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

Informatics Approaches supporting UVCB Chemicals in the US-EPA CompTox Chemicals Dashboard

Antony Williams[1], *Chris Grulke[1] and Emma Schymanski[2]*

Center for Computational Toxicology and Exposure, U.S. Environmental Protection Agency, RTP, NC

Luxembourg Center for Systems Biomedicine, University of Luxembourg, Luxembourg, Europe

November 2019

SETAC, Toronto, Canada

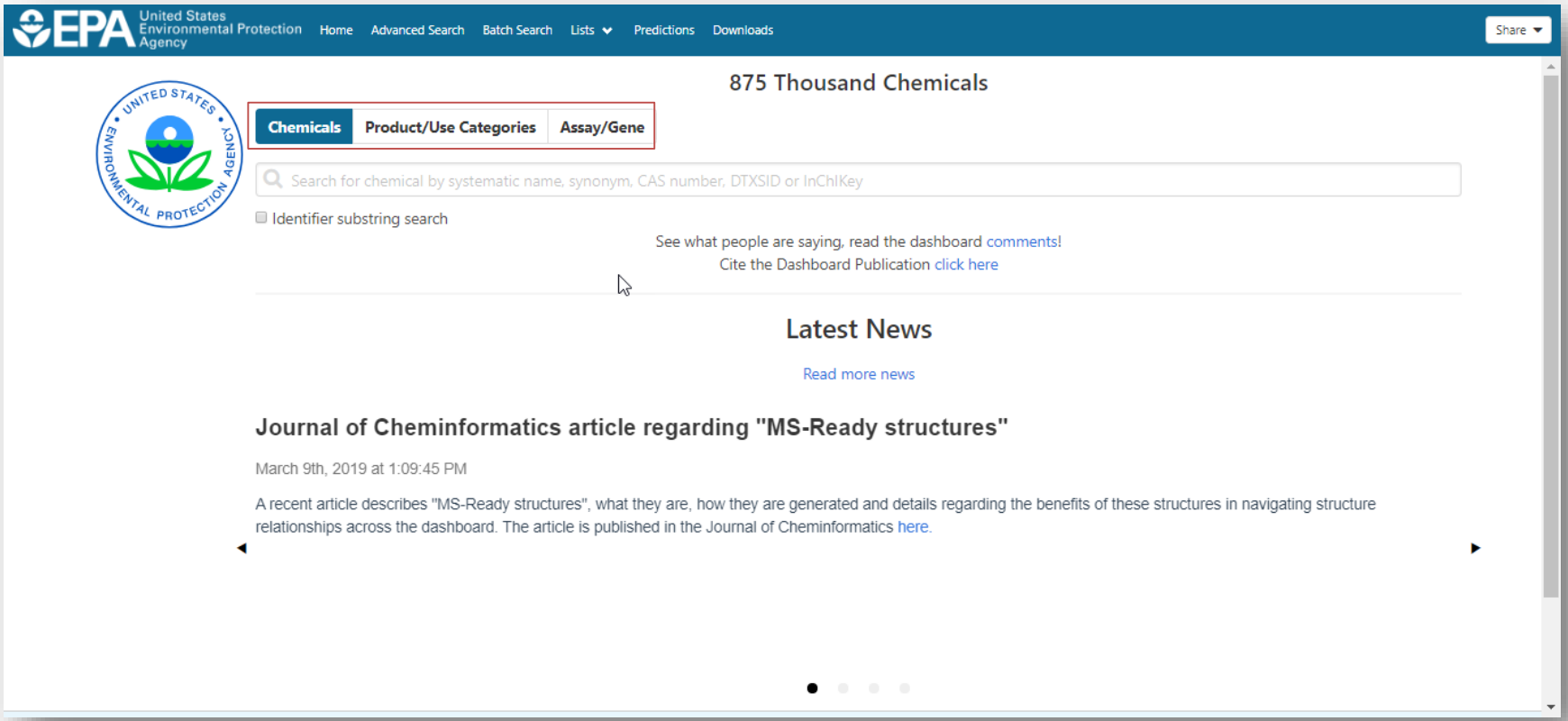
Overview

- The CompTox Chemicals Dashboard - web-based database of 875k substances
- Associated data including:
  - Experimental and predicted physicochemical data
  - *In vivo* hazard data
  - *In vitro* bioactivity screening data
  - Link farm to tens of public resources
- Integrated modules – read-across, lit search
- Data mappings and searches supporting Mass Spectrometry & structure identification

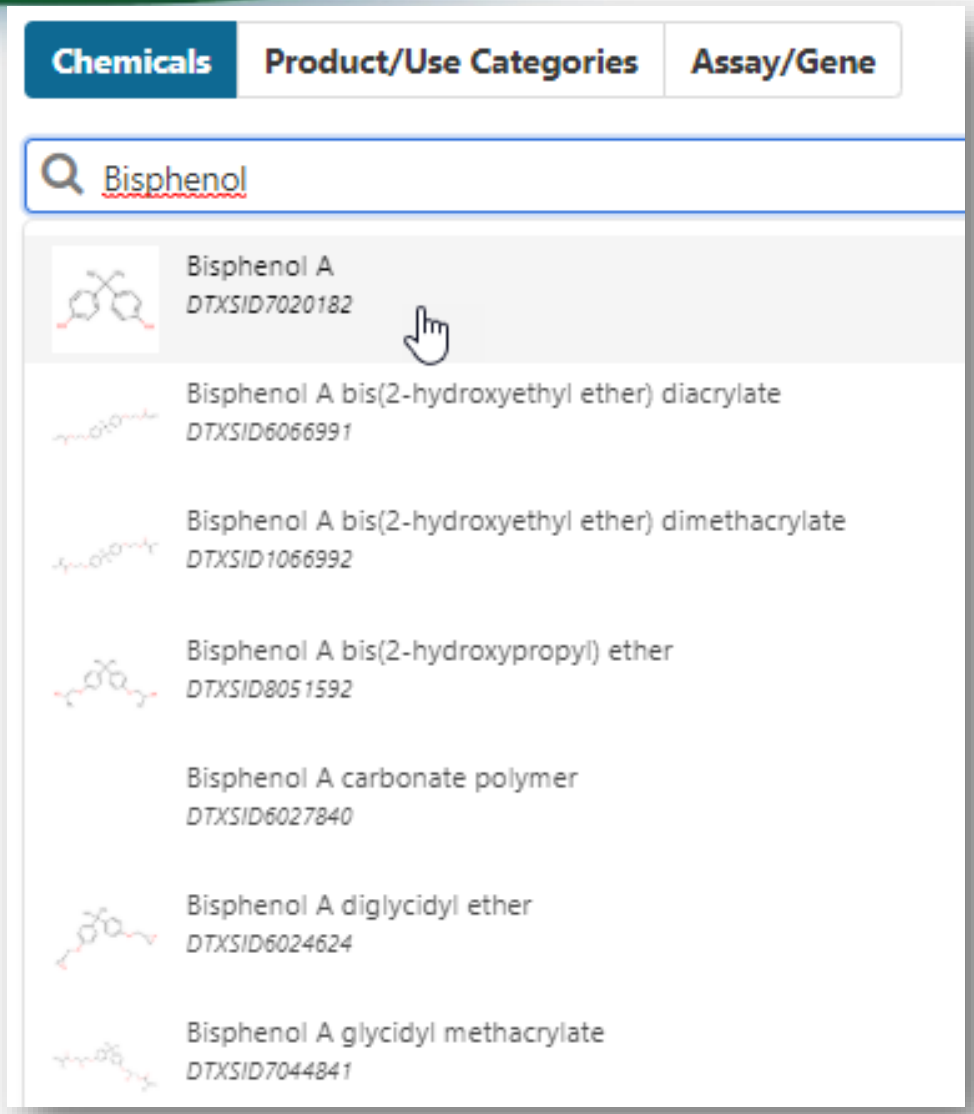
CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard

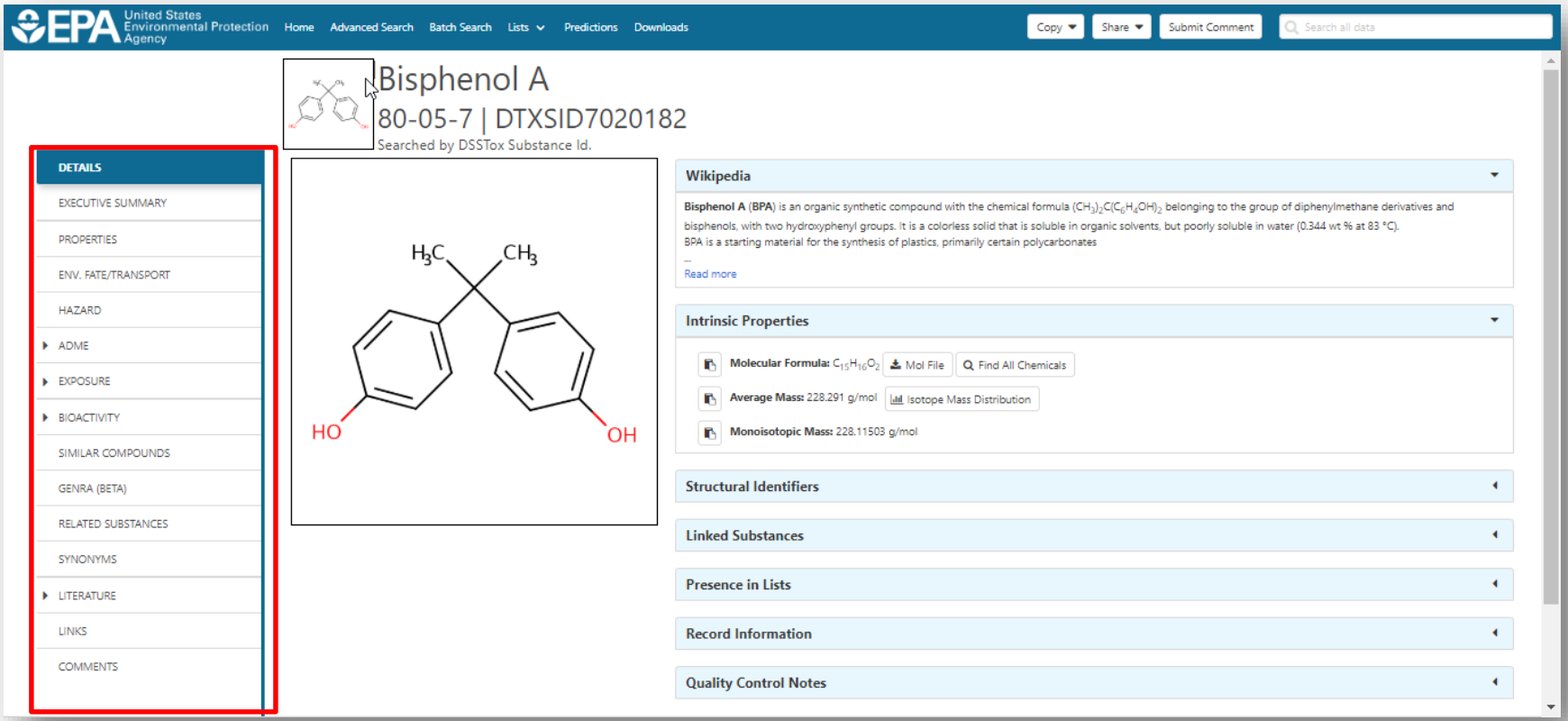
875k Chemical Substances



BASIC Search



Detailed Chemical Pages

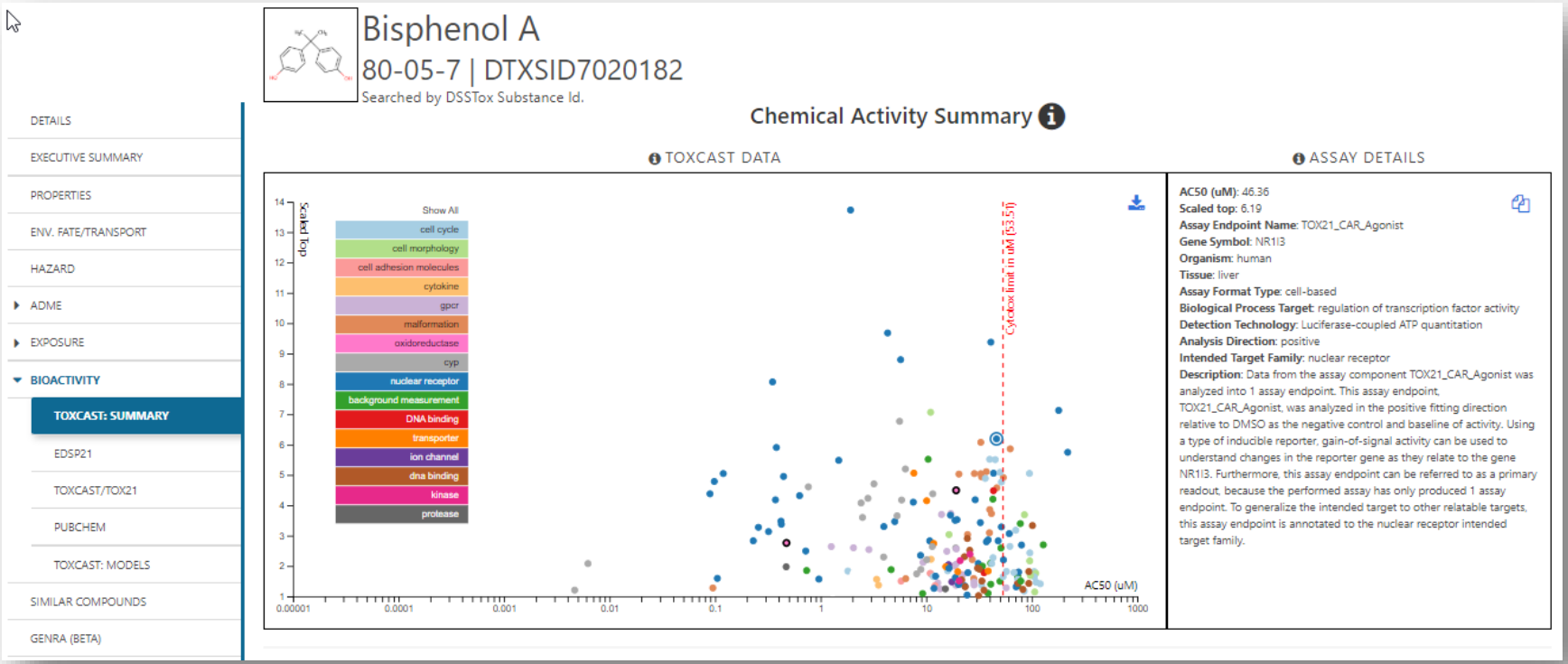


Human and Eco Hazard Data

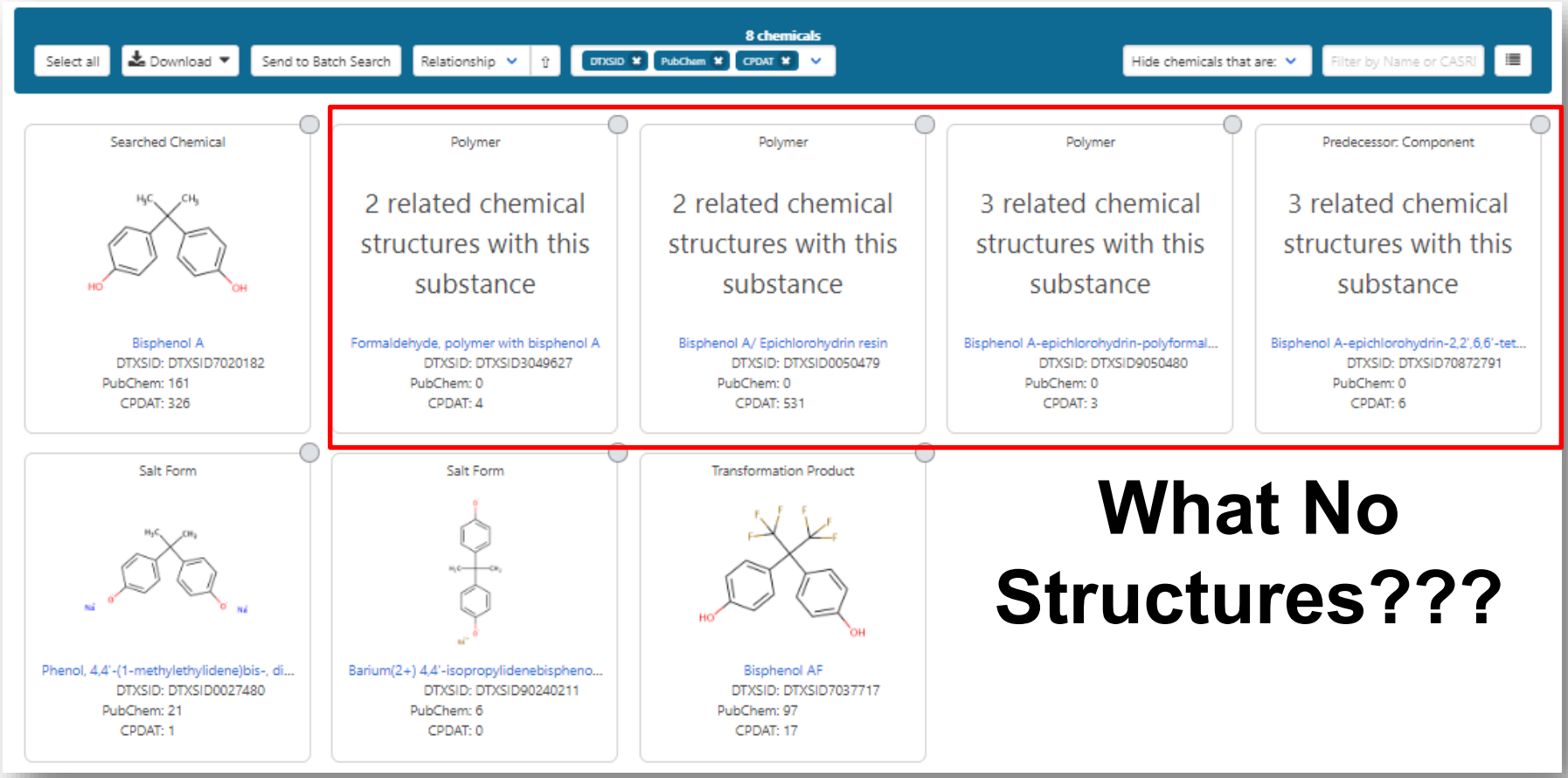
- ToxVal Database contains following data:
  - ~700,000 toxicity values
  - ~30 sources of data
  - ~22,000 sub-sources
  - ~5000 journals cited
  - ~70,000 literature citations

In Vitro Bioassay Screening

ToxCast and Tox21



Related Substances – Transformation Products, “Monomer-Polymer”



"UVCB" Chemical Substances

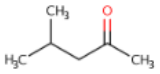
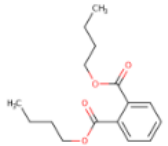
UVCB Chemicals

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.

Lots of UVCBS in Commerce….

1 related chemical structure with this substance Butoxypolypropylene glycol DTXSID: DTXSID3034404 PubChem: 0 CPDAT: 437	0 related chemical structures with this substance Fatty acids, tall-oil DTXSID: DTXSID0028088 PubChem: 0 CPDAT: 436	 Poly(oxy-1,2-ethanediyl), alpha-tridecyl-... DTXSID: DTXSID8043993 PubChem: 0 CPDAT: 435	1 related chemical structure with this substance Polyoxyethylene (20) oleylether DTXSID: DTXSID1027714 PubChem: 0 CPDAT: 435	0 related chemical structures with this substance Soy oil DTXSID: DTXSID8027660 PubChem: 0 CPDAT: 435	0 related chemical structures with this substance Bentonite DTXSID: DTXSID6030782 PubChem: 0 CPDAT: 433
0 related chemical structures with this substance Distillates, petroleum, hydrotreated heav... DTXSID: DTXSID3028217 PubChem: 0 CPDAT: 433	1 related chemical structure with this substance Edifas B DTXSID: DTXSID2020555 PubChem: 0 CPDAT: 433	0 related chemical structures with this substance Coconut oil DTXSID: DTXSID8027664 PubChem: 0 CPDAT: 432	 4-Methyl-2-pentanone DTXSID: DTXSID5021889 PubChem: 173 CPDAT: 428	0 related chemical structures with this substance Distillates, petroleum, solvent-refined he... DTXSID: DTXSID9028172 PubChem: 0 CPDAT: 426	 Dibutyl phthalate DTXSID: DTXSID2021781 PubChem: 138 CPDAT: 421

Example UVCBs with no structures

- Aroclor 1254
- Toxaphene
- Xylenes
- Demeton
- Aroclor 1248
- Asbestos
- Technical chlordane
- Coke oven emissions
- Creosote
- Diesel engine exhaust
- Nickel refinery dust
- Nickel, soluble salts
- Refractory ceramic fibers
- Toluene diisocyanate

What is Toxaphene?

"Complex, but reproducible mixture of at least ***175*** distinct C10-chloro compounds, having an approximate overall empirical formula of C10H10Cl8.; the 2 most active components are a C10H10Cl8 compound and a C10H11Cl7 compound which had been elucidated as 2,2,5-endo,6-exo,8,9,10-heptachlorobornane.

Produced by the chlorination of camphene to 67-69% chlorine by weight and made up of compds. of C10 H8 Cl10, C10 H18-n Cl n (mostly poychloroboranes) and C10 H16-n Cl n (oychloroboranes and/or polychlorotricyclenes) with n = 6 to 9"

Toxaphene has rich data

8001-35-2 | DTXSID7021368

Searched by DSSTox Substance Id.

Chemical Activity Summary

Hazard

DataType

8001-35-2 | DTXSID7021368

Searched by DSSTox Substance Id.

Toxics Release Inventory

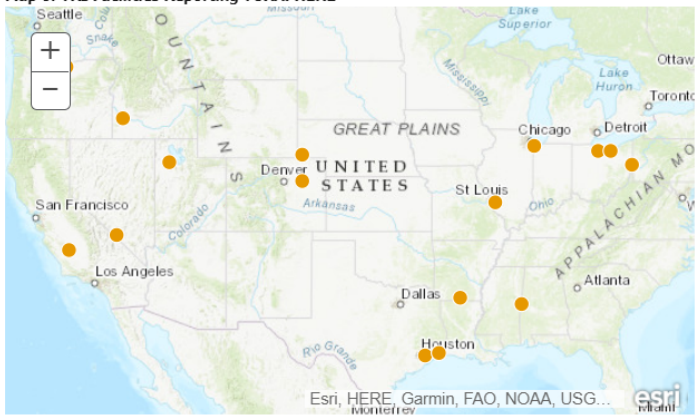
Print Page

2015 TRI Factsheet: Chemical - TOXAPHENE, 008001352

Data Source: 2018 Updated Dataset (released October 2019)

The Toxics Release Inventory (TRI) tracks the management of certain toxic chemicals that may pose a threat to human health and the environment. Certain industrial facilities in the U.S. must report annually how much of each chemical is recycled, combusted for energy recovery, treated for destruction, and disposed of or otherwise released on- and off-site. This information is collectively referred to as production-related waste managed.

Map of TRI Facilities Reporting TOXAPHENE

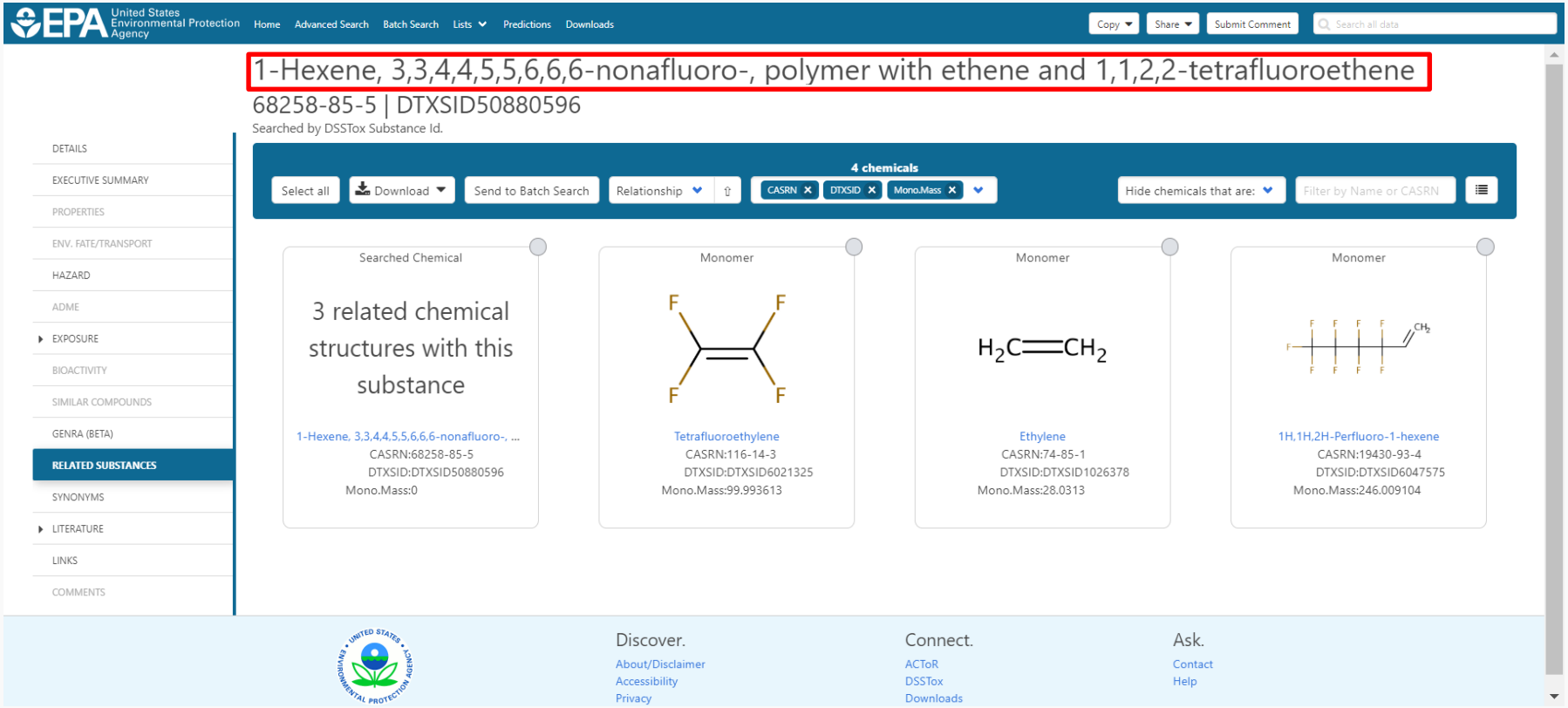


Quick Facts for 2015

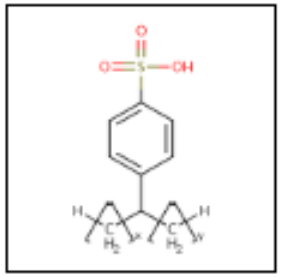
	Chemical	United States
Number of TRI Facilities:	16	22,286
Total Production-Related Waste Managed:	200.2 thousand lbs	28.9 billion lbs
Total On-site and Off-site Disposal or Other Releases:	15.0 thousand lbs	3.4 billion lbs
Total On-site:	14.0 thousand lbs	2.9 billion lbs
• Air:	25 lbs	690.4 million lbs
• Water:	0 lbs	199.3 million lbs
• Land:	14.0 thousand lbs	2.0 billion lbs
Total Off-site:	1.0 thousand lbs	511.4 million lbs

TOXAPHENE ranks **239** out of **499** chemicals reported to TRI in 2015 (Rank 1 = highest releases)

UVCB: Mapped Chemicals
Polymer individual components



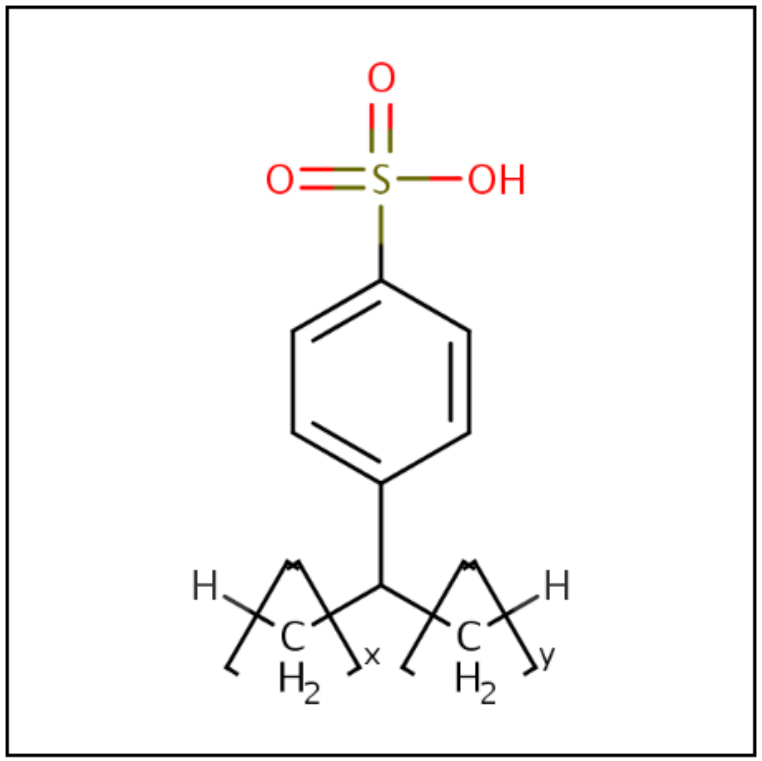
Linear Alkylbenzenesulfonates



Alkylbenzenesulfonate, linear

42615-29-2 | DTXSID3020041

Searched by DSSTox Substance Id.

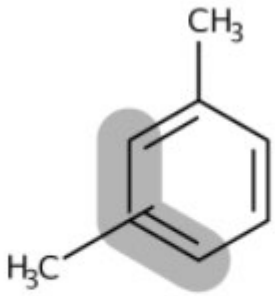


- Ambiguous representations of *some* structures are possible
- "Markush" representations can be valuable for describing and mapping chemicals

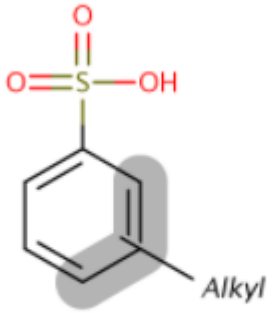
“Markush Structures”

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

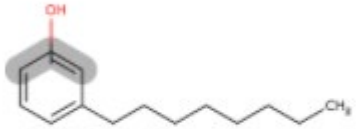
EPA United States Environmental Protection Agency



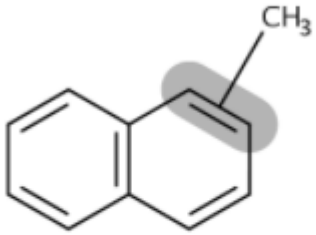
Xylenes
1330-20-7



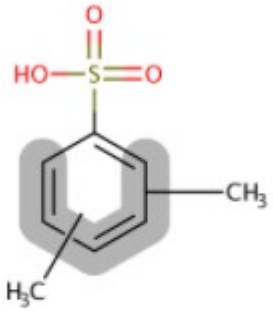
(C10-C16) Alkylbenzenesulfonic acid
68584-22-5



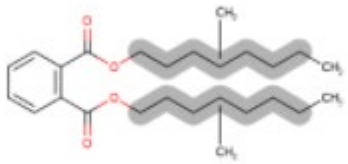
n-Nonylphenol
25154-52-3



Methylnaphthalene
1321-94-4

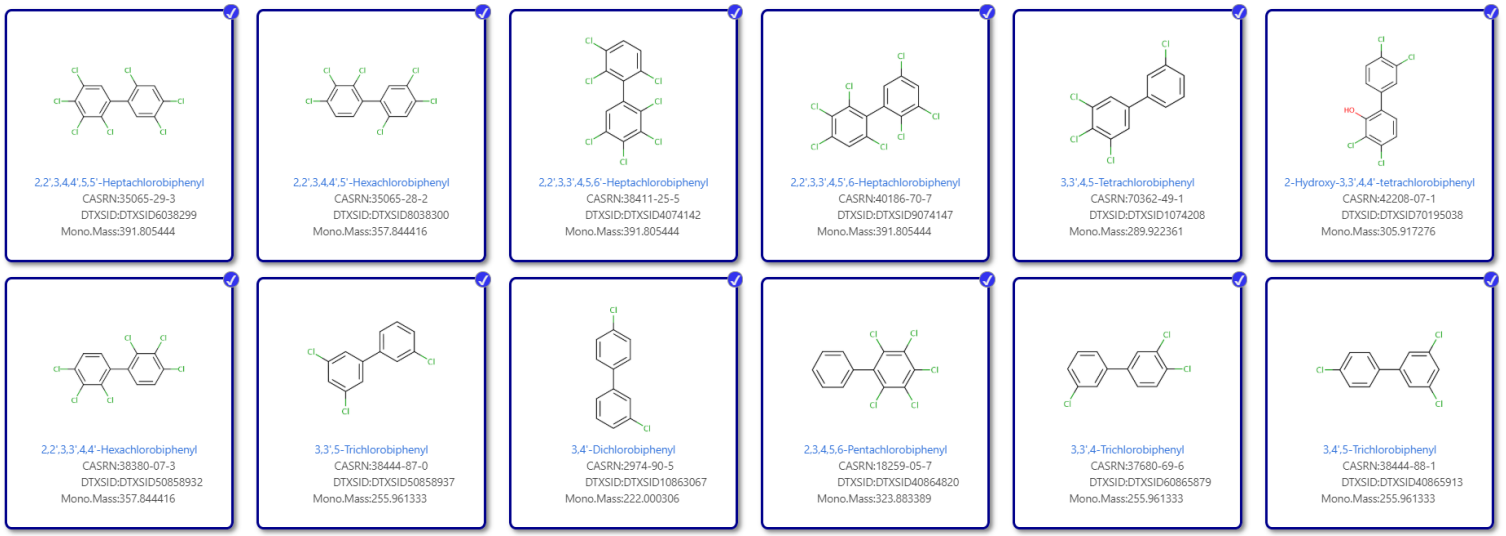


Sodium xylenesulfonate
1300-72-7



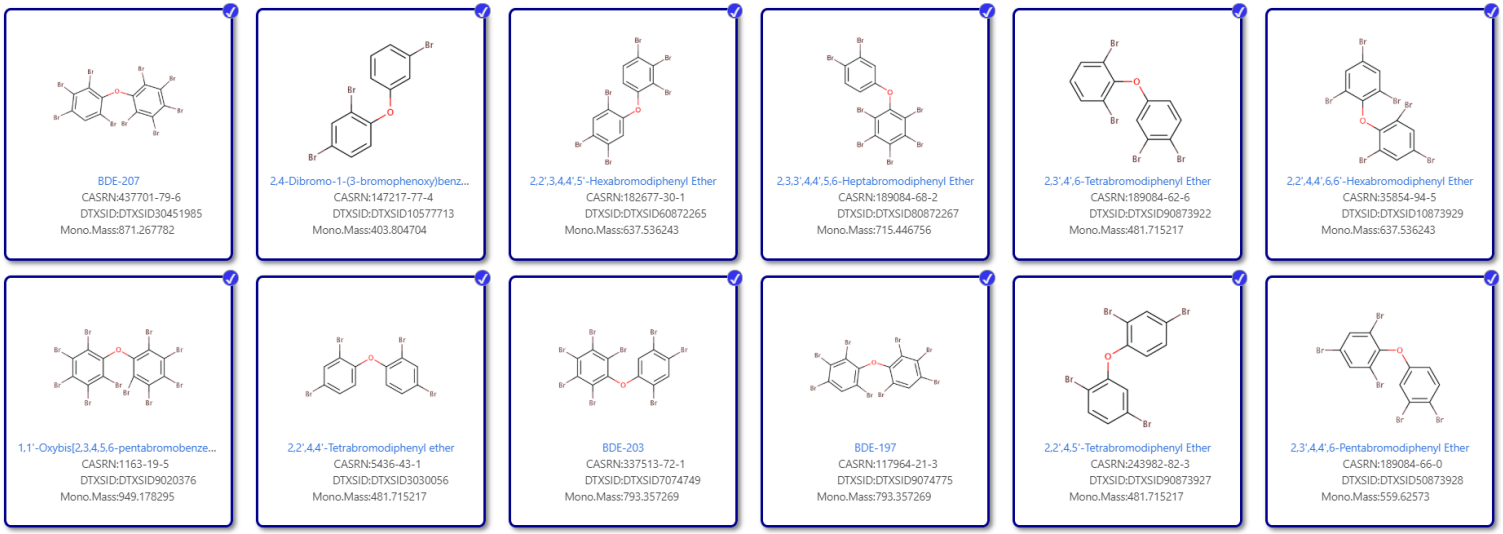
Diisononyl phthalate
28553-12-0

Chemical “grouping” – PCBs, PBDEs

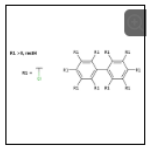


PCBs

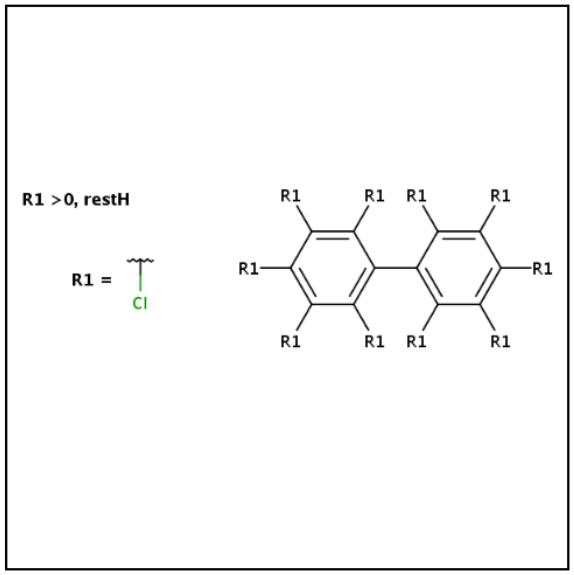
PBDEs



Some Markush chemicals are bounded



Polychlorinated biphenyls
1336-36-3 | DTXSID5024267
Searched by Expert Validated Synonym.

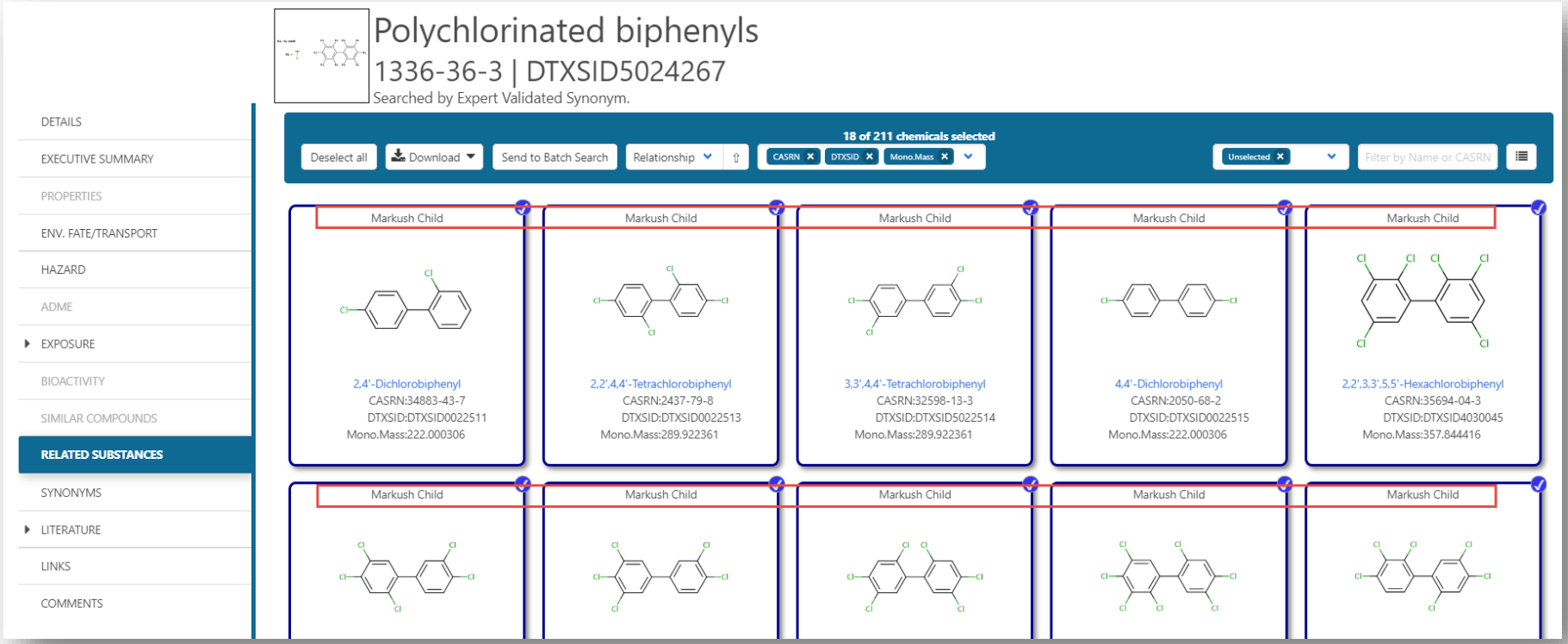


Previously 1-1 mapping

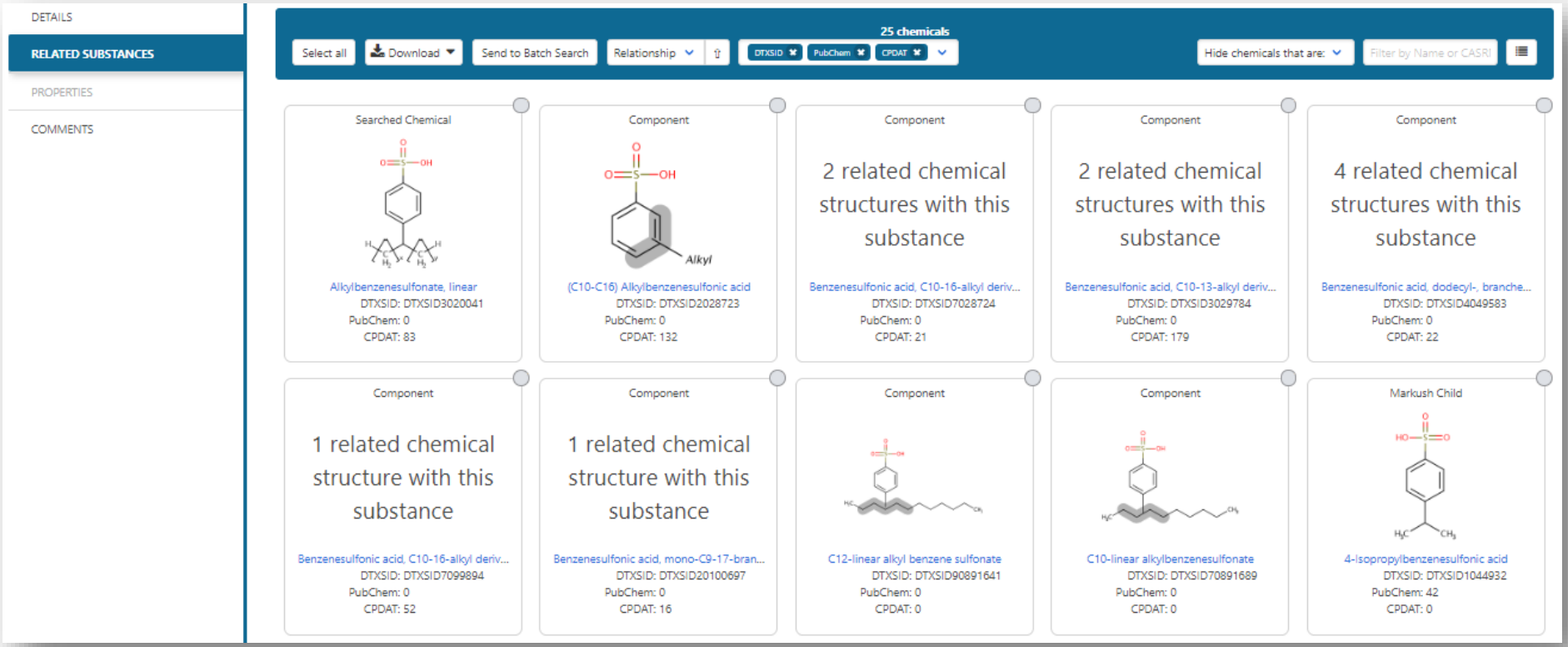
Successor Substances (209)

CAS-RN	Relationship	Source
32774-16-6	is a Representative Isomer of this	STN(DSSTox)
2051-60-7	is a Representative Isomer of this	Public
2051-61-8	is a Representative Isomer of this	Public
2051-62-9	is a Representative Isomer of this	Public
13029-08-8	is a Representative Isomer of this	Public
16605-91-7	is a Representative Isomer of this	Public
25569-80-6	is a Representative Isomer of this	Public
33284-50-3	is a Representative Isomer of this	Public
34883-43-7	is a Representative Isomer of this	Public
34883-39-1	is a Representative Isomer of this	Public

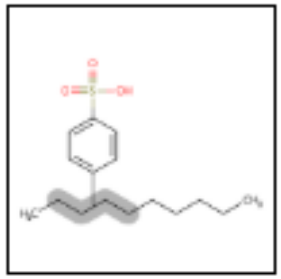
Now Markush Enumeration



UVCB: Complex Surfactants



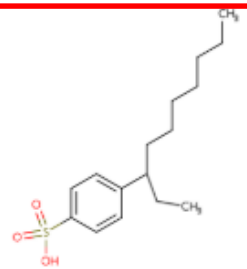
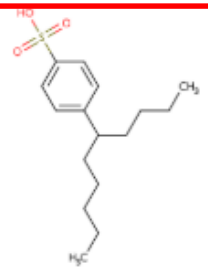
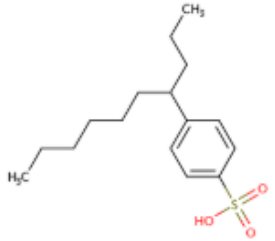
UVCB: Complex Surfactants



C10-linear alkylbenzenesulfonate

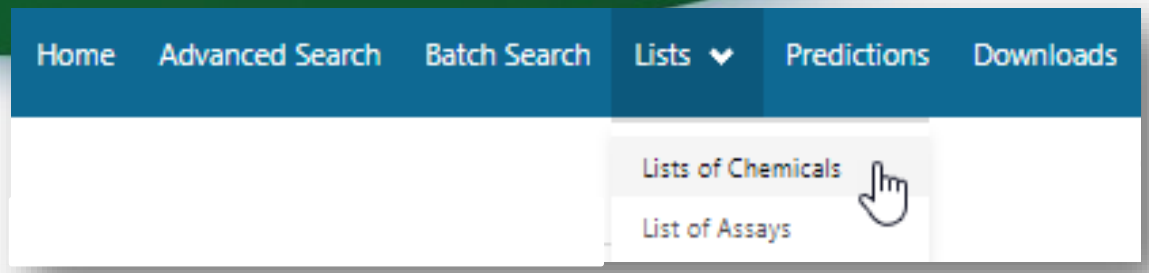
NOCAS_891689 | DTXSID70891689

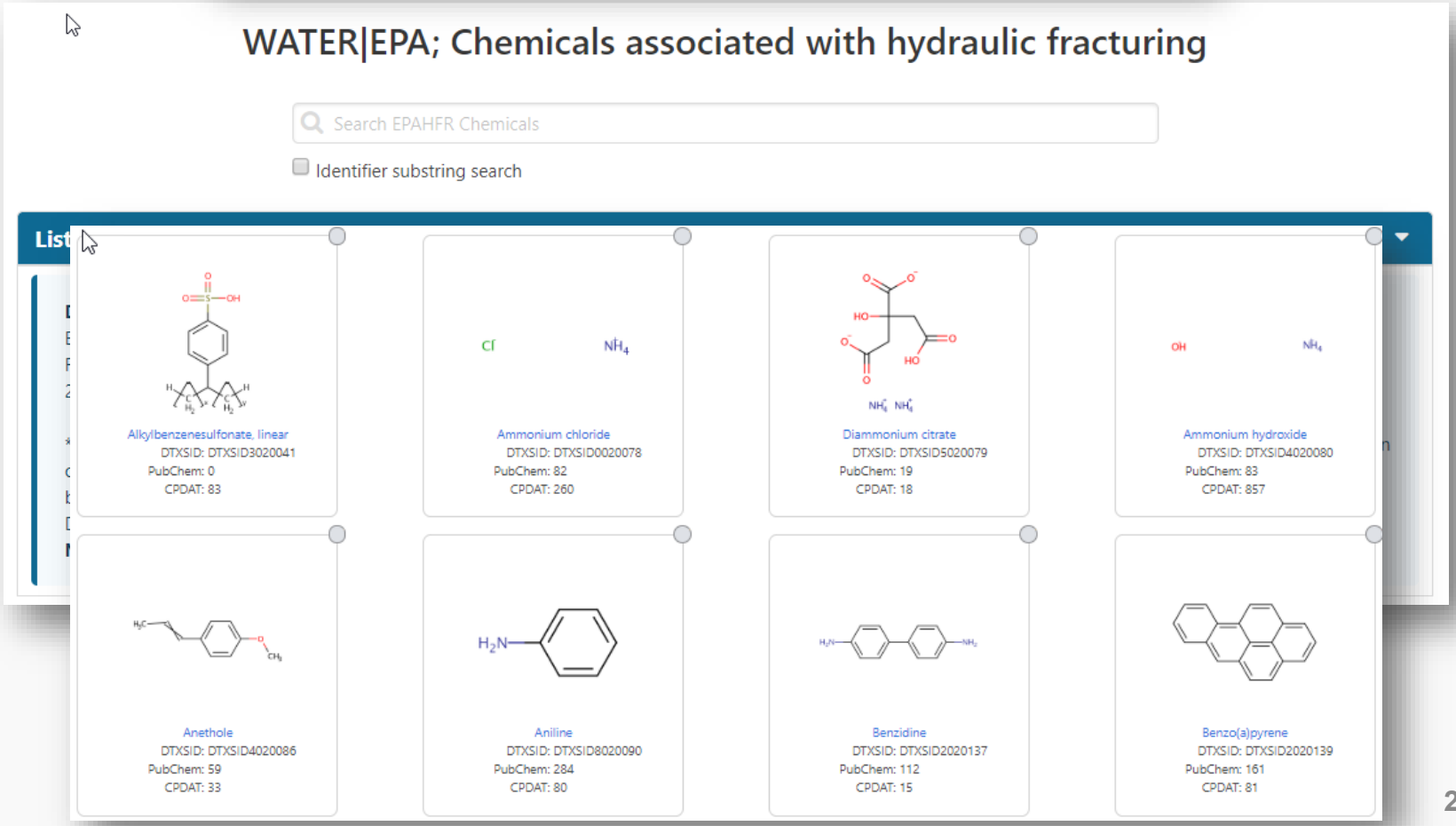
Searched by DSSTox Substance Id.

Markush Child	Markush Child	Markush Child
		
4-(Decan-3-yl)benzene-1-sulfonic acid	4-(decan-5-yl)benzene-1-sulfonic acid	4-(decan-4-yl)benzenesulfonic acid
DTXSID: DTXSID20859618	DTXSID: DTXSID70881146	DTXSID: DTXSID40891333
PubChem: 0	PubChem: 0	PubChem: 0
CPDAT: 0	CPDAT: 0	CPDAT: 0

Chemical Lists and Categories

EPAHFR: Hydraulic Fracturing





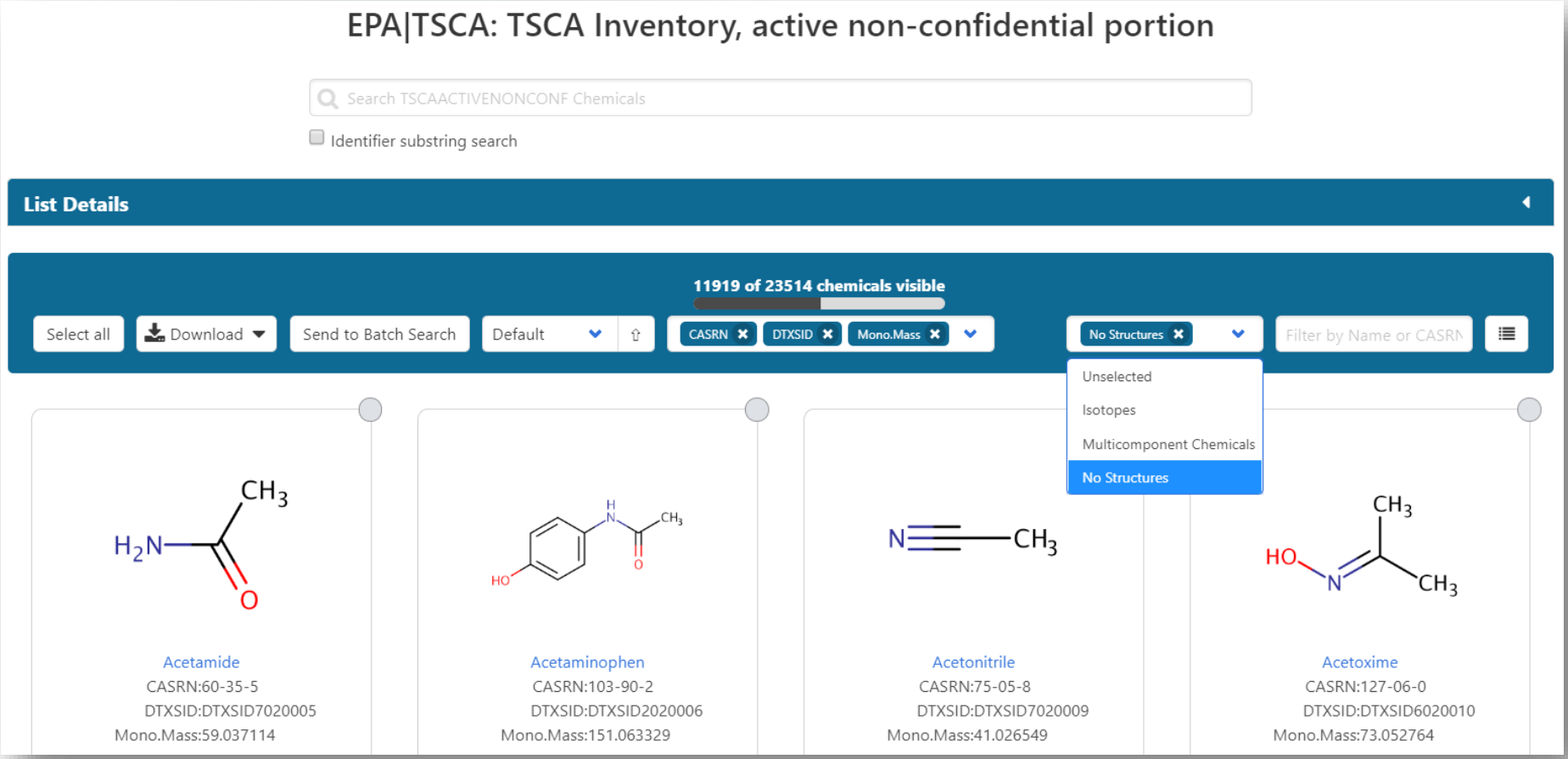
TSCA Inventory

TSCA Chemical Substance Inventory

CONTACT US

The Toxic Substances Control Act (TSCA) Chemical Substance Inventory contains all existing chemical substances manufactured, processed, or imported in the United States that do not qualify for an exemption or exclusion under TSCA. This may be your starting place for interaction with EPA on TSCA regulatory matters.

TSCA Inventory List
Active Non-confidential portion



Over 11,500 out of 23,500 have NO STRUCTURES

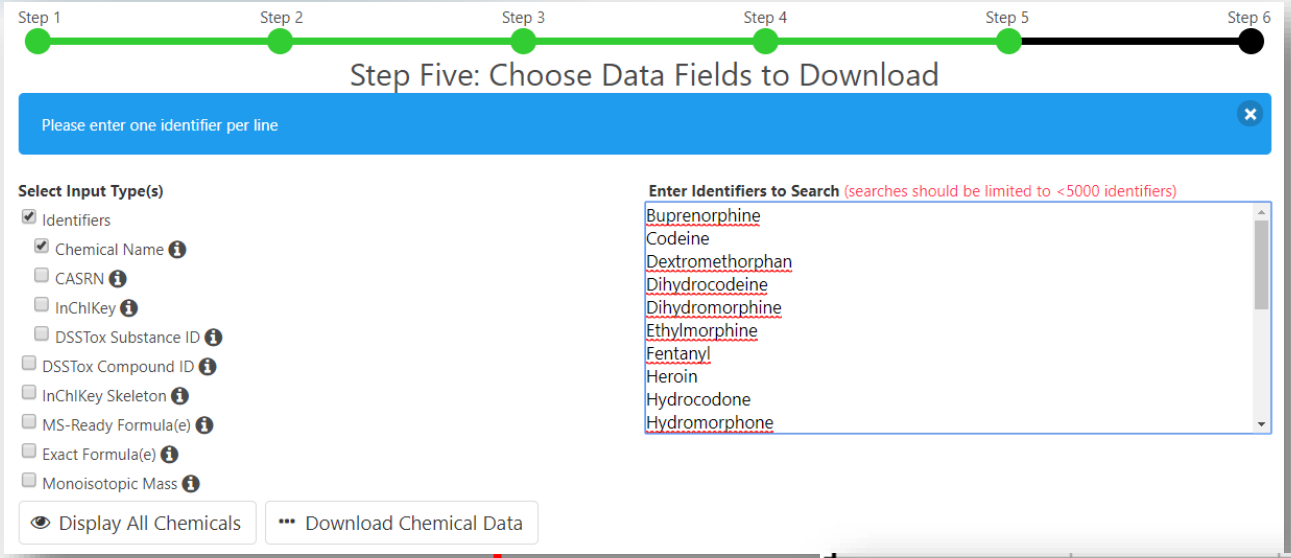
Batch Searching

Batch Searching

- Singleton searches are useful but people generally want data on LOTS of chemicals!

- 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 Search Names



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

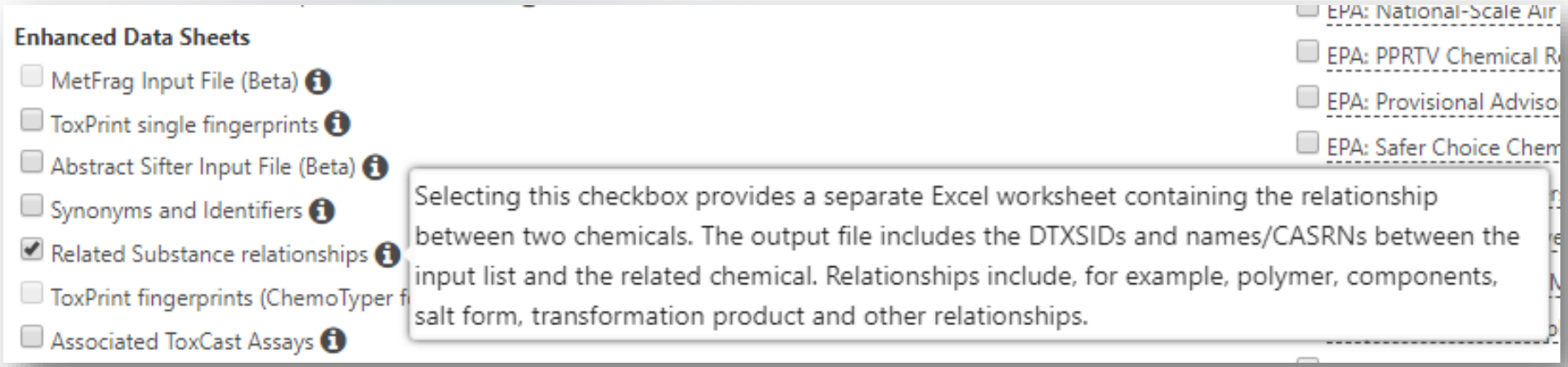
- [ ] Mol File
- [ ] SMILES
- [ ] InChI String
- [x] MS-Ready SMILES
- [ ] QSAR-Ready SMILES

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(=

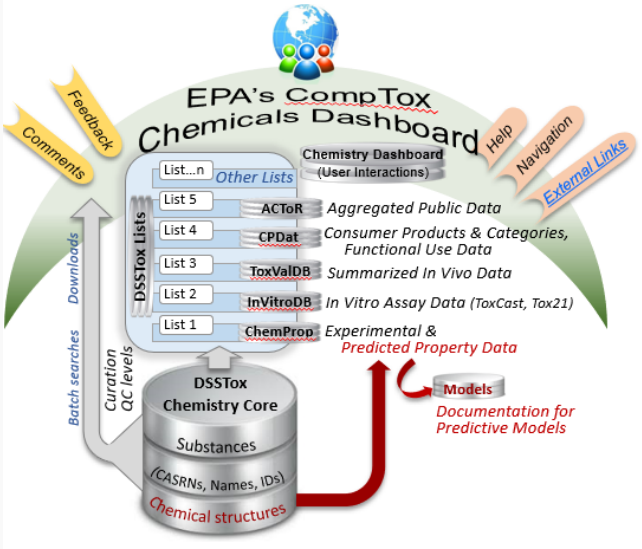
Related Substance Relationships



	A	B	C	D	E	
1	INPUT	DTXSID	PREFERRED_NAME	HAS_RELATIONSHIP_WITH	RELATED_DTXSID	RELATED_PREFERRED_NAM
2	xylenes	DTXSID2021446	Xylenes	Transformation Product	DTXSID40176394	N-Benzoylalanine
3	xylenes	DTXSID2021446	Xylenes	Component	DTXSID6026298	m-Xylene
4	xylenes	DTXSID2021446	Xylenes	Component	DTXSID3021807	o-Xylene
5	xylenes	DTXSID2021446	Xylenes	Component	DTXSID2021868	p-Xylene
6	xylenes	DTXSID2021446	Xylenes	Predecessor: Component	DTXSID9021421	Xylenes; defined mixture 1
7	xylenes	DTXSID2021446	Xylenes	Predecessor: Component	DTXSID7021447	Xylenes; defined mixture 2
8	xylenes	DTXSID2021446	Xylenes	Predecessor: Component	DTXSID30891529	Total Petroleum Hydrocarbons
9	xylenes	DTXSID2021446	Xylenes	Markush Child	DTXSID3021807	o-Xylene
10	xylenes	DTXSID2021446	Xylenes	Markush Child	DTXSID6026298	m-Xylene
11	xylenes	DTXSID2021446	Xylenes	Markush Child	DTXSID2021868	p-Xylene

Conclusion

- Building an integrated hub for environmental chemistry data to serve computational toxicology
- Transparent access to data and models – file downloads, SQL data dumps and web services
- Expansion of functionality to serve all data streams generated by NCCT across the agency & community
- Data **QUALITY** is a key focus - ongoing curation
- Ongoing API development will provide enhanced access to data streams



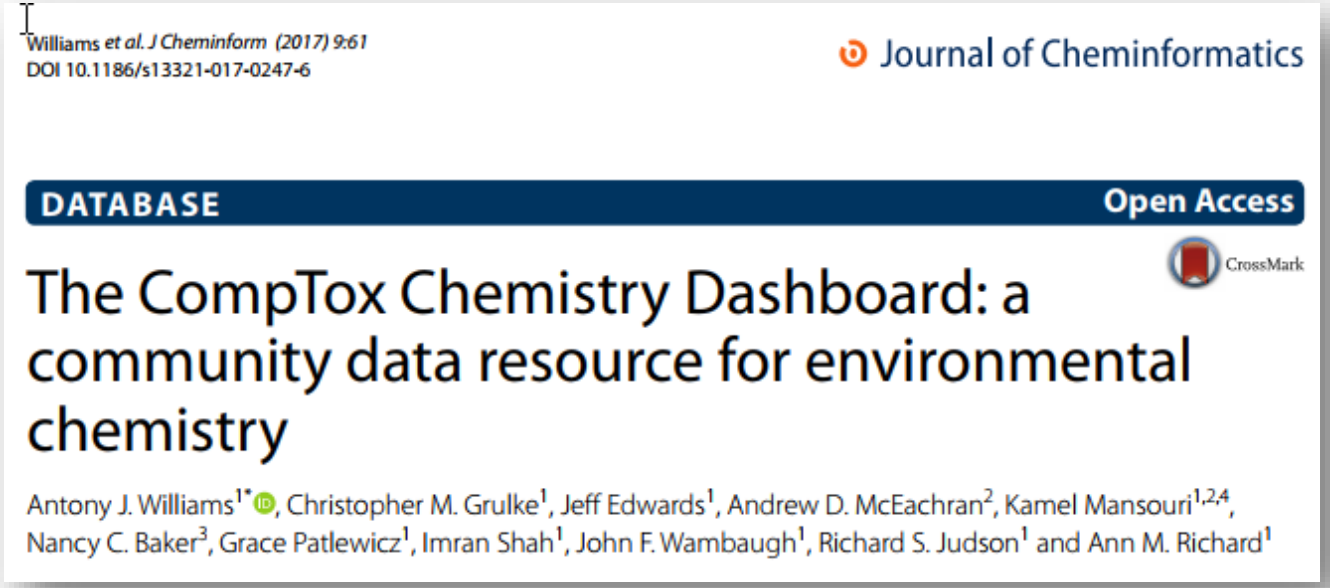
Acknowledgements

EPA-RTP

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

Contact

Antony Williams
CCTE, US EPA Office of Research and Development,
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821



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