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

Accessing information for chemicals in hydraulic fracturing fluids using the US EPA CompTox Chemistry Dashboard

Antony Williams[1], Andrew McEachran[2], Jon Sobus[3], Chris Grulke[1] and Ann Richard[1]

1) *National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC*
2) *Oak Ridge Institute of Science and Education (ORISE) Research Participant, Research Triangle Park, NC*
3) *National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*

March 2018
ACS Spring Meeting, New Orleans

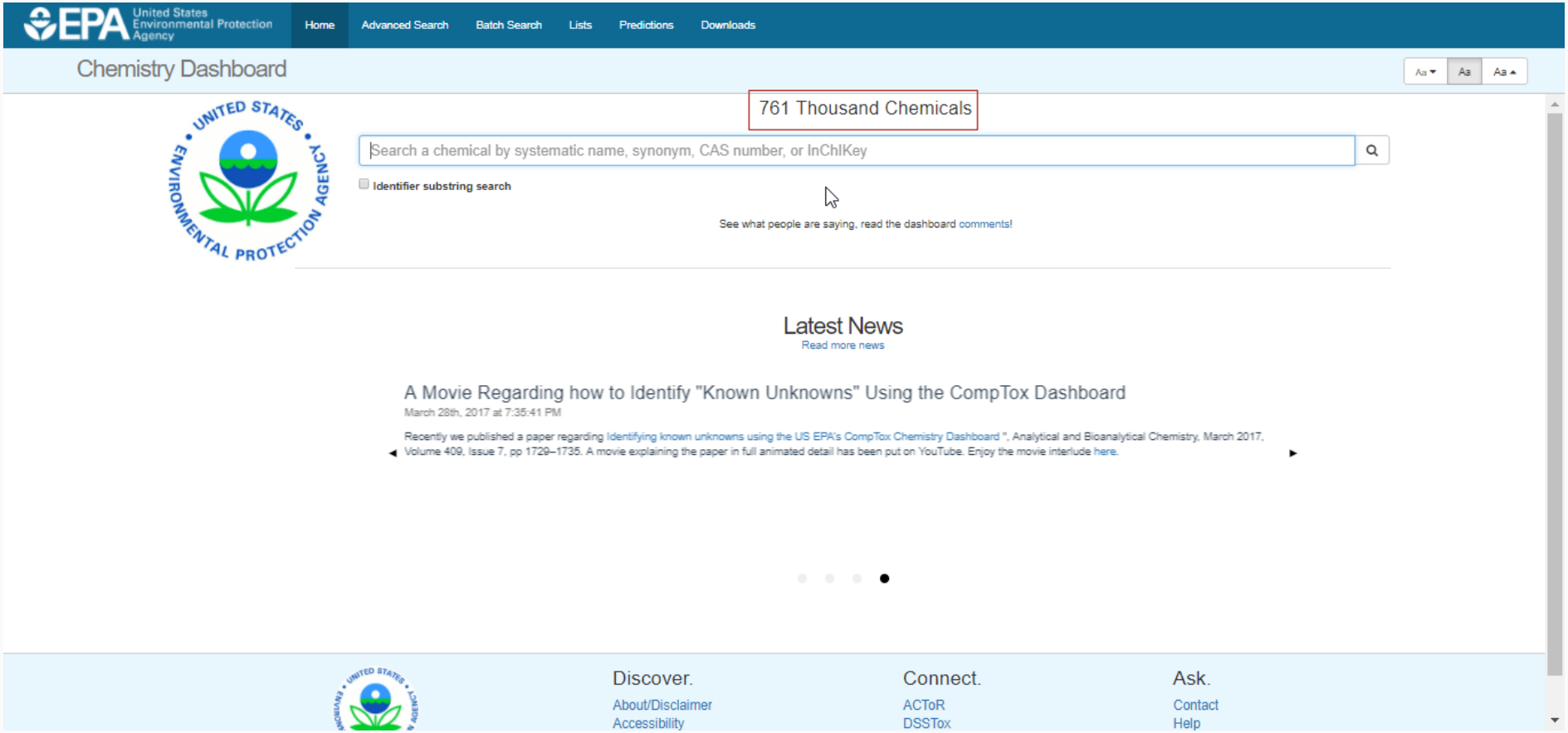
The CompTox Chemistry Dashboard

- A **publicly accessible website** delivering access:
  - ~760,000 chemicals with related property data
  - Experimental and predicted physicochemical property data
  - Experimental Human and Ecological hazard data
  - Integration to “biological assay data” for 1000s of chemicals
  - Information regarding consumer products containing chemicals
  - Links to other agency websites and public data resources
  - “Literature” searches for chemicals using public resources
  - “Batch searching” for thousands of chemicals
  - Real time prediction of physchem and toxicity endpoints

  - **DOWNLOADABLE Open Data for reuse and repurposing**

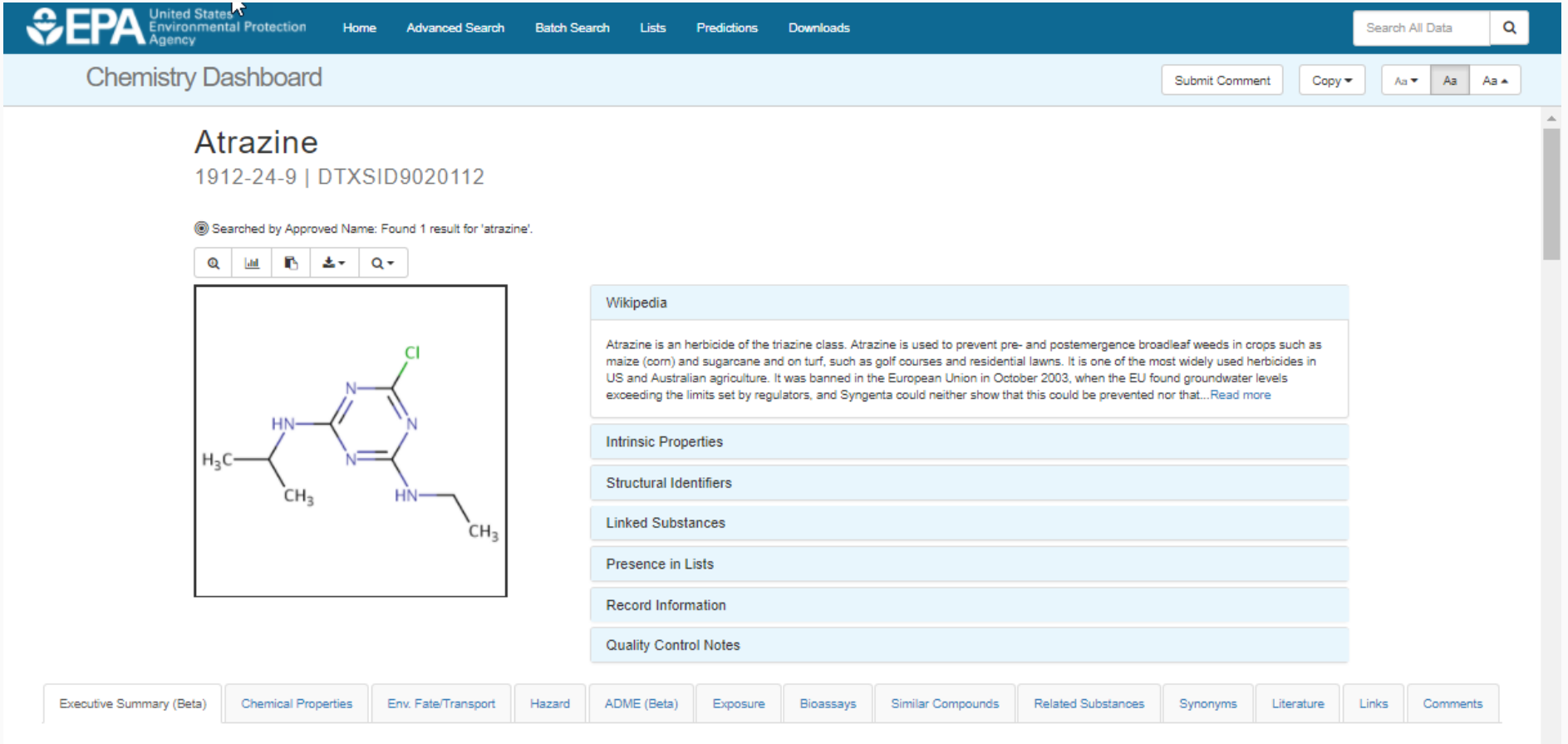
CompTox Chemistry Dashboard

https://comptox.epa.gov/dashboard

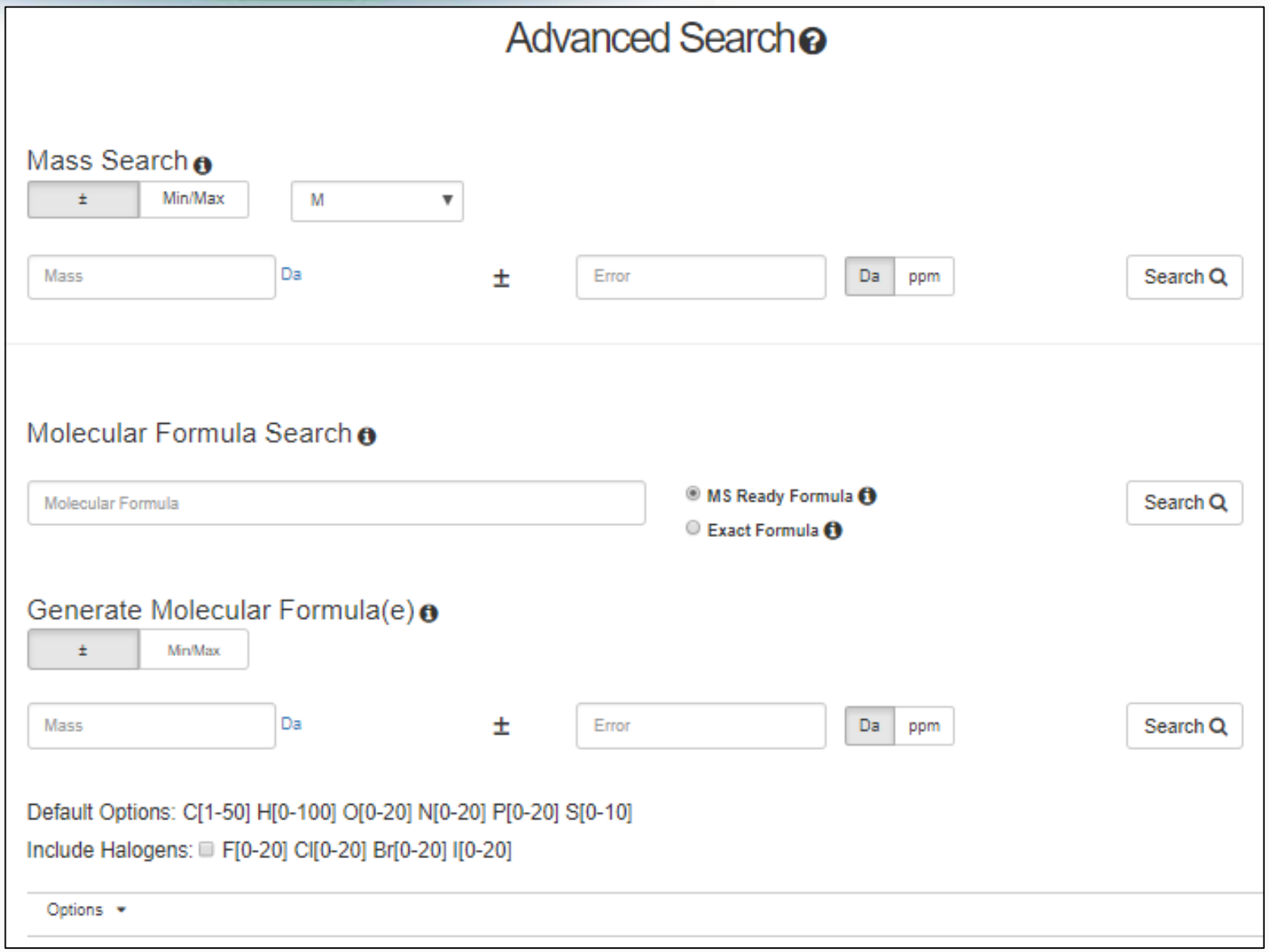


1 of ~761,000 Chemical Pages

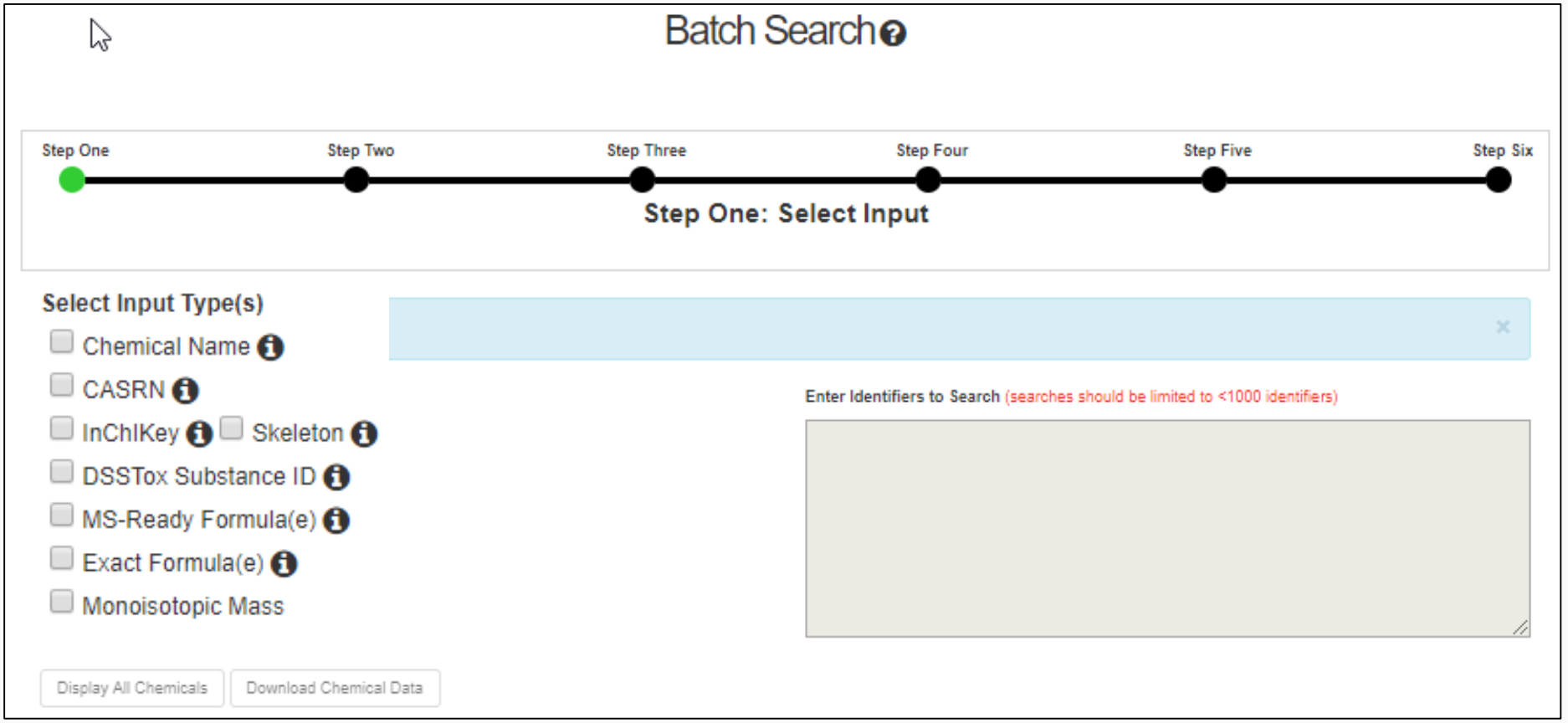
EPA United States Environmental Protection Agency



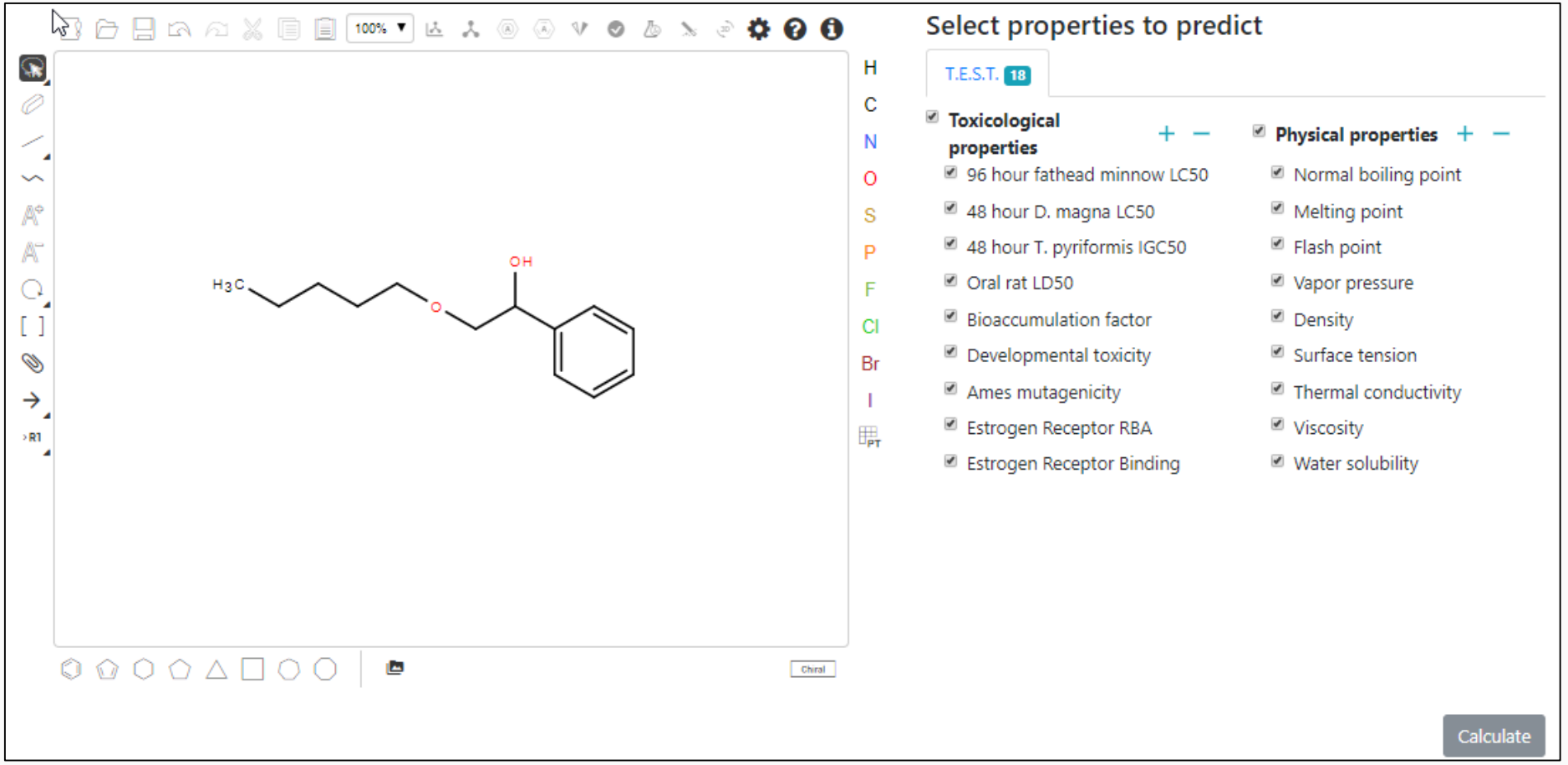
Advanced Searches



Batch Searches



Real-Time Predictions



Real-Time Predictions

Property	Experimental Value	Prediction				
		Consensus	Hierarchical clustering	Single model	Group contribution	Nearest neighbor
96 hour fathead minnow LC50		4.477 -Log10(mol/L) 6.954 mg/L	4.195 -Log10(mol/L) 13.288 mg/L	3.994 -Log10(mol/L) 21.110 mg/L	3.478 -Log10(mol/L) 69.224 mg/L	6.238 -Log10(mol/L) 0.120 mg/L
48 hour D. magna LC50		4.398 -Log10(mol/L) 8.328 mg/L	3.877 -Log10(mol/L) 27.677 mg/L	4.039 -Log10(mol/L) 19.026 mg/L	4.084 -Log10(mol/L) 17.173 mg/L	5.593 -Log10(mol/L) 0.532 mg/L
48 hour T. pyriformis IGC50		4.063 -Log10(mol/L) 18.039 mg/L	3.731 -Log10(mol/L) 38.668 mg/L		3.386 -Log10(mol/L) 85.610 mg/L	5.070 -Log10(mol/L) 1.773 mg/L
Oral rat LD50		1.758 -Log10(mol/kg) 3640.950 mg/kg	1.982 -Log10(mol/kg) 2172.756 mg/kg			1.533 -Log10(mol/kg) 6101.245 mg/kg
Bioaccumulation factor		1.797 Log10 62.700	2.202 Log10 159.310	1.287 Log10 19.346	1.181 Log10 15.157	2.520 Log10 330.834
Developmental toxicity		false	false	false		true
Ames mutagenicity		false	false			false
Estrogen Receptor RBA		-3.075 Log10 $8.418*10^{-4}$	-3.078 Log10 $8.356*10^{-4}$	-3.720 Log10 $1.907*10^{-4}$		-2.427 Log10 0.004
Estrogen Receptor Binding		true	true	true	false	true

Real-Time Predictions

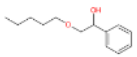
Predicted Water solubility at 25°C for OC(C=1C=CC=CC1)COCCCCC from Consensus method

Prediction results

Endpoint	Experimental value	Predicted value
Water solubility at 25°C -Log10(mol/L)	N/A	2.46
Water solubility at 25°C mg/L	N/A	723.26

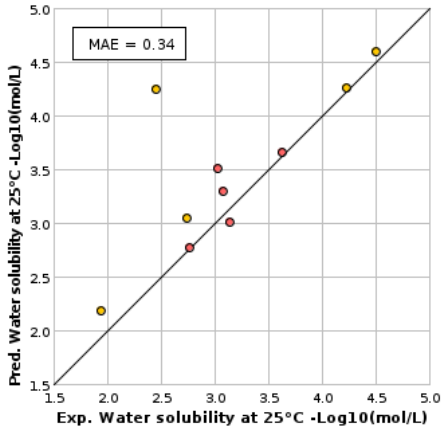
Individual Predictions

Method	Predicted value -Log10(mol/L)
Hierarchical clustering	2.42
Group contribution	2.32
Nearest neighbor	2.64



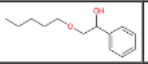
Predictions for the test chemical and for the most similar chemicals

Prediction results (colors defined in table below)

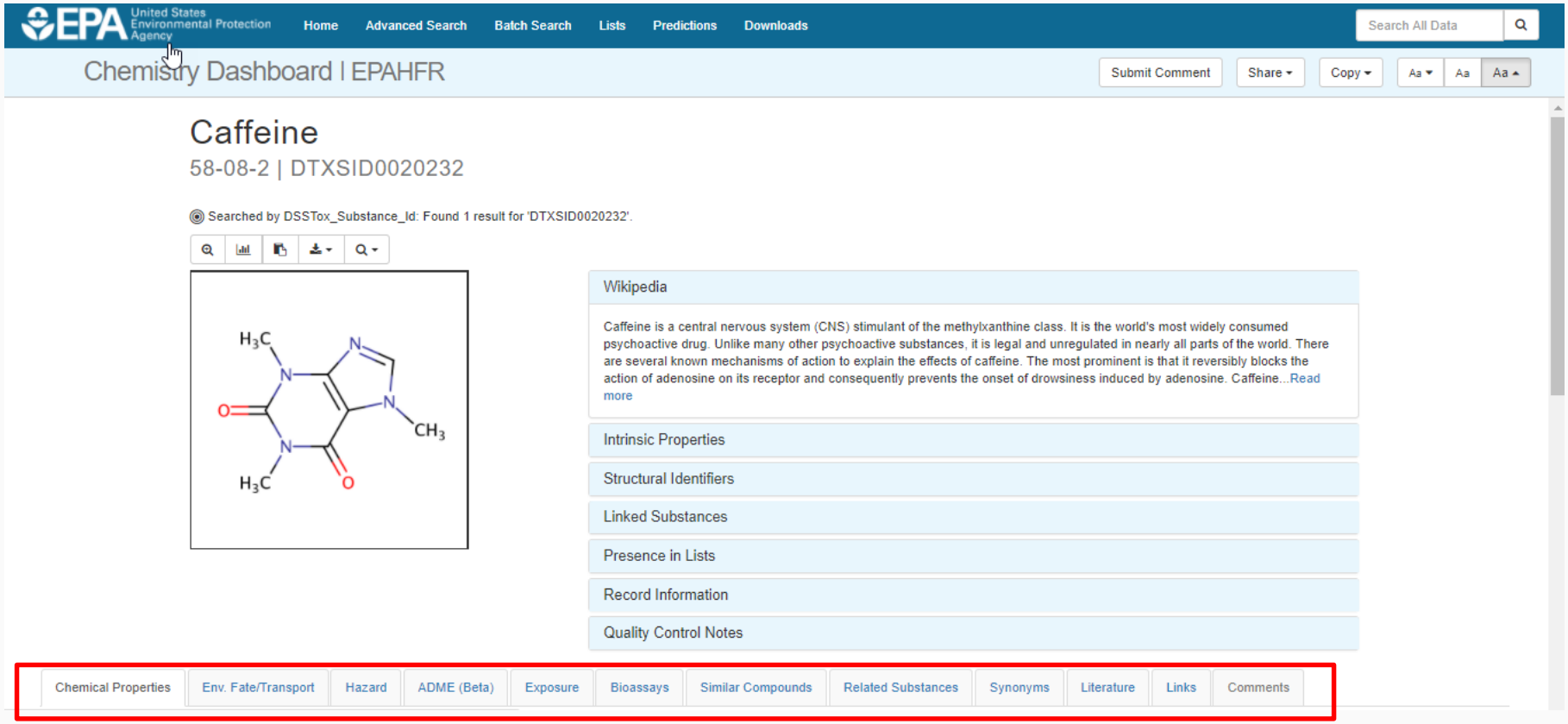


Chemicals	MAE*
Entire set	0.58
Similarity coefficient ≥ 0.5	0.34

*Mean absolute error in -Log10(mol/L)

CAS	Structure	Similarity Coefficient	Experimental value -Log10(mol/L)	Predicted value -Log10(mol/L)
OC(C=1C=CC=CC1)COCCCCC (test chemical)			N/A	2.46
104-40-5		0.68	4.50	4.60
1219-38-1		0.67	4.22	4.26

What's under the **Data Tabs**?
Focus on Hydraulic Fracturing Chemicals



Surfacing Lists of Chemicals

- Specific subsets of chemicals, “lists”, can be displayed on the dashboard
- If there are chemicals that map together then these link to existing:
  - Property data
  - Hazard data
  - Exposure data
  - *In vitro* bioassay data
  - Documents and Literature

EPA's Hydraulic Fracturing Site

https://www.epa.gov/hfstudy

EPA United States Environmental Protection Agency

EPA United States Environmental Protection Agency

Environmental Topics | Laws & Regulations | About EPA | Search EPA.gov

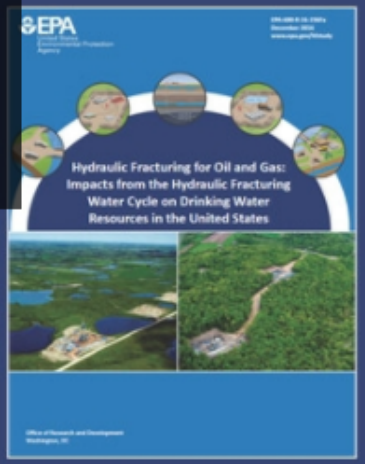
EPA's Study of Hydraulic Fracturing for Oil and Gas and Its Potential Impact on Drinking Water Resources

CONTACT US

SHARE

Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States

Read the final assessment.



1 2 3

- Read EPA's blog on the final assessment.
- View the presentation slides from EPA's December 14, 2016 public webinar.

Draft Database

- Early draft database is available associated with the June 2015 draft report
- Contained list of CAS Numbers/Names aggregated from various sources

What were the original sources?

Text Box 9-2. The EPA's List of Chemicals Identified in Hydraulic Fracturing Fluids and/or Produced Water.

This chemical list progressed through multiple iterations as the EPA's hydraulic fracturing study was developed, culminating in the list of 1,606 chemicals presented in this report.

The first iteration of this chemical list was published in the interim progress report (U.S. EPA, 2012h), and included 1,026 chemicals that were identified from ten sources of information. Seven of these information sources were documents from federal and state governmental units—including the EPA (U.S. EPA, 2011a, e, 2004a; Material Safety Data Sheets), the U.S. House of Representatives (House of Representatives, 2011), the New York State Department of Environmental Conservation (NYSDEC, 2011), and the Pennsylvania Department of Environmental Protection (PA DEP, 2010a)—which obtained data directly from industry. This includes a list of chemicals provided directly to the EPA by nine well operating companies, representing chemicals used in hydraulic fracturing fluids between 2005 and 2009, and a list of chemicals detected by these companies in produced water from 81 wells. The remaining three sources are as follows: a technical report prepared by the Gas Technology Institute for the Marcellus Shale Coalition, which is a drilling industry trade group (Hayes, 2009); a peer-reviewed journal article by Colborn et al. (2011); and the FracFocus Chemical Disclosure Registry, which is a national hydraulic fracturing chemical registry developed by the Ground Water Protection Council and the Interstate Oil and Gas Compact Commission (GWPC, 2012).

- http://ofmpub.epa.gov/eims/eimscomm.getfile?p_download_id=530159

Delivering a Better Database

- An improved database would provide:
  - Curated CAS Number-Name mappings
  - Chemical structures
  - Access to related data of interest

Registration and Curation

- Consolidation and registration of the original chemical lists into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs

Names to CASRN Mappings

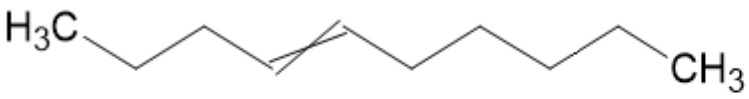
Substance Mapping

(1 of 66) 1 2 3 4 5 6 7 8 9 10 25

Source Casrn	Source Name	Hit Substance_ID	Hit Casrn	Hit Name	
19398-89-1	4-Decene	DTXSID50876156	19689-18-0	4-Decene	Remove Validation
112926-00-8	silica gel, cryst. -free	DTXSID9029851	112926-00-8	Hydrated silica	Remove Validation
124-28-7	1-Octadecanamine, N,N-dimethyl-	DTXSID4027026	124-28-7	N,N-Dimethyl-1-octadecanamine	Remove Validation
1330-43-4	Boron sodium oxide	DTXSID2034388	1330-43-4	Sodium tetraborate	Remove Validation
13492-26-7	Mono- and di-potassium salts of phosphorous acid	DTXSID9035961	13492-26-7	Phosphonic acid, potassium salt (1:2)	Remove Validation
135-37-5	Glycine, N-(carboxymethyl)-N-(2-hydroxyethyl)-, disodium salt	DTXSID8042008	135-37-5	Ethanoldiglycine disodium salt	Remove Validation

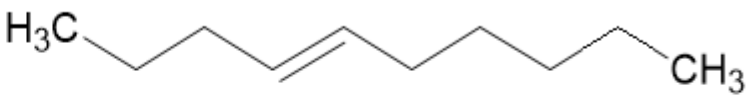
Subtleties

“4-Decene”



19398-89-1

E/Z-stereochemistry



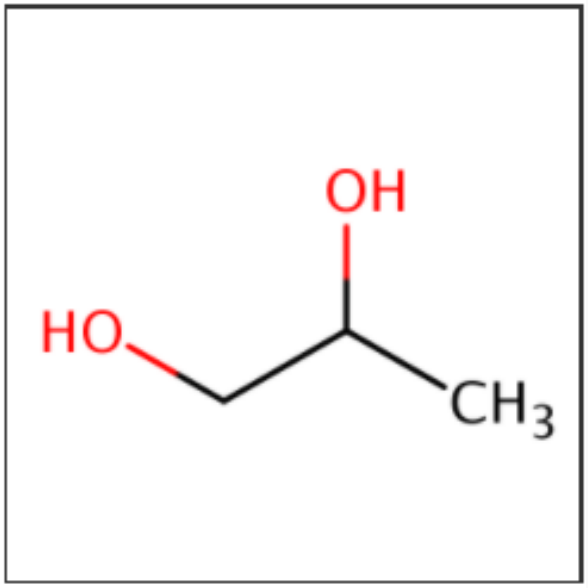
19689-18-0

E-stereochemistry

Registration and Curation

- Registration of the chemical list into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs
  - CASRN checking – Active, Alternate and Deleted

CAS Registry Numbers



1,2-Propylene glycol	Propylenglycol
Propane-1,2-diol	Sentry Propylene Glycol
1,2-Propanediol	Trimethyl glycol
57-55-6 Active CAS-RN	Ucar 35
alpha-Propylene glycol	a-Propylene glycol
(+/-) 1,2-Propanediol	alpha-propyleneglycol
(RS)-1,2-Propanediol	methyl glycol
dl-Propylene glycol	methylethyl glycol
3-01-00-02142 Beilstein Registry Number	1194046-20-2 Deleted CAS-RN
1,2-Propanediol	190913-75-8 Deleted CAS-RN
(.+-.)-1,2-Propanediol	4254-16-4 Deleted CAS-RN
(.+-.)-Propylene glycol	63625-56-9 Deleted CAS-RN

Registration and Curation

- Registration of the chemical list into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs
  - CASRN checking – Active, Alternate and Deleted
  - Misspellings, alternative synonyms, misassociations

Alternative Synonyms

Substance Mapping

(1 of 66) 1 2 3 4 5 6 7 8 9 10 25

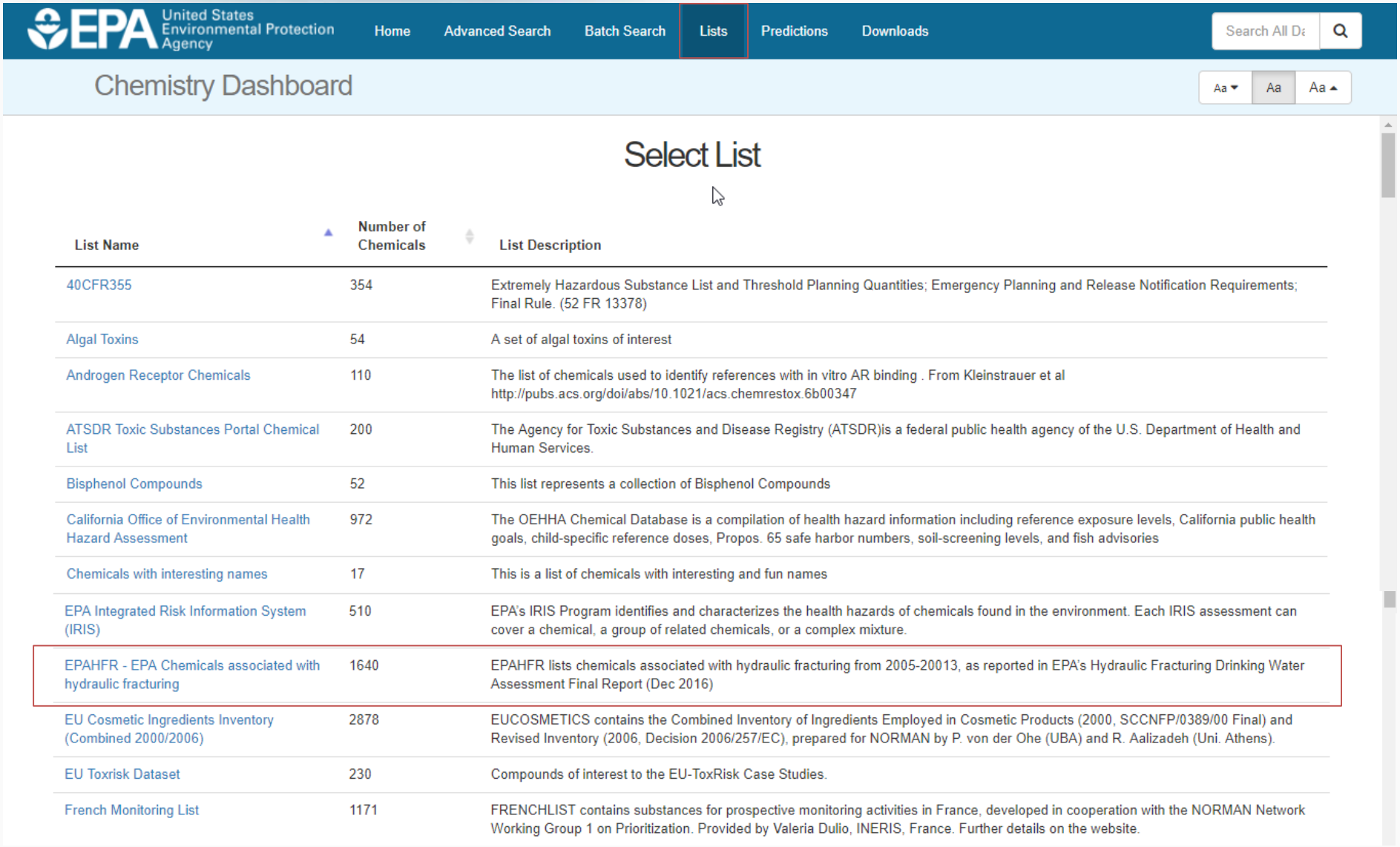
Source Casrn	Source Name	Hit Substance_ID	Hit Casrn	Hit Name	
19398-89-1	4-Decene	DTXSID50876156	19689-18-0	4-Decene	Remove Validation
112926-00-8	silica gel, cryst. - free	DTXSID9029851	112926-00-8	Hydrated silica	Remove Validation
124-28-7	1-Octadecanamine, N,N-dimethyl-	DTXSID4027026	124-28-7	N,N-Dimethyl-1-octadecanamine	Remove Validation
1330-43-4	Boron sodium oxide	DTXSID2034388	1330-43-4	Sodium tetraborate	Remove Validation
13492-26-7	Mono- and di-potassium salts of phosphorous acid	DTXSID9035961	13492-26-7	Phosphonic acid, potassium salt (1:2)	Remove Validation
135-37-5	Glycine, N-(carboxymethyl)-N-(2-hydroxyethyl)-, disodium salt	DTXSID8042008	135-37-5	Ethanoldiglycine disodium salt	Remove Validation

Registration and Curation

- Registration of the chemical list into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs
  - CASRN checking – Active, Alternate and Deleted
  - Misspellings, alternative synonyms, misassociations
  - Iterative checking on how the list was assembled

A List of Lists of Chemicals

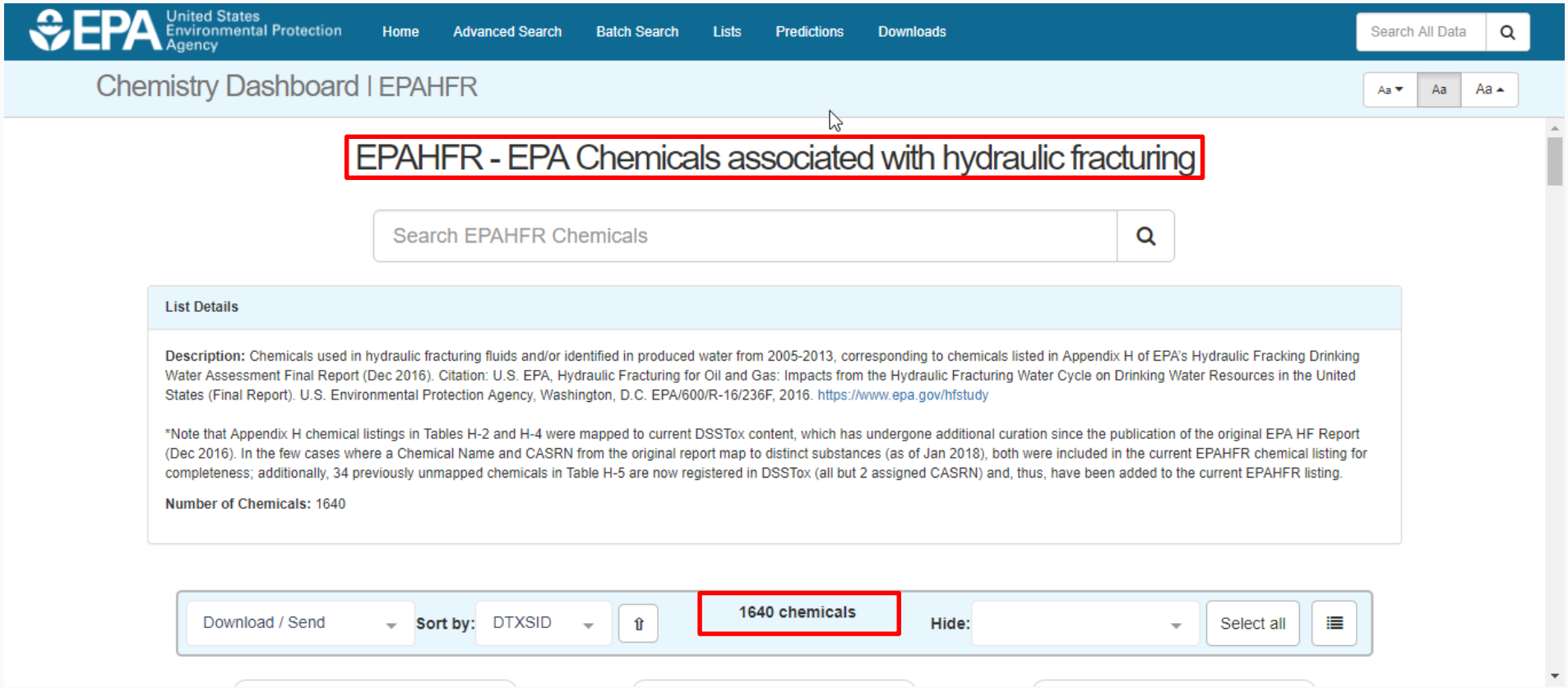
https://comptox.epa.gov/dashboard/chemical_lists



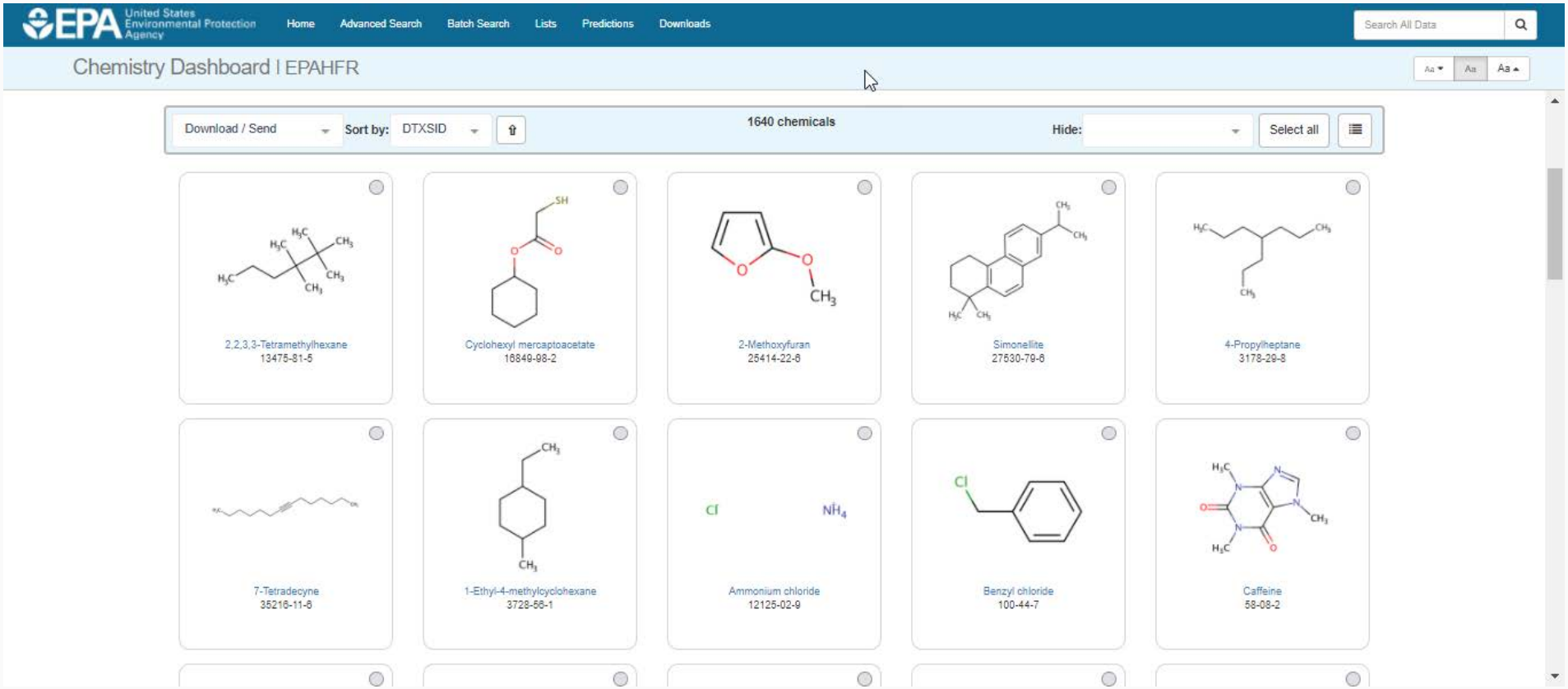
Select List

List Name	Number of Chemicals	List Description
40CFR355	354	Extremely Hazardous Substance List and Threshold Planning Quantities; Emergency Planning and Release Notification Requirements; Final Rule. (52 FR 13378)
Algal Toxins	54	A set of algal toxins of interest
Androgen Receptor Chemicals	110	The list of chemicals used to identify references with in vitro AR binding . From Kleinstrauer et al http://pubs.acs.org/doi/abs/10.1021/acs.chemrestox.6b00347
ATSDR Toxic Substances Portal Chemical List	200	The Agency for Toxic Substances and Disease Registry (ATSDR)is a federal public health agency of the U.S. Department of Health and Human Services.
Bisphenol Compounds	52	This list represents a collection of Bisphenol Compounds
California Office of Environmental Health Hazard Assessment	972	The OEHHA Chemical Database is a compilation of health hazard information including reference exposure levels, California public health goals, child-specific reference doses, Propos. 65 safe harbor numbers, soil-screening levels, and fish advisories
Chemicals with interesting names	17	This is a list of chemicals with interesting and fun names
EPA Integrated Risk Information System (IRIS)	510	EPA's IRIS Program identifies and characterizes the health hazards of chemicals found in the environment. Each IRIS assessment can cover a chemical, a group of related chemicals, or a complex mixture.
EPAHFR - EPA Chemicals associated with hydraulic fracturing	1640	EPAHFR lists chemicals associated with hydraulic fracturing from 2005-20013, as reported in EPA's Hydraulic Fracturing Drinking Water Assessment Final Report (Dec 2016)
EU Cosmetic Ingredients Inventory (Combined 2000/2006)	2878	EUCOSMETICS contains the Combined Inventory of Ingredients Employed in Cosmetic Products (2000, SCCNFP/0389/00 Final) and Revised Inventory (2006, Decision 2006/257/EC), prepared for NORMAN by P. von der Ohe (UBA) and R. Aalizadeh (Uni. Athens).
EU Toxrisk Dataset	230	Compounds of interest to the EU-ToxRisk Case Studies.
French Monitoring List	1171	FRENCHLIST contains substances for prospective monitoring activities in France, developed in cooperation with the NORMAN Network Working Group 1 on Prioritization. Provided by Valeria Dulio, INERIS, France. Further details on the website.

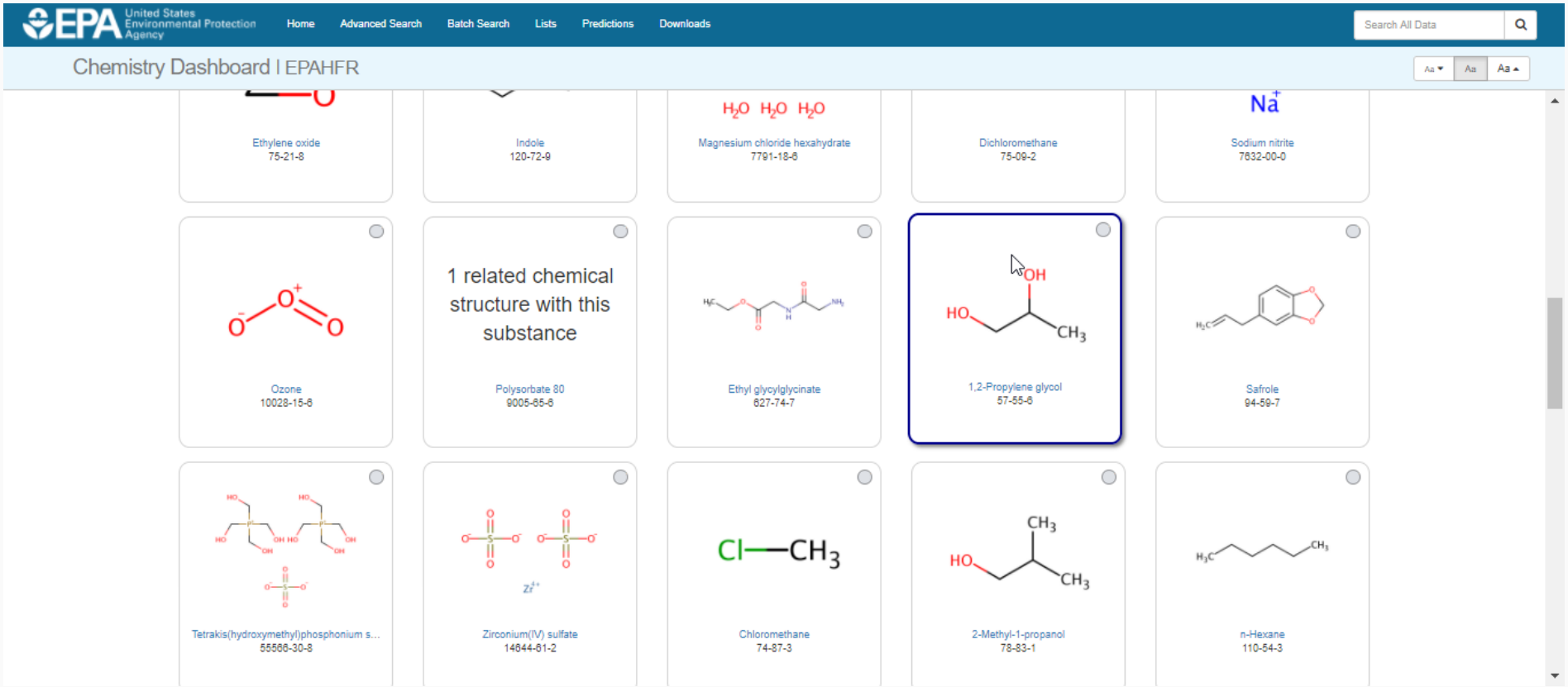
The EPA List of Hydraulic Fracturing Chemicals



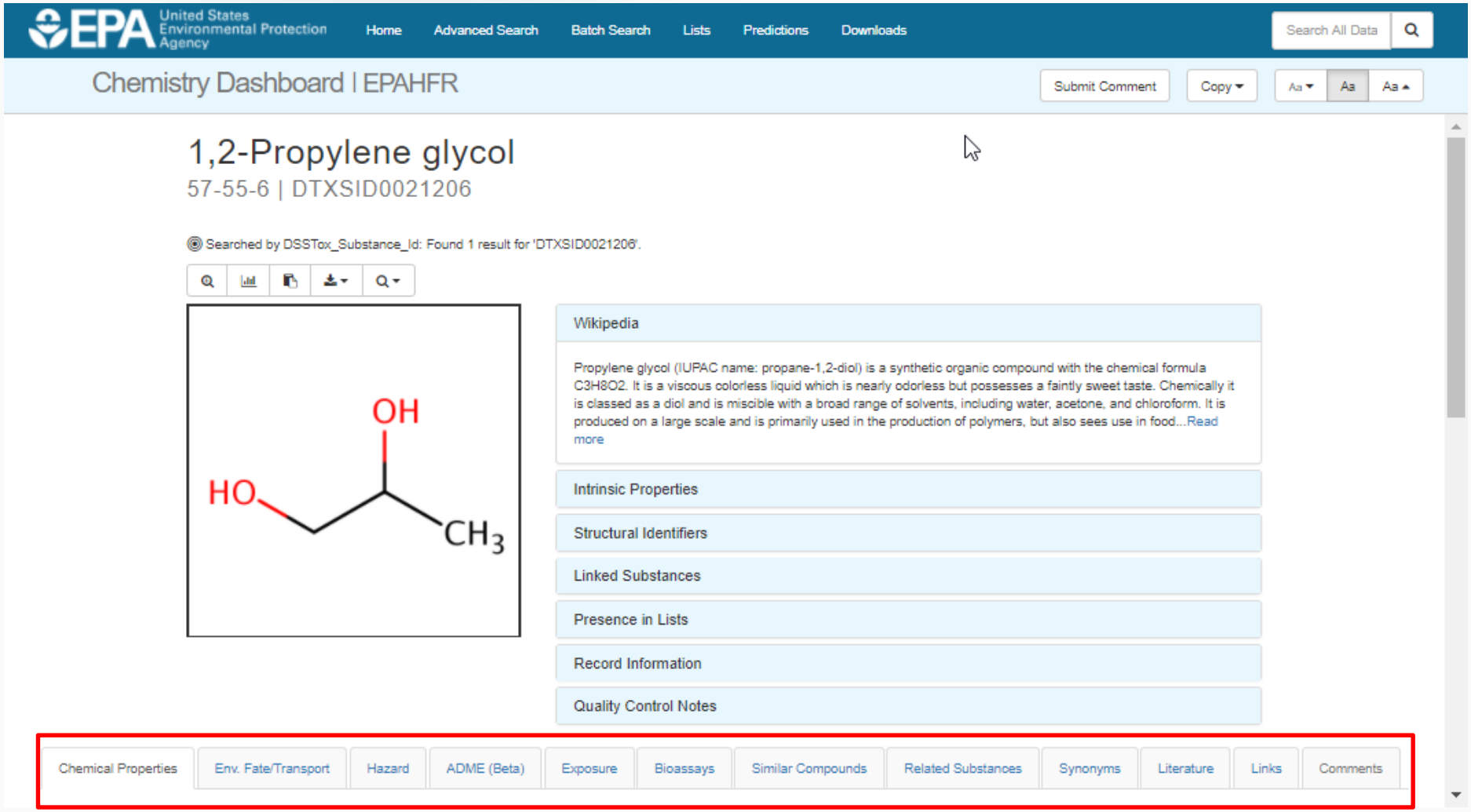
The List as “Structures”



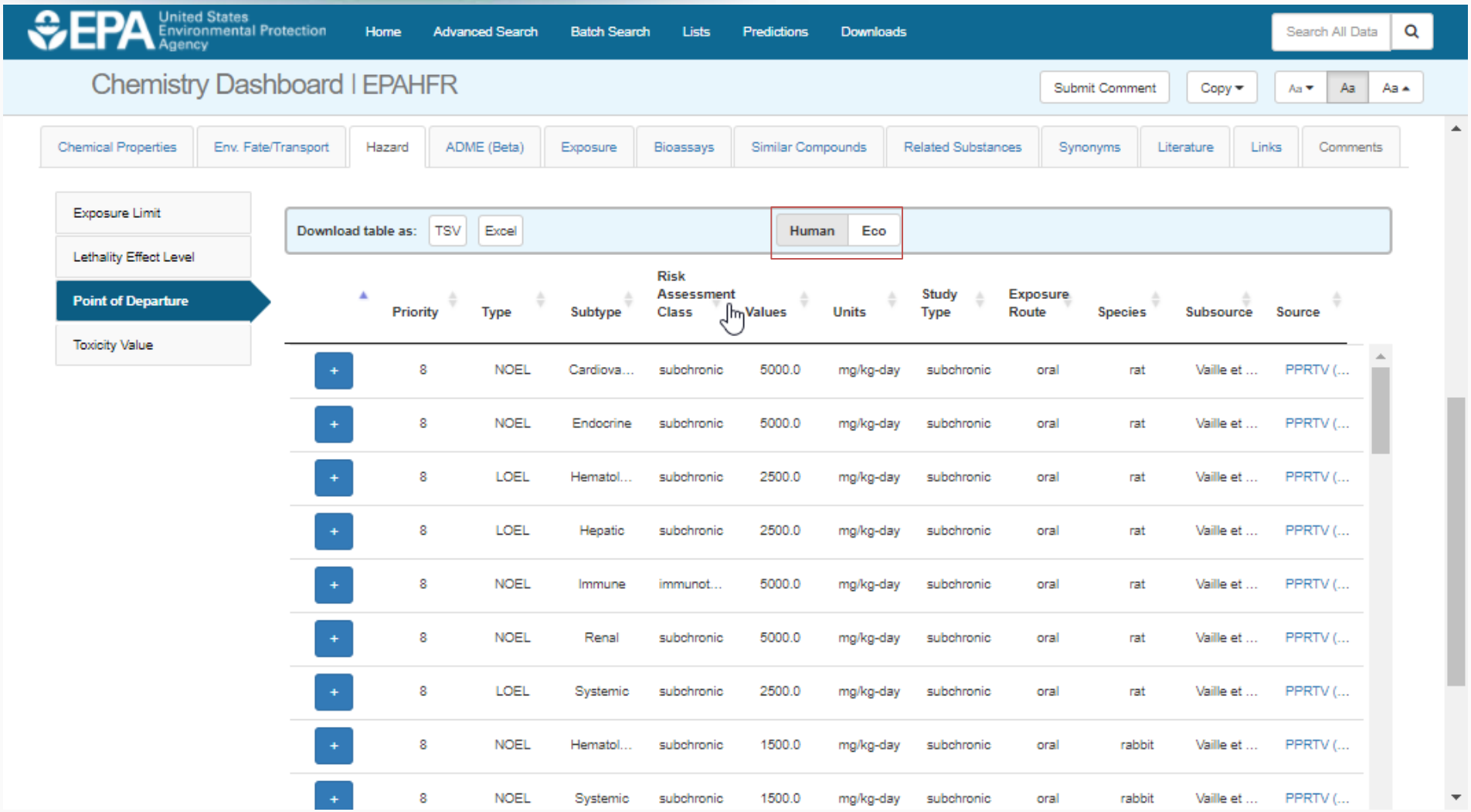
Browsing the List of Chemicals



Detailed Chemical Pages



Access to Chemical Hazard Data



	Priority	Type	Subtype	Risk Assessment Class	Values	Units	Study Type	Exposure Route	Species	Subsource	Source
+	8	NOEL	Cardiova...	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Endocrine	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Hematol...	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Hepatic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Immune	immunot...	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Renal	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Systemic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Hematol...	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaille et ...	PPRTV (...
+	8	NOEL	Systemic	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaille et ...	PPRTV (...

In Vitro Bioassay Screening

ToxCast and Tox21

Sources of Exposure to Chemicals

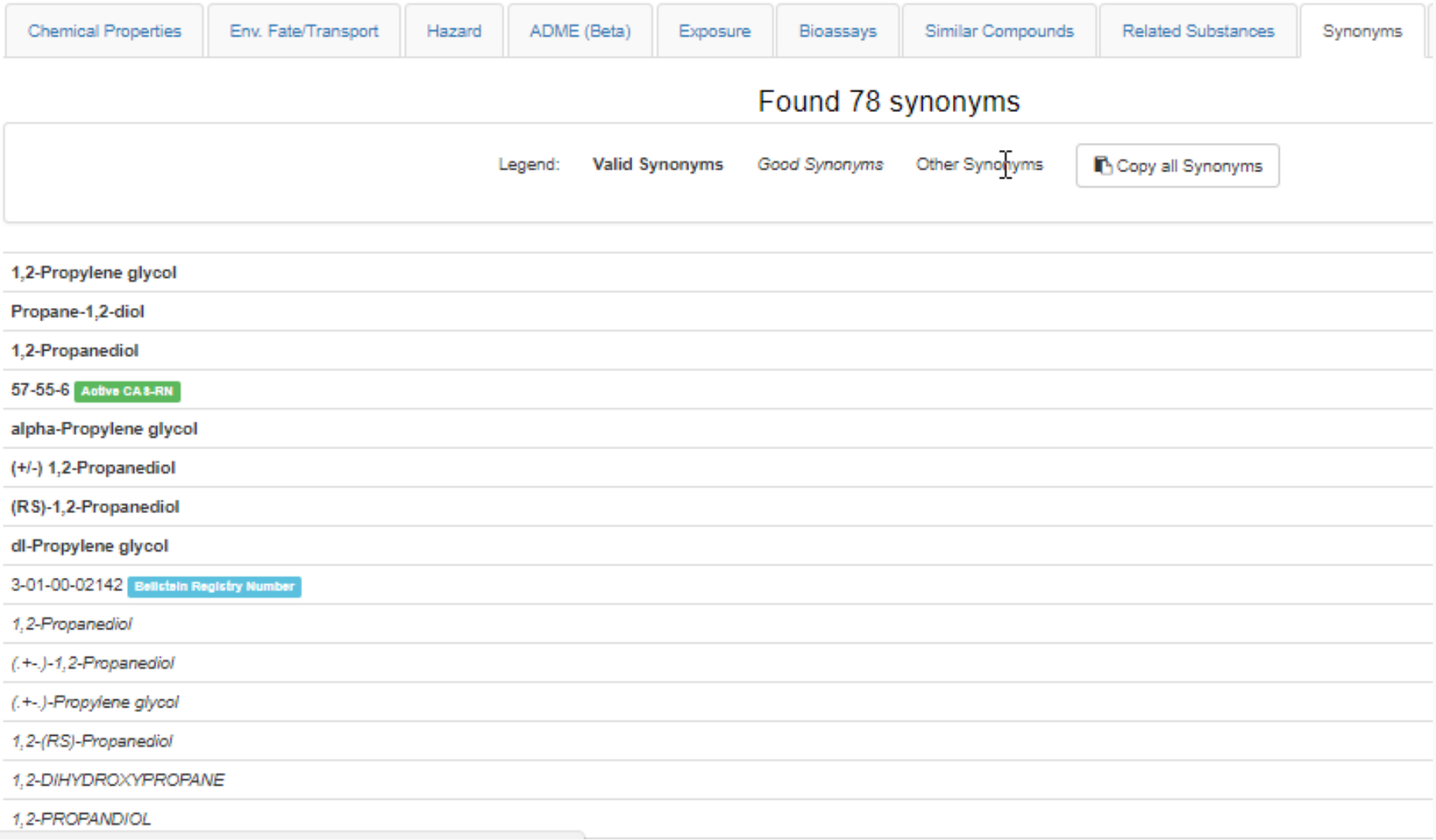
- **Product & Use Categories**
- Chemical Weight Fraction
- Chemical Functional Use
- Monitoring Data
- Exposure Predictions
- Production Volume

osure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

Product & Use Categories (PUCs)

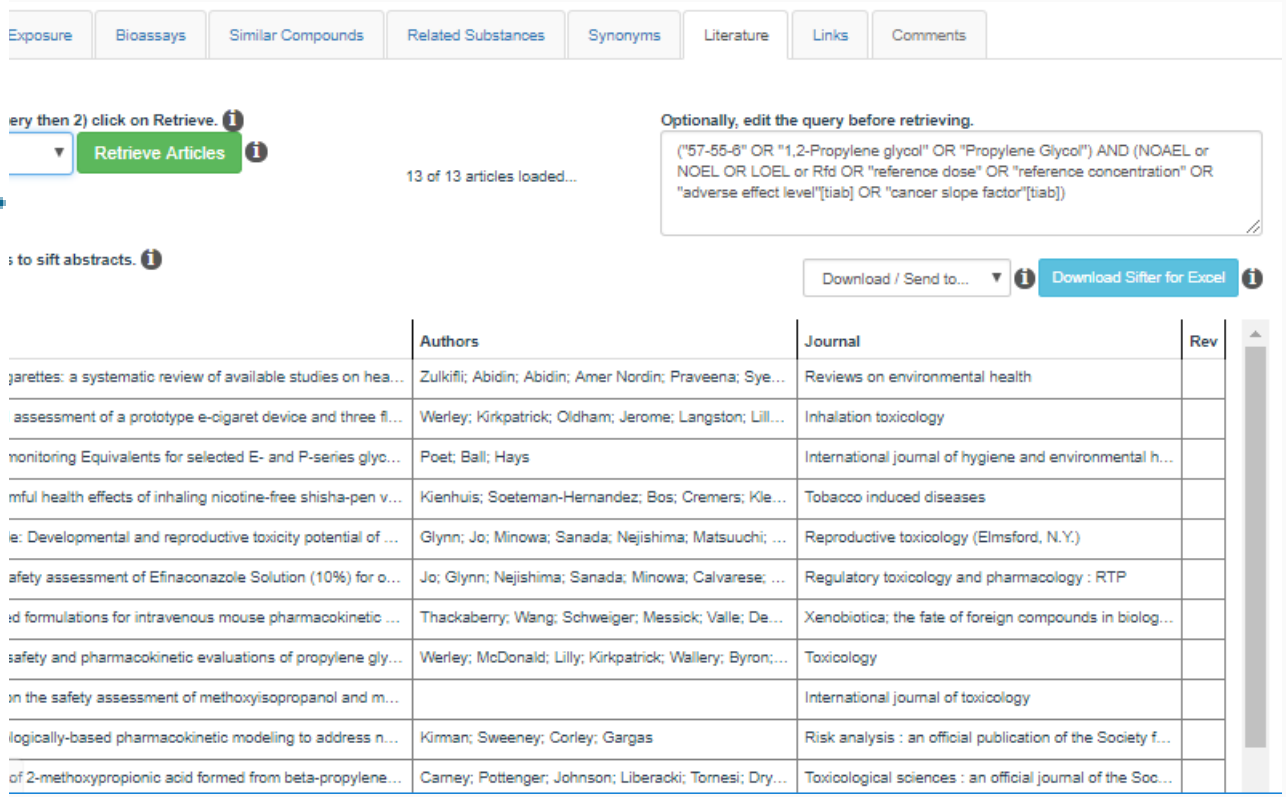
Categorization type	Number of Unique Products
PUC	288
PUC	206
PUC	117
PUC	107
PUC	107
PUC	101
PUC	101
PUC	90
PUC	89

Identifiers to Support Searches

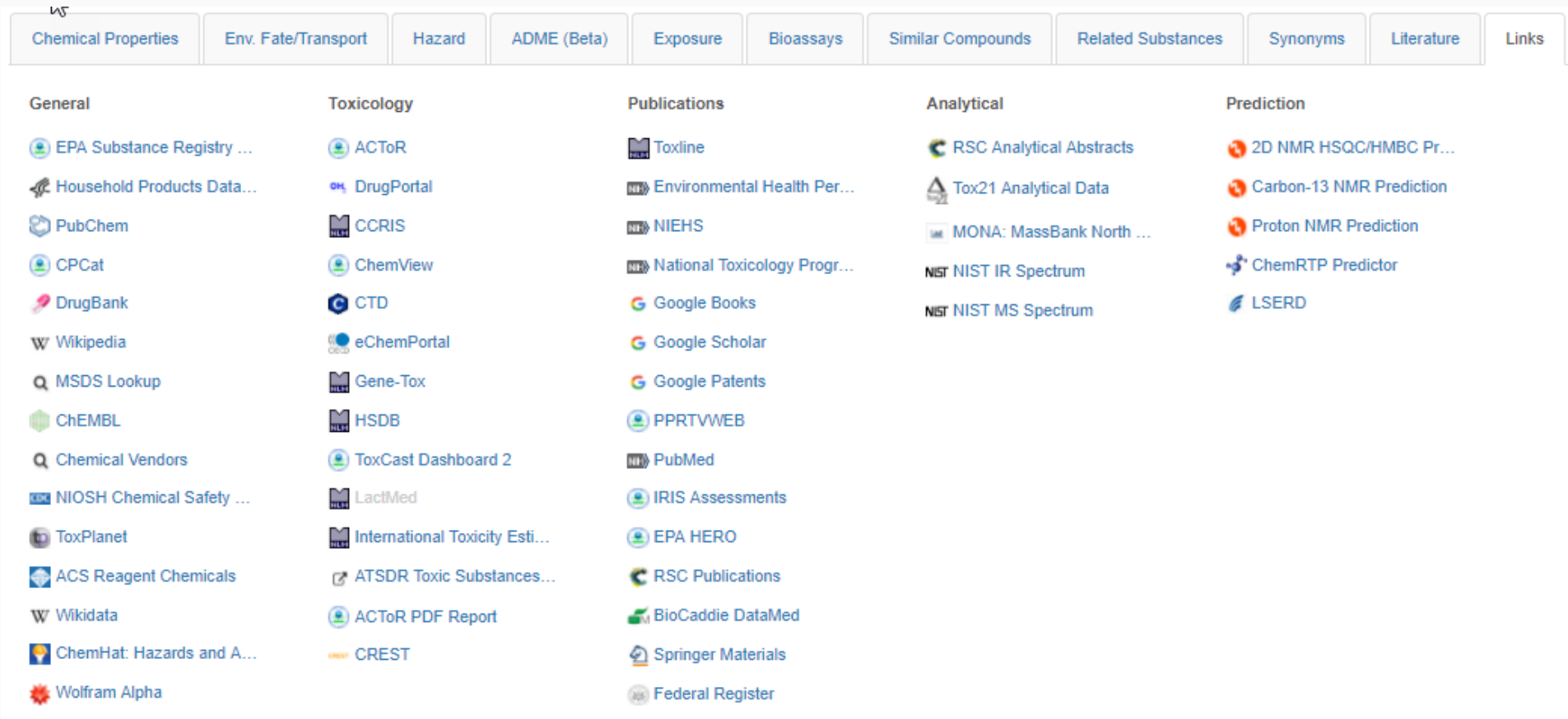


Literature Searches and Links

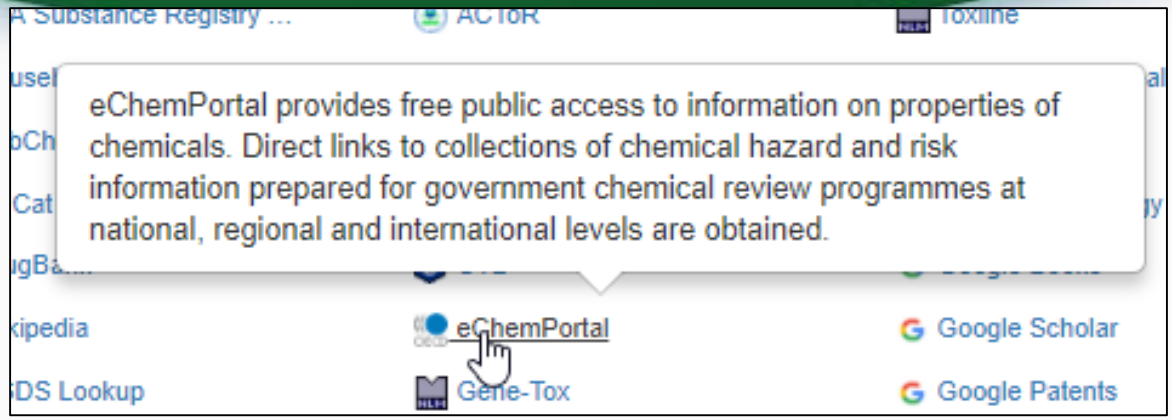
- Google Scholar
- **PubMed Abstract Sifter**
- PubChem Articles
- PubChem Patents
- PPRTV
- IRIS

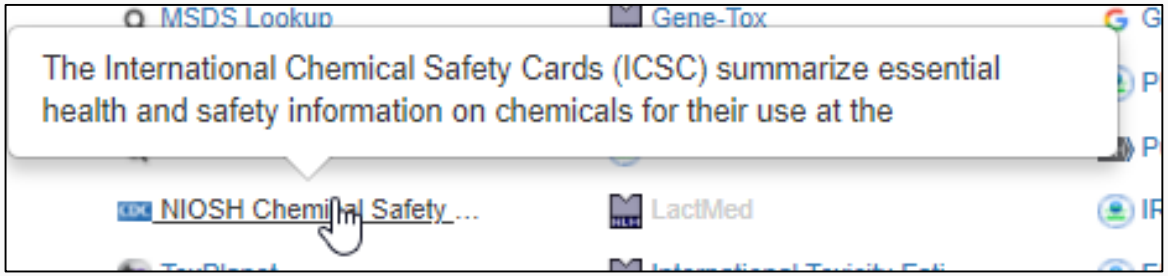


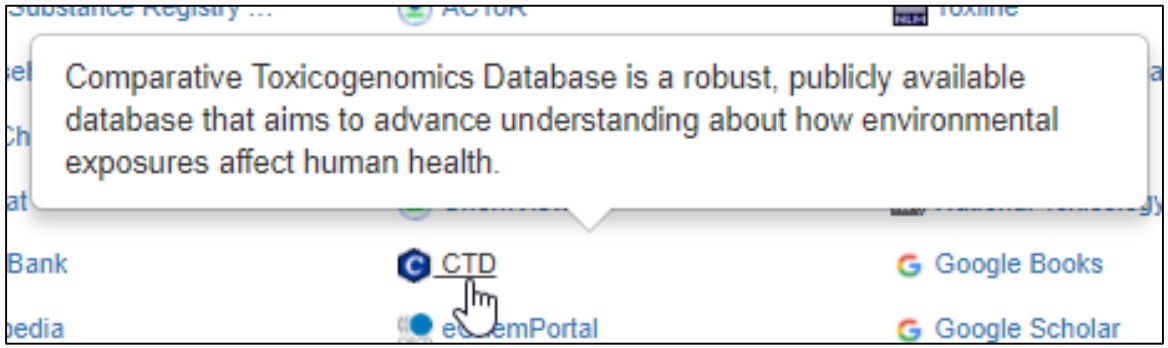
External Links to Data and Services



Integrated Linkouts



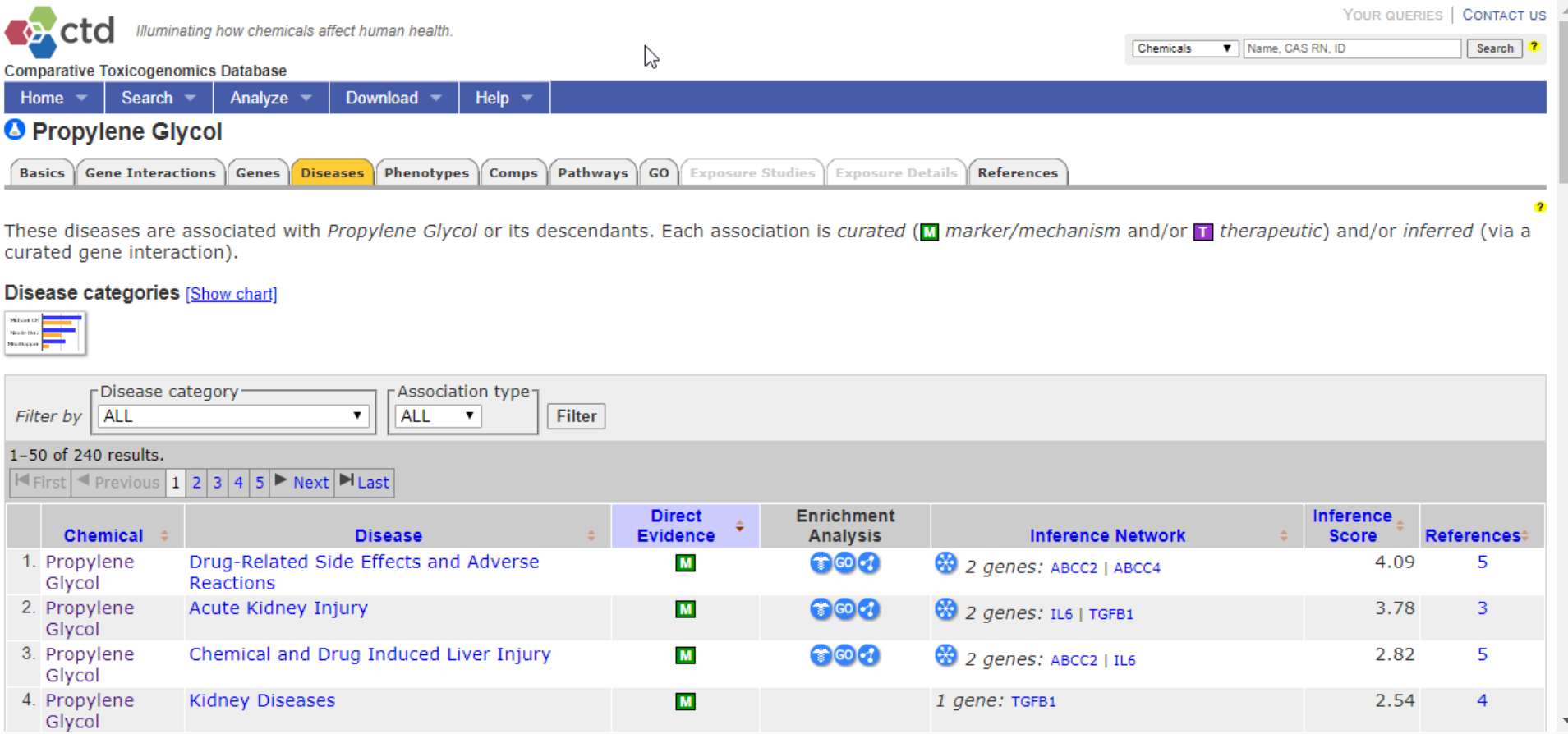




Integrated Linkouts

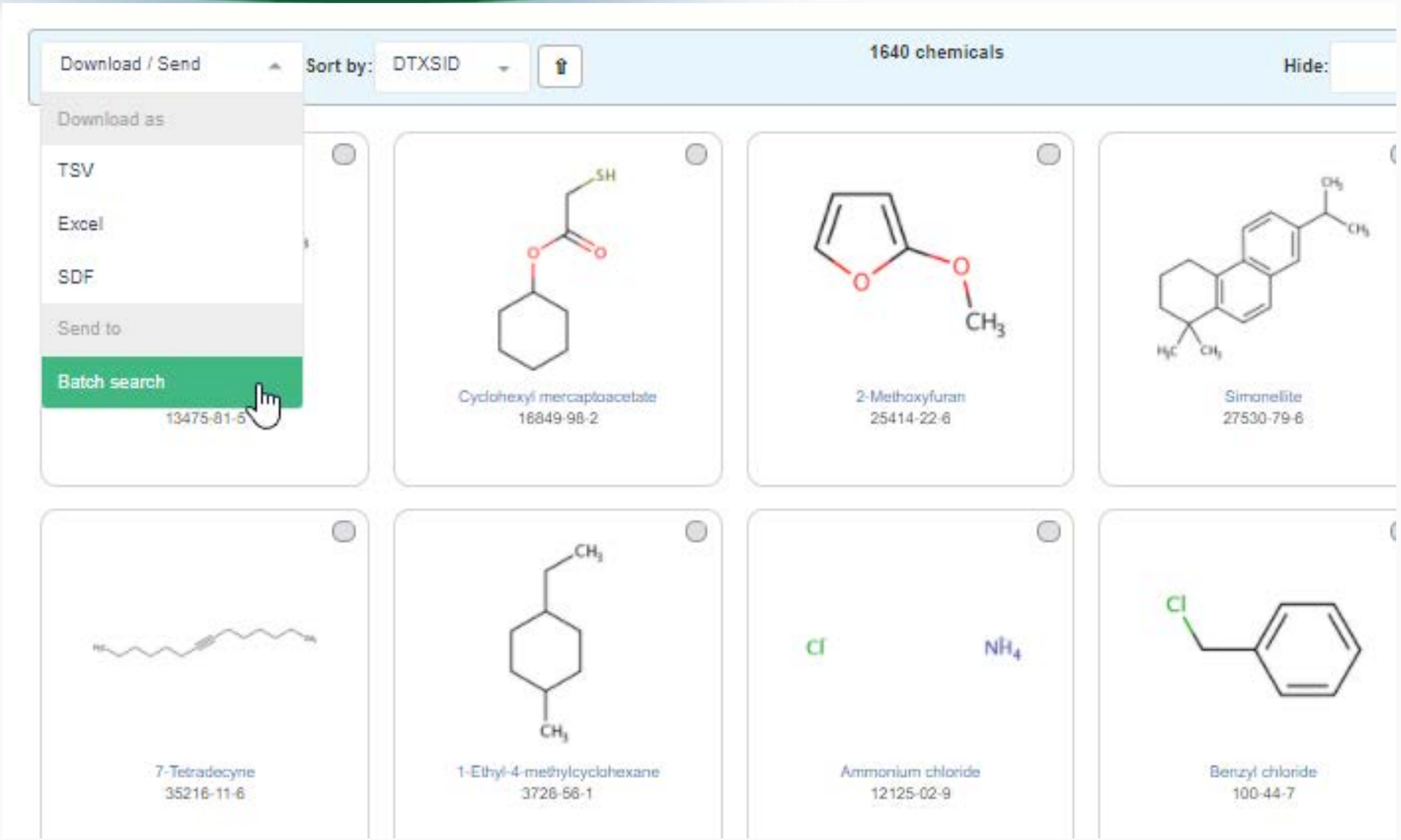
Comparative Toxicogenomics DB

EPA United States Environmental Protection Agency

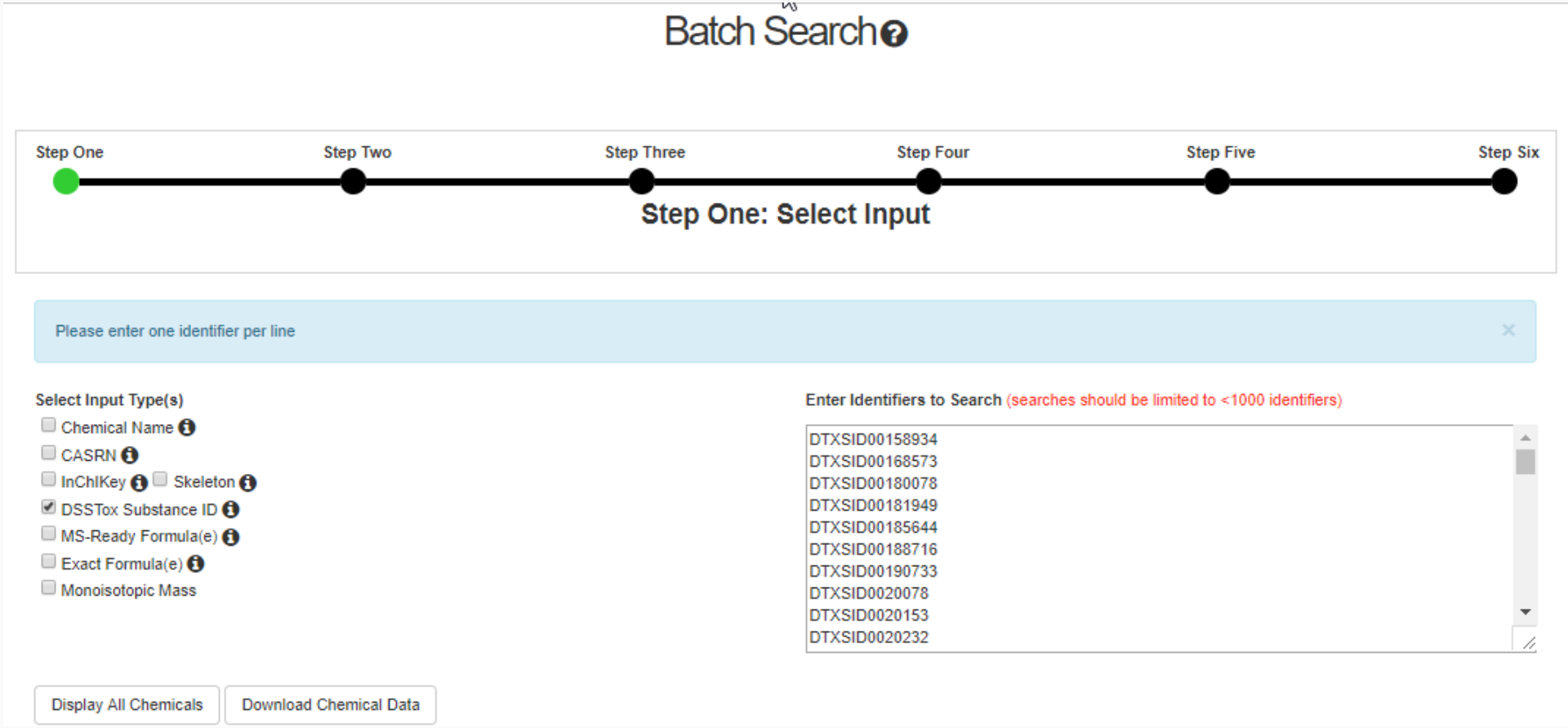


	Chemical	Disease	Direct Evidence	Enrichment Analysis	Inference Network	Inference Score	References
1.	Propylene Glycol	Drug-Related Side Effects and Adverse Reactions	M		2 genes: ABCC2 | ABCC4	4.09	5
2.	Propylene Glycol	Acute Kidney Injury	M		2 genes: IL6 | TGFB1	3.78	3
3.	Propylene Glycol	Chemical and Drug Induced Liver Injury	M		2 genes: ABCC2 | IL6	2.82	5
4.	Propylene Glycol	Kidney Diseases	M		1 gene: TGFB1	2.54	4

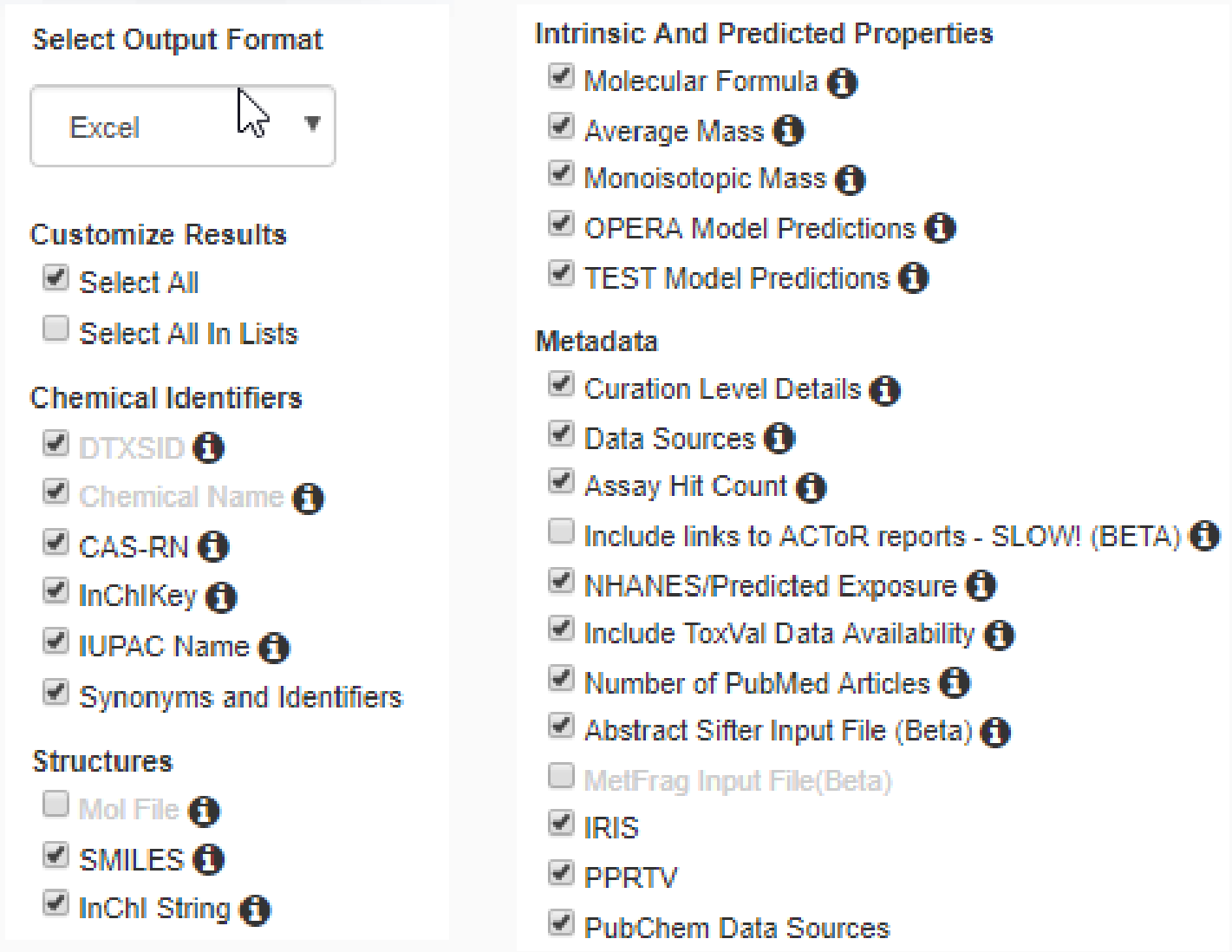
Batch Access to Dashboard Data



Batch Search Through Dashboard



Batch Search Through Dashboard

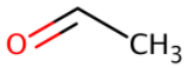
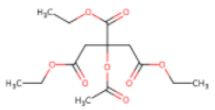
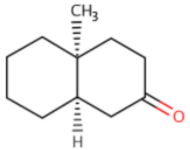
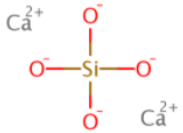
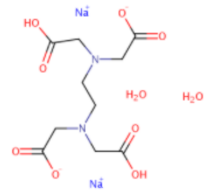


Batch Access to Dashboard Data

DTXSID	PREFERR	CASRN	INCHIKEY	IUPAC_NAME	MOLECUL	AVERAGE_	MONOISO	EXPOCAS	EXPOCAS	DATA_SOU	TOXVAL_D	TOXCAST_	TOXCAST_
DTXSID001	2,2,3,3-Tetr	13475-81-	RMQHJMM	2,2,3,3-Tetrameth	C10H22	142.286	142.17215	-	-	14	-	-	-
DTXSID001	Cyclohexyl	16849-98-	DHQYDHVI	Cyclohexyl sulfan	C8H14O2S	174.26	174.07145	-	-	6	-	-	-
DTXSID001	2-Methoxyfu	25414-22-	OXCGHDN	2-Methoxyfuran	C5H6O2	98.101	98.036779	-	-	16	-	-	-
DTXSID001	Simonellite	27530-79-	XZDCNNO	1,1-Dimethyl-7-(p	C19H24	252.401	252.1878	-	-	6	-	-	-
DTXSID001	4-Propylhep	3178-29-8	ABYGSZMC	4-Propylheptane	C10H22	142.286	142.17215	-	-	13	Y	-	-
DTXSID001	7-Tetradecy	35216-11-	AFNWSIIBA	Tetradec-7-yne	C14H26	194.362	194.20345	-	-	8	-	-	-
DTXSID001	1-Ethyl-4-m	3728-56-1	CYISMTMR	1-Ethyl-4-methylc	C9H18	126.243	126.14085	-	-	12	Y	-	-
DTXSID002	Ammonium	12125-02-	NLXLAEXVI	Ammonium chlori	ClH4N	53.49	53.003227	-	-	51	Y	-	-
DTXSID002	Benzyl chlo	100-44-7	KCXMKQU	(Chloromethyl)ber	C7H7Cl	126.58	126.02363	2.39e-06	Y	94	Y	2.65	3/113
DTXSID002	Caffeine	58-08-2	RYYVLZVU	1,3,7-Trimethyl-3,	C8H10N4O	194.194	194.08038	2.42e-05	Y	117	Y	3.85	21/546
DTXSID002	Ethylene ox	75-21-8	IAYPIBMAS	Oxirane	C2H4O	44.053	44.026215	-	-	85	Y	-	-
DTXSID002	Indole	120-72-9	SIKJAQJRH	1H-Indole	C8H7N	117.151	117.05785	9.05e-06	Y	82	Y	1.68	5/297
DTXSID002	Magnesium	7791-18-6	DHRRIBDT	Dichloridomagnes	Cl2H12MgO	203.3	201.98614	-	-	25	Y	2.65	3/113
DTXSID002	Dichlorome	75-09-2	YMWUJEA	Dichloromethane	CH2Cl2	84.93	83.953356	2.02e-06	Y	110	Y	0.88	1/113
DTXSID002	Sodium nitr	7632-00-0	LPXPTNMV	Sodium nitrite	NNaO2	68.995	68.982673	3.28e-05	Y	68	Y	0.56	3/537
DTXSID002	Ozone	10028-15-	CBENFWS	Trioxid-2-en-2-iun	O3	47.997	47.984744	-	-	49	Y	-	-
DTXSID002	Polysorbate	9005-65-6	-	-	-	-	-	0.000158	Y	54	Y	3.37	10/297
DTXSID002	Ethyl glycyl	627-74-7	LFAVEINQL	Ethyl glycylglycina	C6H12N2O	160.173	160.08479	-	-	4	-	-	-
DTXSID002	1,2-Propyle	57-55-6	DNIAPMSP	Propane-1,2-diol	C3H8O2	76.095	76.052429	6.29e-05	Y	110	Y	2.04	11/539
DTXSID002	Safrole	94-59-7	ZMQAAUB	5-(Prop-2-en-1-yl)	C10H10O2	162.188	162.06808	1.75e-07	Y	76	Y	1.47	8/546
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Many Hydraulic Fracturing Chemicals are “Complex”

EPA United States Environmental Protection Agency

 Acetaldehyde 75-07-0	0 related chemical structures with this substance Amines, dicoco alkyl 61789-76-2	0 related chemical structures with this substance Amines, coco alkyldimethyl 61788-93-0	1 related chemical structure with this substance Cristobalite 14464-46-1	 Copper(I) iodide 7681-65-4
0 related chemical structures with this substance Alcohols, C12-13, ethoxylated 66455-14-9	1 related chemical structure with this substance Alcohols, C12-15, ethoxylated 68131-39-5	0 related chemical structures with this substance Alcohols, C14-15, ethoxylated 68951-67-7	 Iron 7439-89-6	 Acetyltriethyl citrate 77-89-4
 (4aS,8aR)-4a-Methyloctahydronaphthal... 938-06-7	 Dicalcium silicate 10034-77-2	0 related chemical structures with this substance Diethylenetriamine reaction product wit... 68647-57-4	2 related chemical structures with this substance Di-sec-butylphenol 31291-60-8	 Disodium ethylenediaminetetraacetate ... 6381-92-6

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

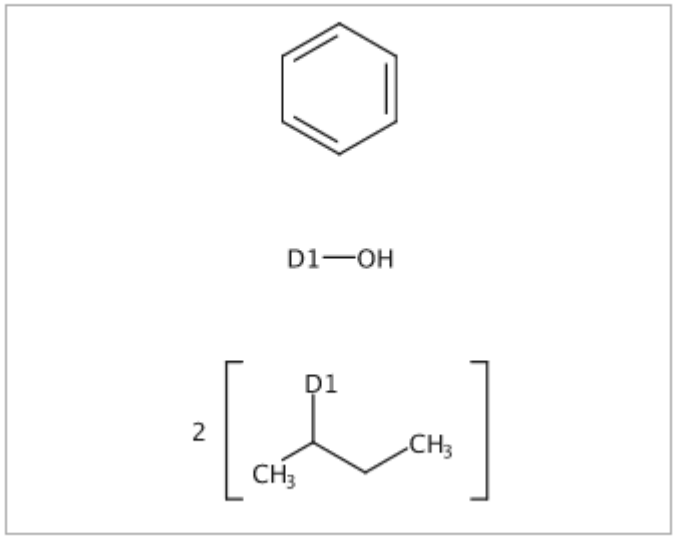
This paper is a compendium of information related to the broad class of chemical substances referred to as UVCBs for the Toxic Substances Control Act (TSCA) Chemical Substance Inventory. These chemical substances cannot be represented by unique structures and molecular formulas.

Di-sec-butylphenol

CAS Representation

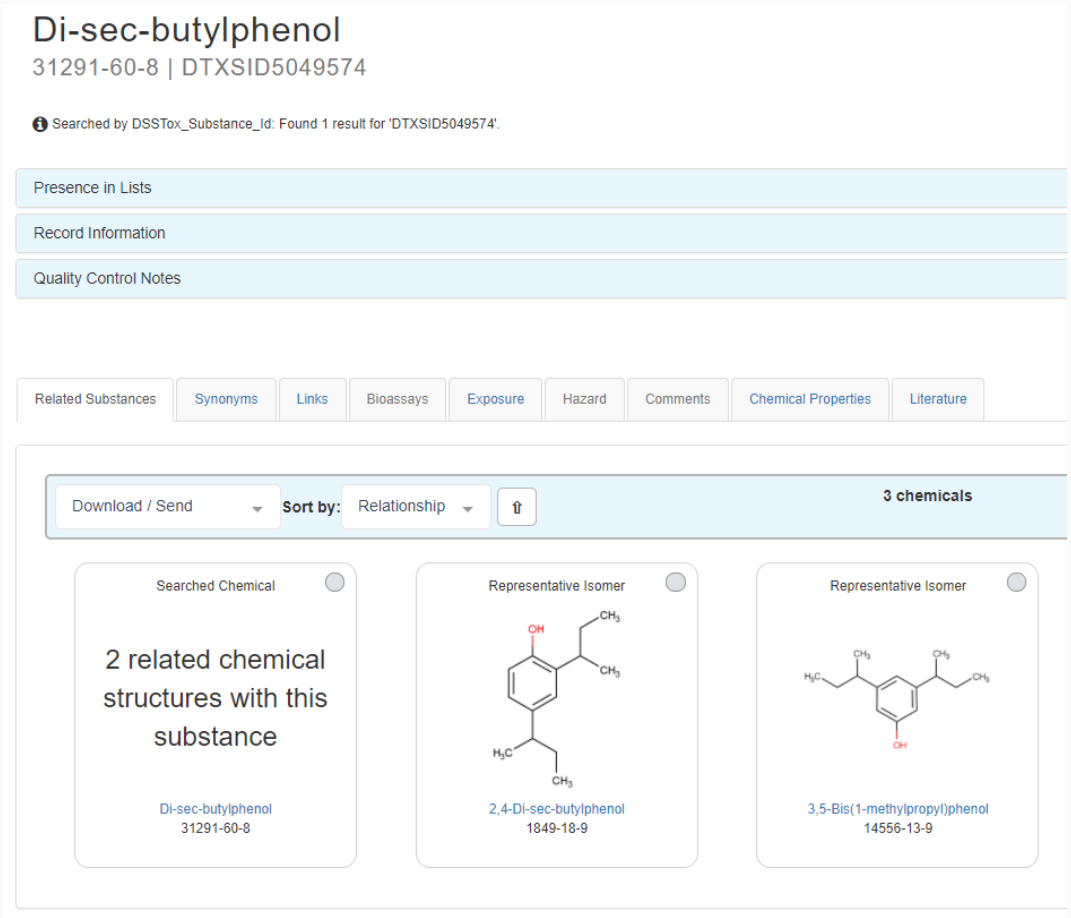
1. 31291-60-8

~31 ~4



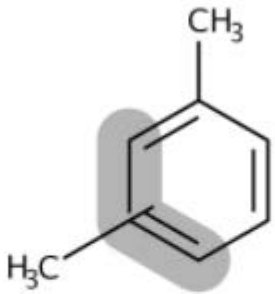
$C_{14}H_{22}O$
Phenol, bis(1-methylpropyl)-

Dashboard Representation

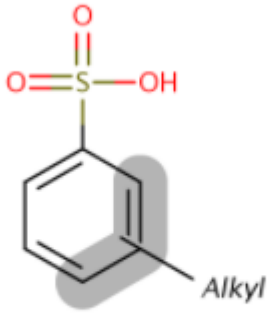


“Markush Structures”

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

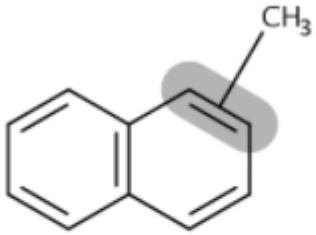


Xylenes
1330-20-7

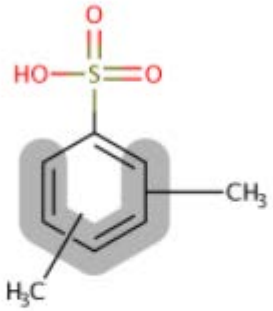


(C10-C16) Alkylbenzenesulfonic acid
68584-22-5

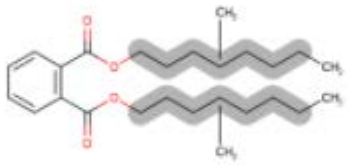
n-Nonylphenol
25154-52-3



Methylnaphthalene
1321-94-4



Sodium xylenesulfonate
1300-72-7



Diisononyl phthalate
28553-12-0

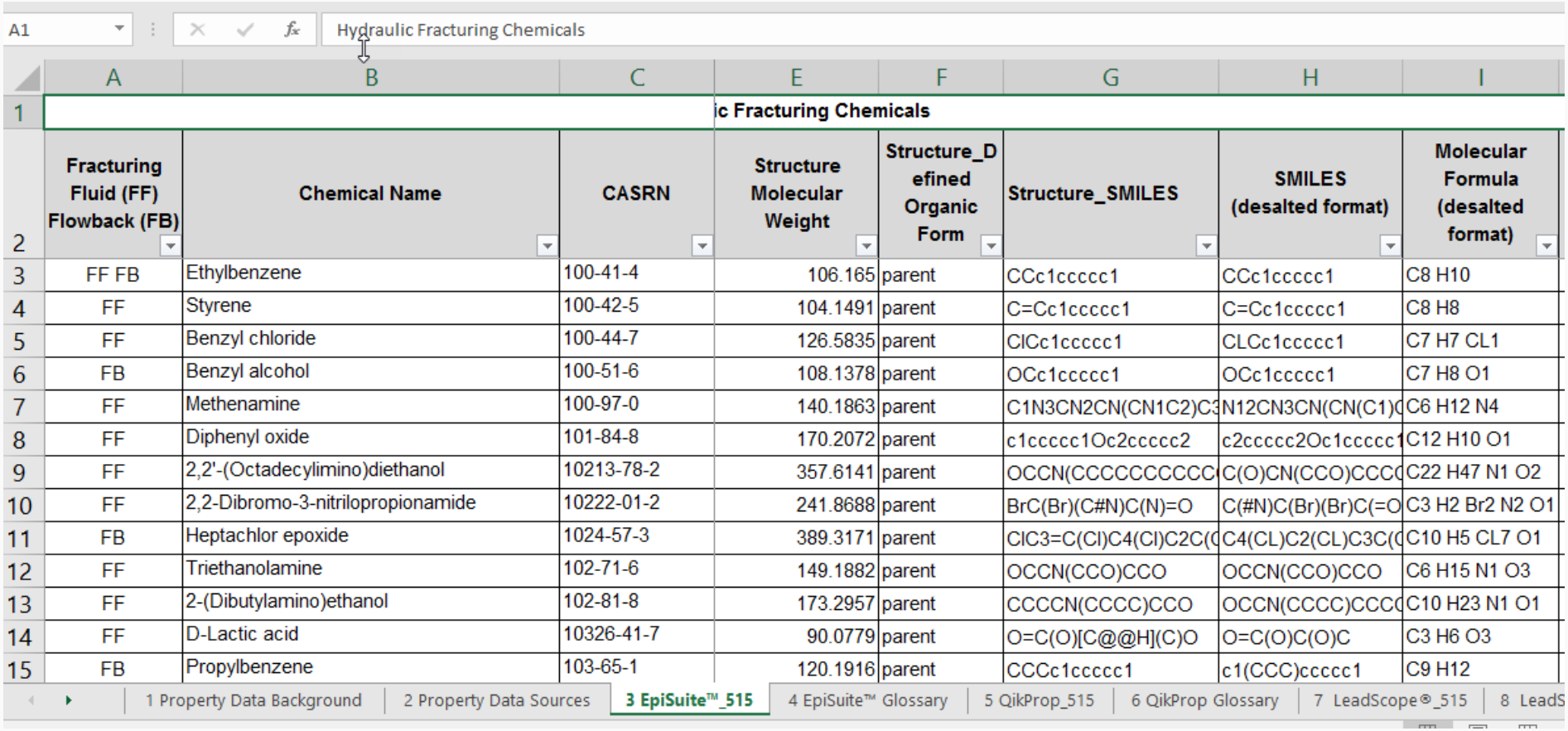
AFTER CURATION
1640 Chemicals

- 528/1640 chemicals **without** explicit structures
- 483/1640 chemicals examined in **ToxCast** – 94 not active in any assay
- 1036/1640 chemicals with **Hazard Data**
- 141/1640 chemicals with **IRIS Assessment** Reports (Integrated Risk Information System)

Curated Database for Download

https://www.epa.gov/hfstudy

Download Files from EPA Site



A1 | Hydraulic Fracturing Chemicals

	A	B	C	E	F	G	H	I
1				ic Fracturing Chemicals				
2	Fracturing Fluid (FF) Flowback (FB)	Chemical Name	CASRN	Structure Molecular Weight	Structure_Defined Organic Form	Structure_SMILES	SMILES (desalted format)	Molecular Formula (desalted format)
3	FF FB	Ethylbenzene	100-41-4	106.165	parent	CCc1ccccc1	CCc1ccccc1	C8 H10
4	FF	Styrene	100-42-5	104.1491	parent	C=Cc1ccccc1	C=Cc1ccccc1	C8 H8
5	FF	Benzyl chloride	100-44-7	126.5835	parent	ClCc1ccccc1	CLCc1ccccc1	C7 H7 CL1
6	FB	Benzyl alcohol	100-51-6	108.1378	parent	OCc1ccccc1	OCc1ccccc1	C7 H8 O1
7	FF	Methenamine	100-97-0	140.1863	parent	C1N3CN2CN(CN1C2)C3	N12CN3CN(CN(C1)C	C6 H12 N4
8	FF	Diphenyl oxide	101-84-8	170.2072	parent	c1ccccc1Oc2ccccc2	c2ccccc2Oc1ccccc1	C12 H10 O1
9	FF	2,2'-(Octadecylimino)diethanol	10213-78-2	357.6141	parent	OCCN(CCCCCCCCCC(	C(O)CN(CCO)CCC(	C22 H47 N1 O2
10	FF	2,2-Dibromo-3-nitrilopropionamide	10222-01-2	241.8688	parent	BrC(Br)(C#N)C(N)=O	C(#N)C(Br)(Br)C(=O	C3 H2 Br2 N2 O1
11	FB	Heptachlor epoxide	1024-57-3	389.3171	parent	ClC3=C(Cl)C4(Cl)C2C(	C4(CL)C2(CL)C3C(	C10 H5 CL7 O1
12	FF	Triethanolamine	102-71-6	149.1882	parent	OCCN(CCO)CCO	OCCN(CCO)CCO	C6 H15 N1 O3
13	FF	2-(Dibutylamino)ethanol	102-81-8	173.2957	parent	CCCCN(CCCC)CCO	OCCN(CCCC)CCC(	C10 H23 N1 O1
14	FF	D-Lactic acid	10326-41-7	90.0779	parent	O=C(O)[C@@H](C)O	O=C(O)C(O)C	C3 H6 O3
15	FB	Propylbenzene	103-65-1	120.1916	parent	CCCc1ccccc1	c1(CCC)ccccc1	C9 H12

1 Property Data Background | 2 Property Data Sources | 3 EpiSuite™_515 | 4 EpiSuite™ Glossary | 5 QikProp_515 | 6 QikProp Glossary | 7 LeadScope®_515 | 8 LeadS

Suspect Screening and Non-Targeted Analysis Workflow

EPA United States Environmental Protection Agency

Suspect Screening

- Raw Samples
- Extracted Samples
- Raw Features
- "Molecular Features"
- DSSTox Chemical Database
- Matched Formulas
- Mapped Structures
- Prioritized Structures (using ToxPi)
- Confirmed Structures (using ToxCast standards)
- Predicted Concentrations

Non-Targeted Analysis

- Processed Features
- Prioritized Features
- Predicted Formulas
- Candidate Structures
- Sorted Structures
- Predicted Retention Times
- Predicted Mass Spectra
- Predicted/Observed Functional Use
- Predicted/Observed Media Occurrence
- Methodological Concordance
- Top Candidate Structure(s)

Color Key

Red = Analytical Chemistry

Blue = Data Processing & Analysis

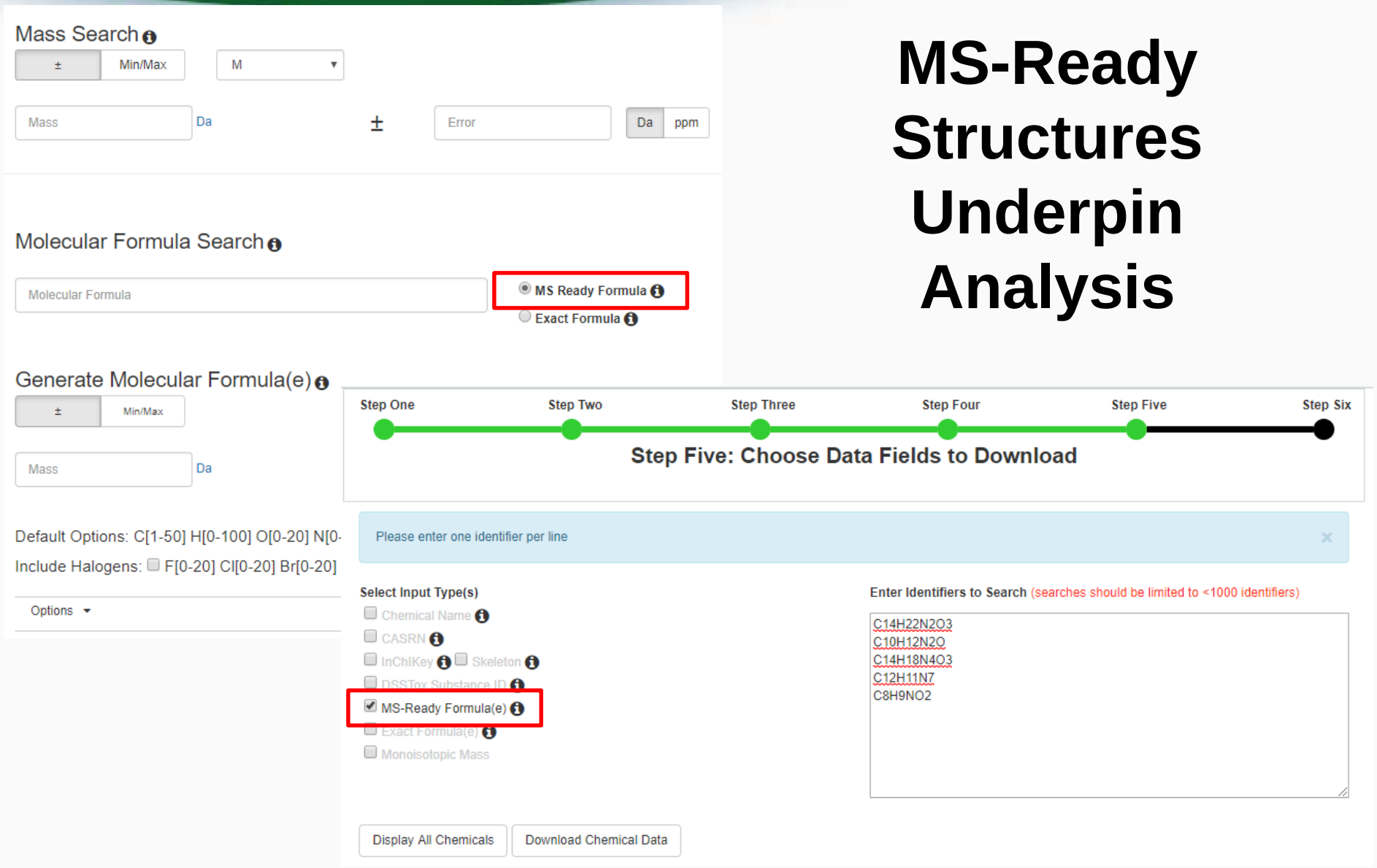
Purple = Mathematical & QSPR Modeling

Green = Informatics & Web Services

Content from J. Sobus, US-EPA-NERL

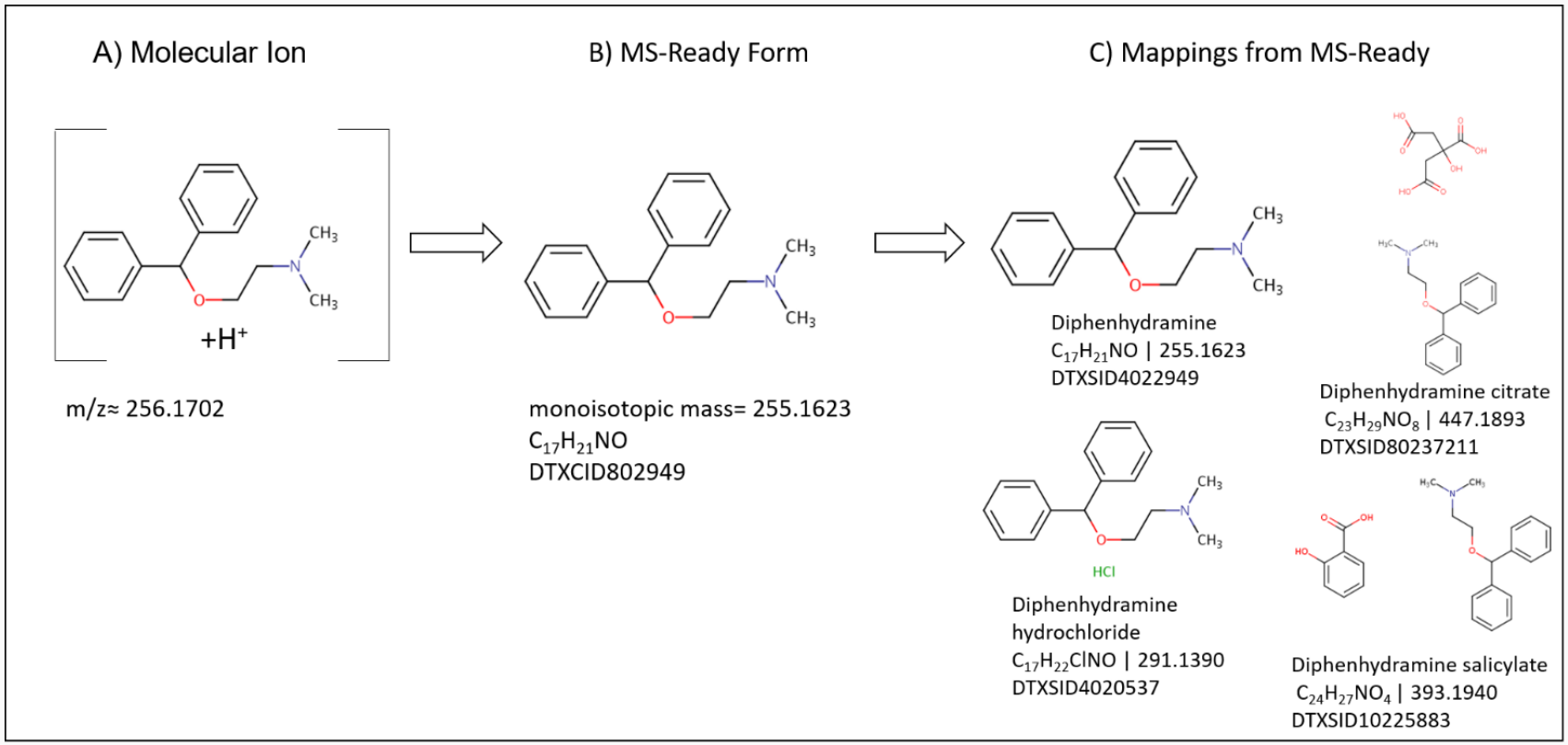
The Dashboard to Support MS-Analysis

EPA United States Environmental Protection Agency



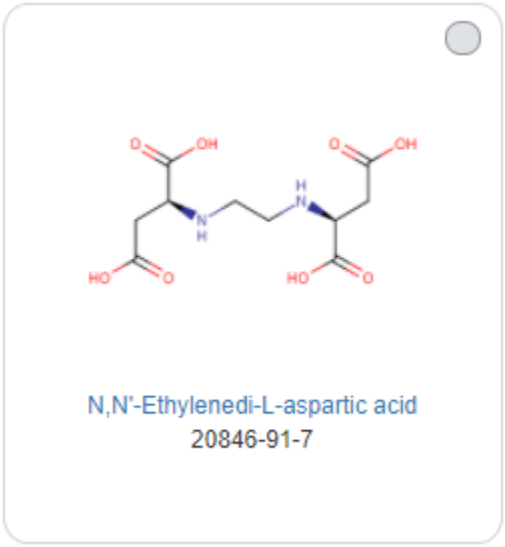
MS-Ready Structures Underpin Analysis

Specific Data-Mappings "MS-Ready Structures"



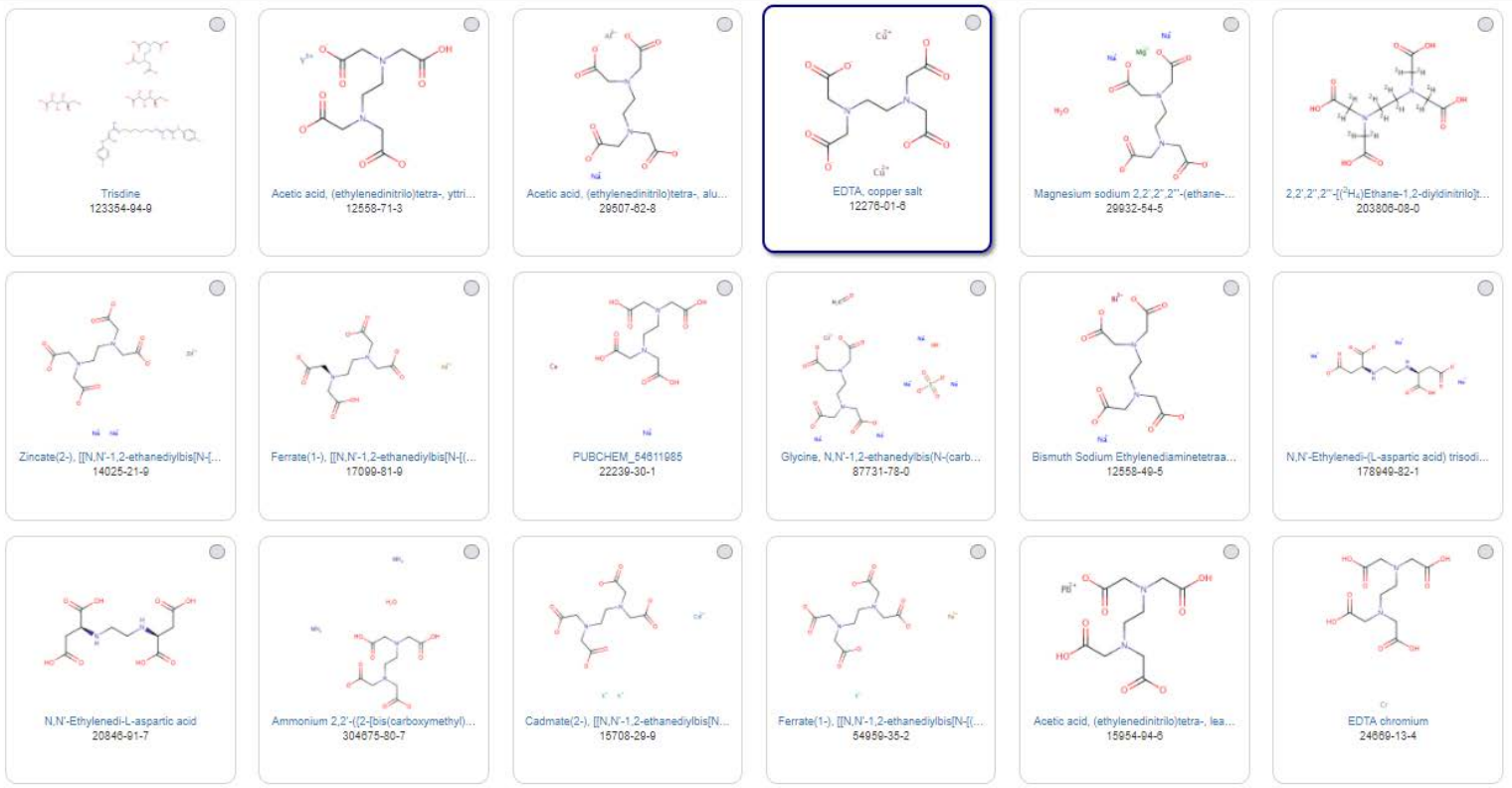
MS-Ready Mappings

- Input Formula: C10H16N2O8
- EXACT Formula Search: 3 Chemicals



MS-Ready Mappings

- **Same** Input Formula: C10H16N2O8
- MS Ready Formula Search: 88 Chemicals



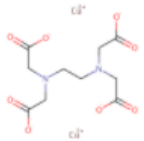
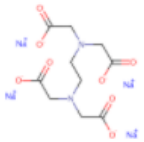
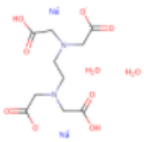
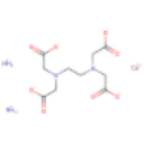
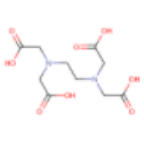
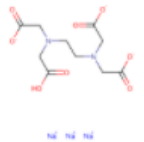
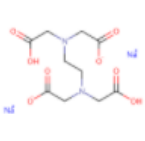
Complexity to Simplicity
88 Chemicals – 7 in EPAHFR

EPA United States Environmental Protection Agency

	INPUT	DTXCID_INDIVIDI	FORMULA_	SMILES_I	DTXSID	CASRN	EXPOCAST_M	EXPOCAST	DATA_SOURCI	TOXVAL_D	TOXCAST_	TOXCAST_	# OF_PUBMED	PUBCHEM	EPAHFR
2	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6022977	60-00-4	7.96e-05	Y	71	Y	2.65	3/113	25251	158	Y
3	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9027073	139-33-3	-	-	41	Y	-	-	25251	56	Y
4	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3026350	64-02-8	-	-	37	Y	-	-	-	57	Y
5	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7020556	150-38-9	-	-	30	Y	-	-	-	33	Y
6	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5049609	67989-88-2	-	-	20	Y	-	-	-	8	Y
7	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5049576	6381-92-6	-	-	19	Y	-	-	25251	31	Y
8	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID0034564	12276-01-6	-	-	11	-	-	-	-	8	Y
9	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5027774	15708-41-5	-	-	48	Y	1.98	6/303	241	53	-
10	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2036409	62-33-9	4.64e-06	Y	37	Y	0.0	0/64	25251	42	-
11	C10H16N2O8	DTXCID00197424	C10H16N2(	OC(=O)C(	DTXSID1051852	20846-91-7	-	-	36	Y	-	-	89	25	-
12	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6042107	15375-84-5	-	-	25	Y	-	-	97	25	-
13	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3036442	5964-35-2	-	-	23	Y	-	-	25251	25	-
14	C10H16N2O8	DTXCID00197424	C10H16N2(	OC(=O)C(	DTXSID1051806	178949-82-1	-	-	22	Y	-	-	-	5	-
15	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID0065696	14025-21-9	-	-	22	Y	-	-	-	43	-
16	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9027813	20824-56-0	-	-	21	Y	-	-	-	12	-
17	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9027815	21265-50-9	-	-	20	Y	-	-	241	24	-
18	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5058272	17421-79-3	-	-	19	Y	-	-	25251	25	-
19	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3058612	2001-94-7	-	-	18	Y	-	-	25251	19	-
20	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID8027820	22473-78-5	-	-	16	Y	-	-	-	11	-
21	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID8058324	17572-97-3	-	-	15	-	-	-	-	36	-
22	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID8028343	67859-51-2	-	-	14	Y	-	-	-	5	-
23	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID4051328	13235-36-4	-	-	14	-	-	-	-	18	-
24	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6070980	68015-77-0	-	-	14	Y	-	-	-	13	-
25	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID9058317	15934-01-7	-	-	11	-	-	-	-	5	-
26	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID0066163	17099-81-9	-	-	11	-	-	-	241	14	-
27	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID1068988	54959-35-2	-	-	11	-	-	-	241	14	-
28	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5074266	60816-63-9	-	-	11	-	-	-	1	10	-
29	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID4048197	39208-15-6	-	-	10	-	-	-	-	28	-
30	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2065830	14931-83-0	-	-	10	-	-	-	47	9	-
31	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID70189997	36499-65-7	-	-	10	-	-	-	25298	26	-
32	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7051420	61916-40-3	-	-	9	-	-	-	-	4	-
33	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2051425	73513-47-0	-	-	8	Y	-	-	-	3	-
34	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7051426	73637-19-1	-	-	8	Y	-	-	-	5	-
35	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID2051427	73637-20-4	-	-	8	Y	-	-	-	-	-
36	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID3058741	10378-23-1	-	-	8	Y	-	-	-	31	-
37	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6065925	15708-48-2	-	-	8	-	-	-	-	19	-
38	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID20217976	6766-87-6	-	-	8	-	-	-	-	13	-
39	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID5065807	14689-29-3	-	-	7	-	-	-	-	12	-
40	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID6069408	60544-70-9	-	-	7	-	-	-	-	12	-
41	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID00153984	123354-94-9	-	-	7	-	-	-	2	6	-
42	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID70190705	37209-61-3	-	-	7	-	-	-	6	9	-
43	C10H16N2O8	DTXCID902977	C10H16N2(	OC(=O)CI	DTXSID7051424	67401-50-7	-	-	6	-	-	-	-	4	-

Complexity to Simplicity
88 Chemicals – 7 in the list

EPA United States Environmental Protection Agency

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Source	PubMed Data Source	Monoisotopic Mass
	DTXSID0034564	EDTA, copper salt	12276-01-6	Level 1	10	11	8	0	413.918561
	DTXSID3026350	Ethylenediaminetetraacetic acid tetrasodium salt	64-02-8	Level 1	1227	37	57	0	380.018442
	DTXSID5049576	Disodium ethylenediaminetetraacetate dihydrate	6381-92-6	Level 1	93	19	31	25251	372.075683
	DTXSID5049609	Ethylenediaminetetraacetic acid, diammonium copper salt	67989-88-2	Level 2	9	20	8	0	387.057712
	DTXSID6022977	Ethylenediaminetetraacetic acid	60-00-4	Level 1	346	71	158	25251	292.090665
	DTXSID7020556	Trisodium ethylenediaminetetraacetate	150-38-9	Level 1	85	30	33	0	358.036498
	DTXSID9027073	Ethylenediaminetetraacetic acid, disodium salt	139-33-3	Level 1	1358	41	56	25251	336.054554

Searching batches
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	DTXSID6025667	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	50

Conclusions

- The CompTox Chemistry Dashboard is useful for delivering access to lists
  - The EPA Hydraulic Fracturing Chemical List
  - The ToxCast Bioassay Screening List
  - NIOSH International Chemical Safety Cards
  - Algal Toxins
- Lists can be updated with new chemicals

Conclusions

- Registering chemical lists gives access to:
  - Human and Ecological Hazard data
  - ToxCast/Tox21 *in vitro* screening data
  - Exposure data – consumer products & functional use
  - Physicochemical and environmental fate and transport experimental and predicted data
  - Integrated mappings into multiple public sites
- Curation is time-consuming, iterative and high-value

Acknowledgments

- Hydraulic Fracturing Drinking Water Study Team
- The CompTox Chemistry Dashboard team
- Todd Martin (TEST Predictions), US-EPA-NRMRL
- NERL colleagues:
  - Jon Sobus, Elin Ulrich, Mark Strynar, Seth Newton (NTA Analysis)
  - Katherine Phillips, Kathie Dionisio, Kristin Isaacs (Consumer Products Database)
- Emma Schymanski – Luxembourg Center for Systems Biomedicine (MS-ready/NTA)

Contact

Antony Williams

US EPA Office of Research and Development

National Center for Computational Toxicology (NCCT)

Williams.Antony@epa.gov

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