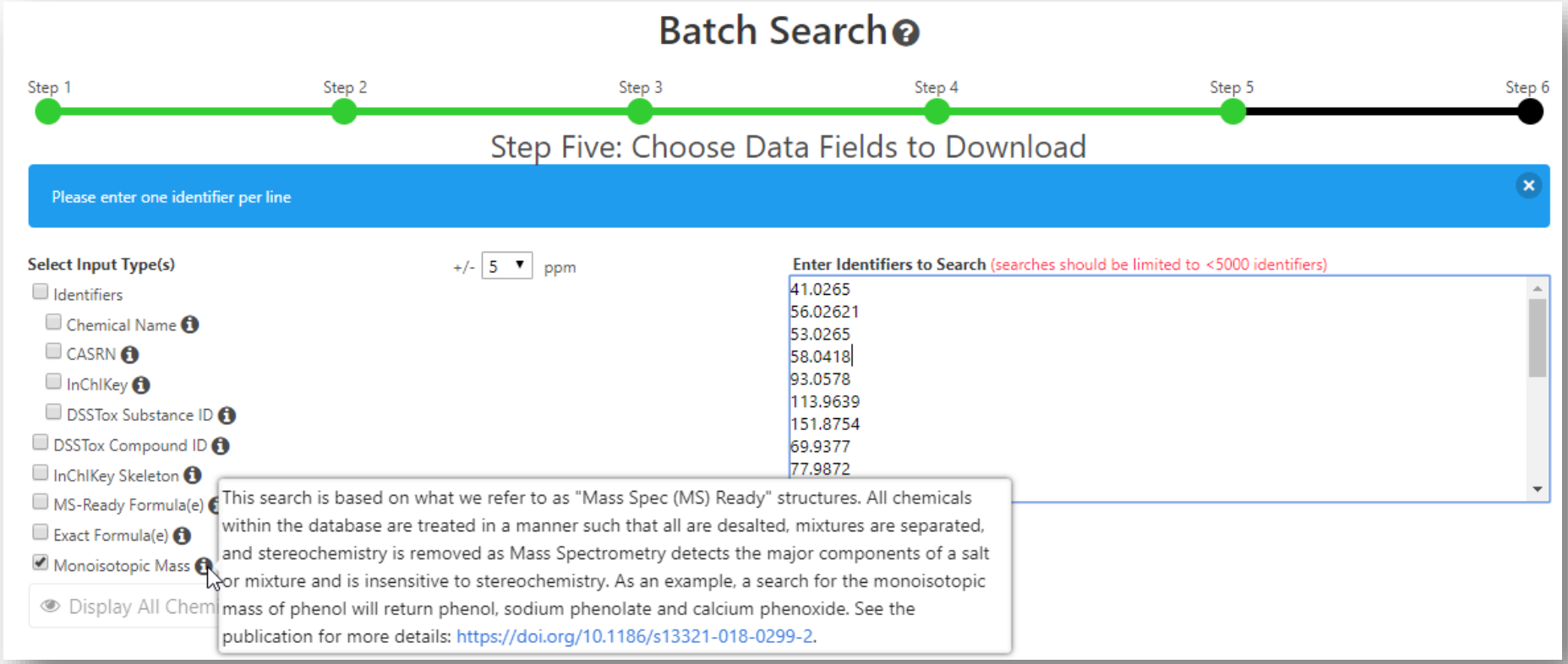
- [ ] Mol File
- [ ] SMILES
- [ ] InChI String
- [x] MS-Ready SMILES
- [ ] QSAR-Ready SMILES

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(=

Batch Searching Formula/Mass



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

EPA United States Environmental Protection Agency

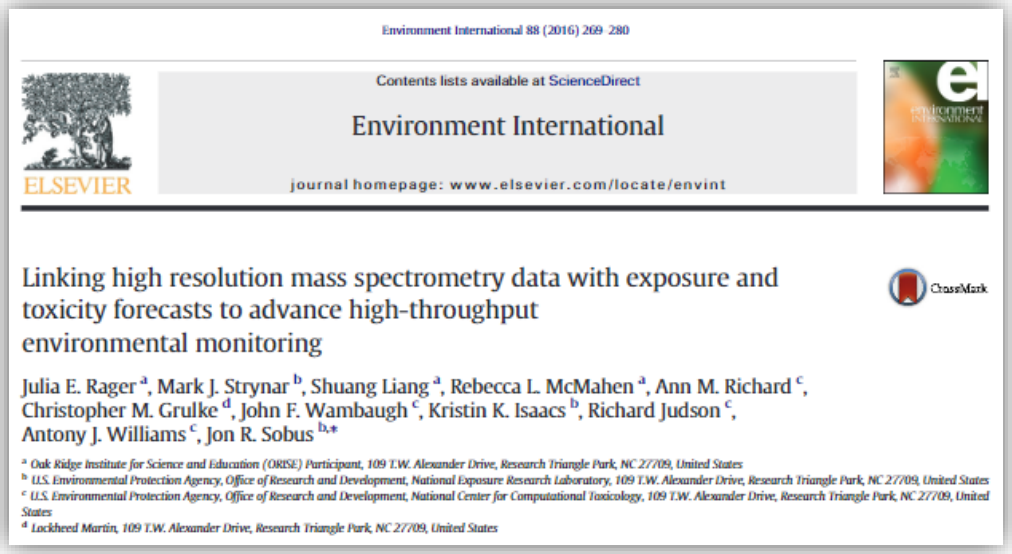
	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576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

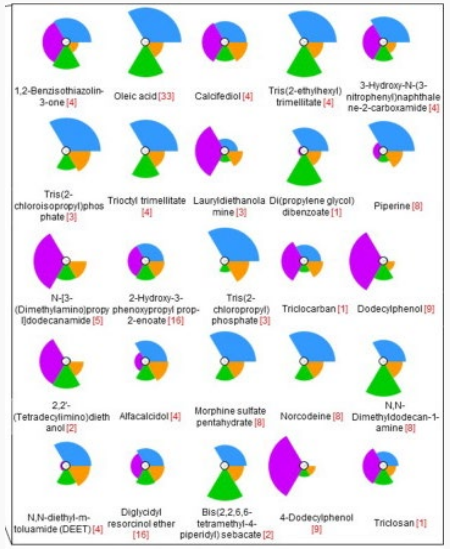
Environment International

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

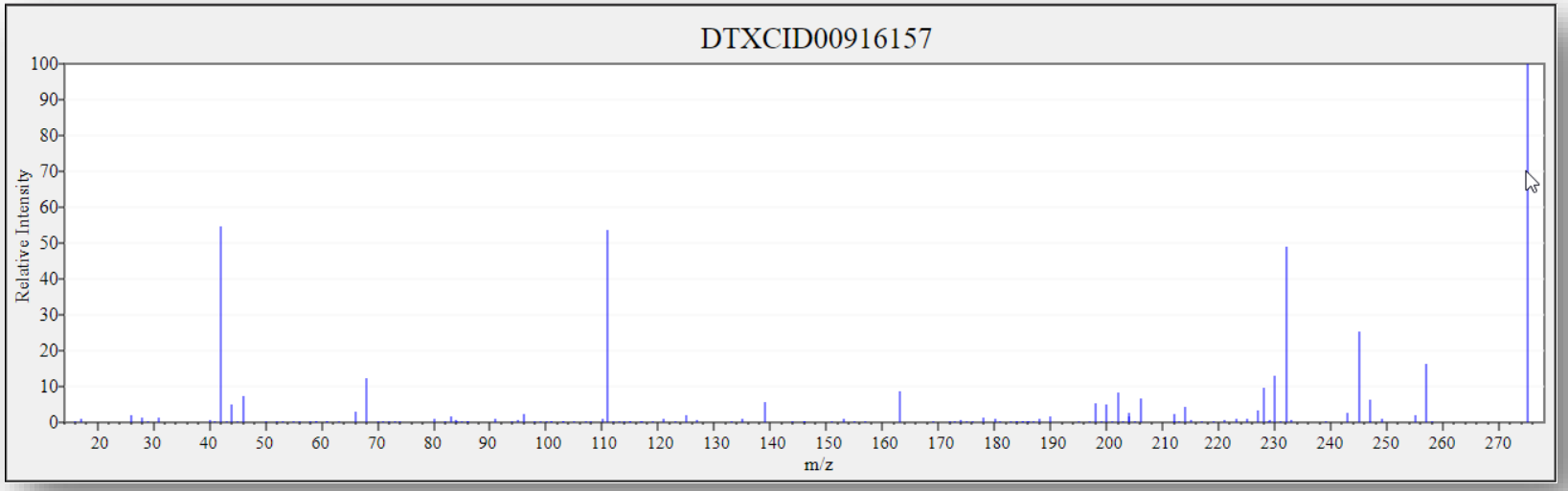


Dashboard Data for in silico spectral prediction

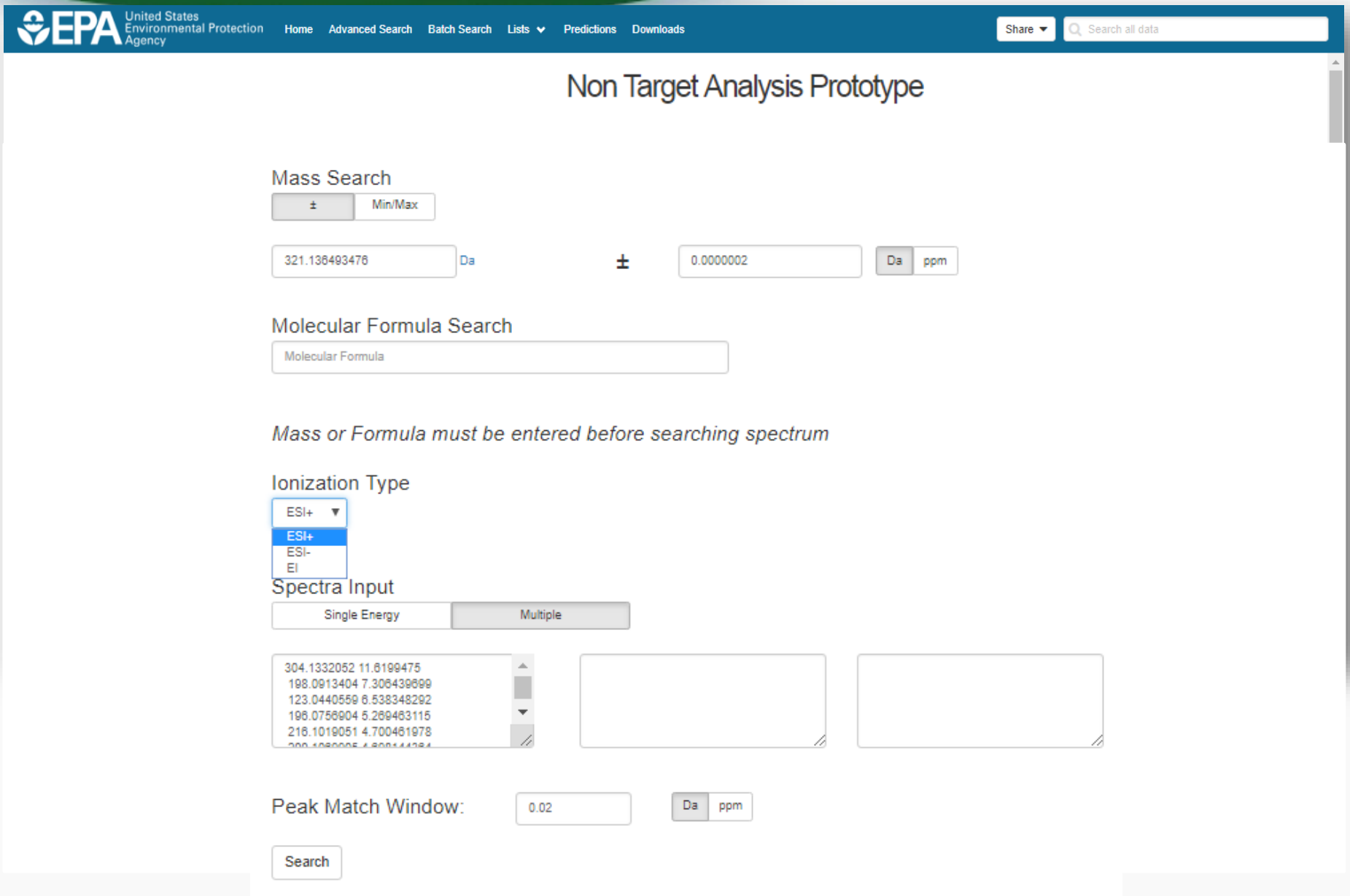
Predicted Mass Spectra

http://cfmid.wishartlab.com/

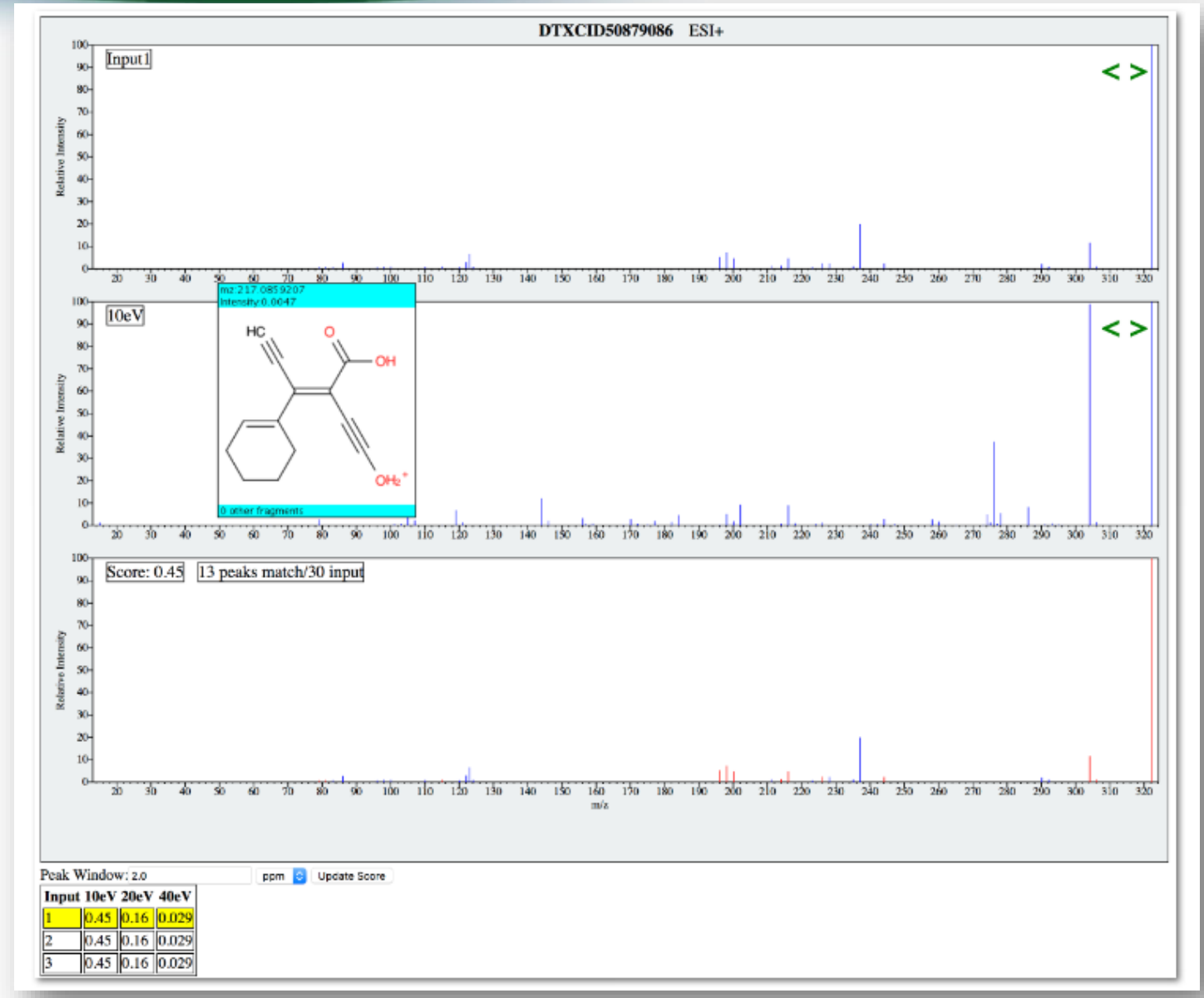
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra



Spectral Viewer Comparison



Published: Alex Chao et al

Analytical and Bioanalytical Chemistry

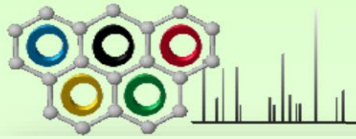
RESEARCH PAPER

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

CASMI 2012-2017 revisited

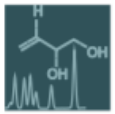


CASMI

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!



metabolites

MDPI

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.

United States Environmental Protection Agency

CATEGORIES
- Toxicology

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

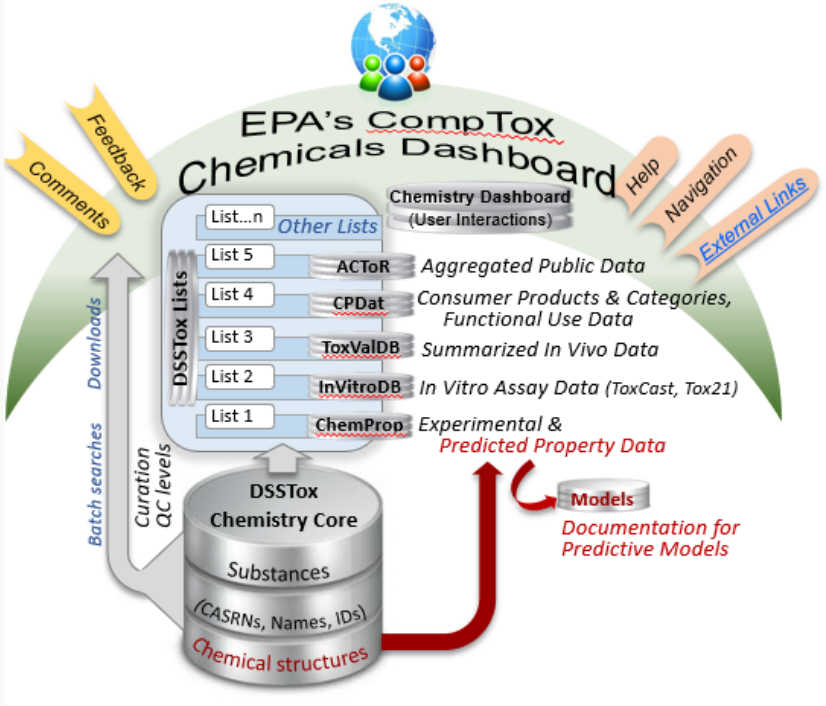
LICENCE
CC0

EXPORT
RefWorks
BibTeX
Ref. manager
Endnote

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

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

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