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

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

Antony Williams

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

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

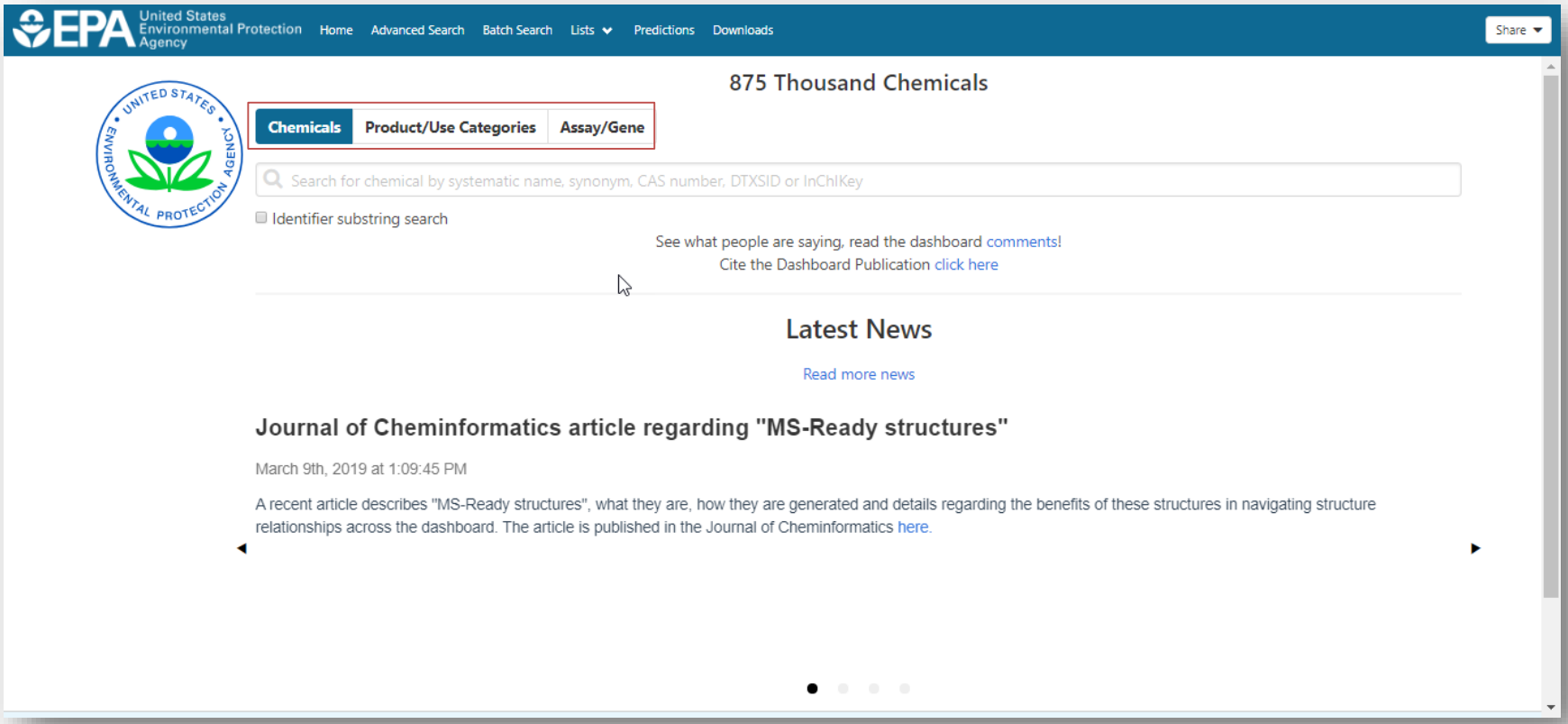
Unravelling the Exposome
ASMS-Sanibel Conference, January 2020

Overview

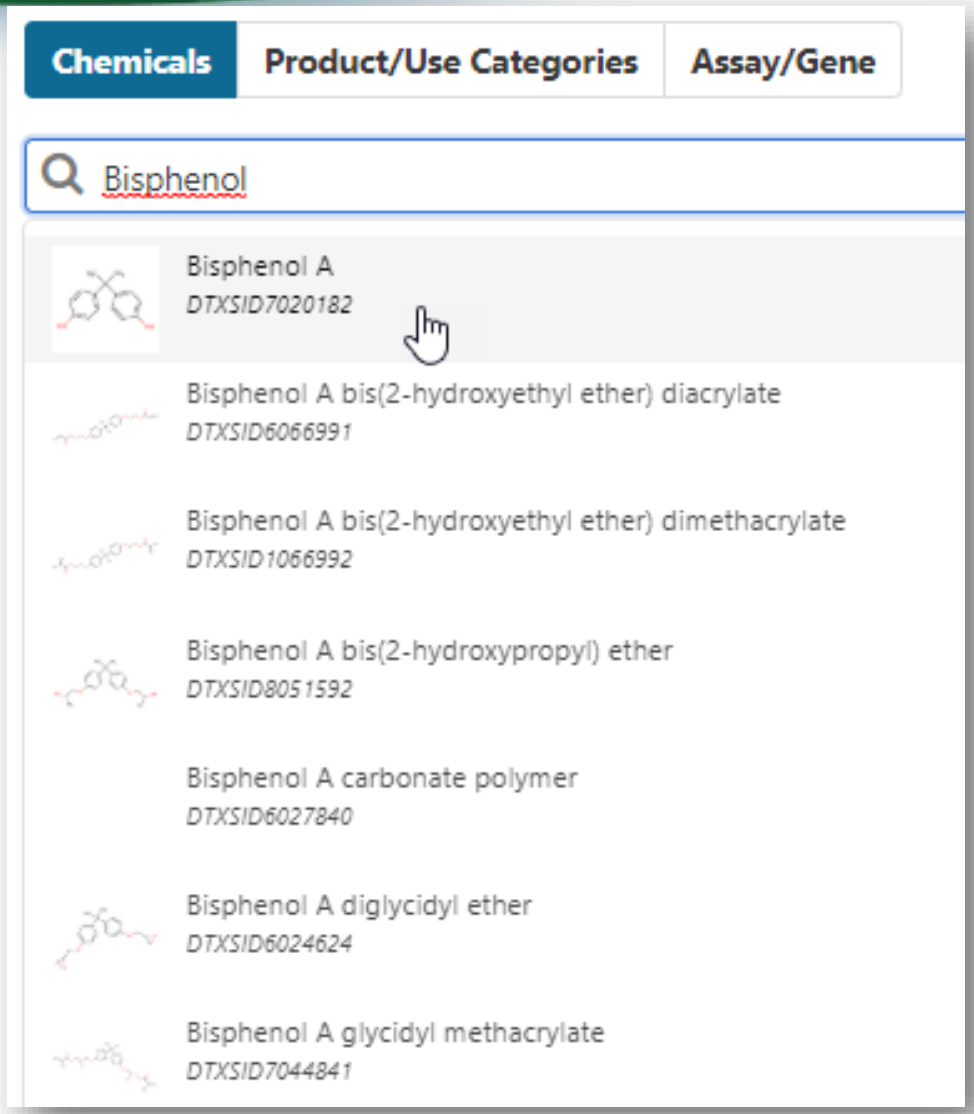
- The CompTox Chemicals Dashboard - web-based database of 875k substances
- Associated toxicity data including:
  - Experimental and predicted physicochemical data
  - *In vivo* hazard data
  - *In vitro* bioactivity screening data
  - Link farm to tens of public resources
- Integrated modules – read-across,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

CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard



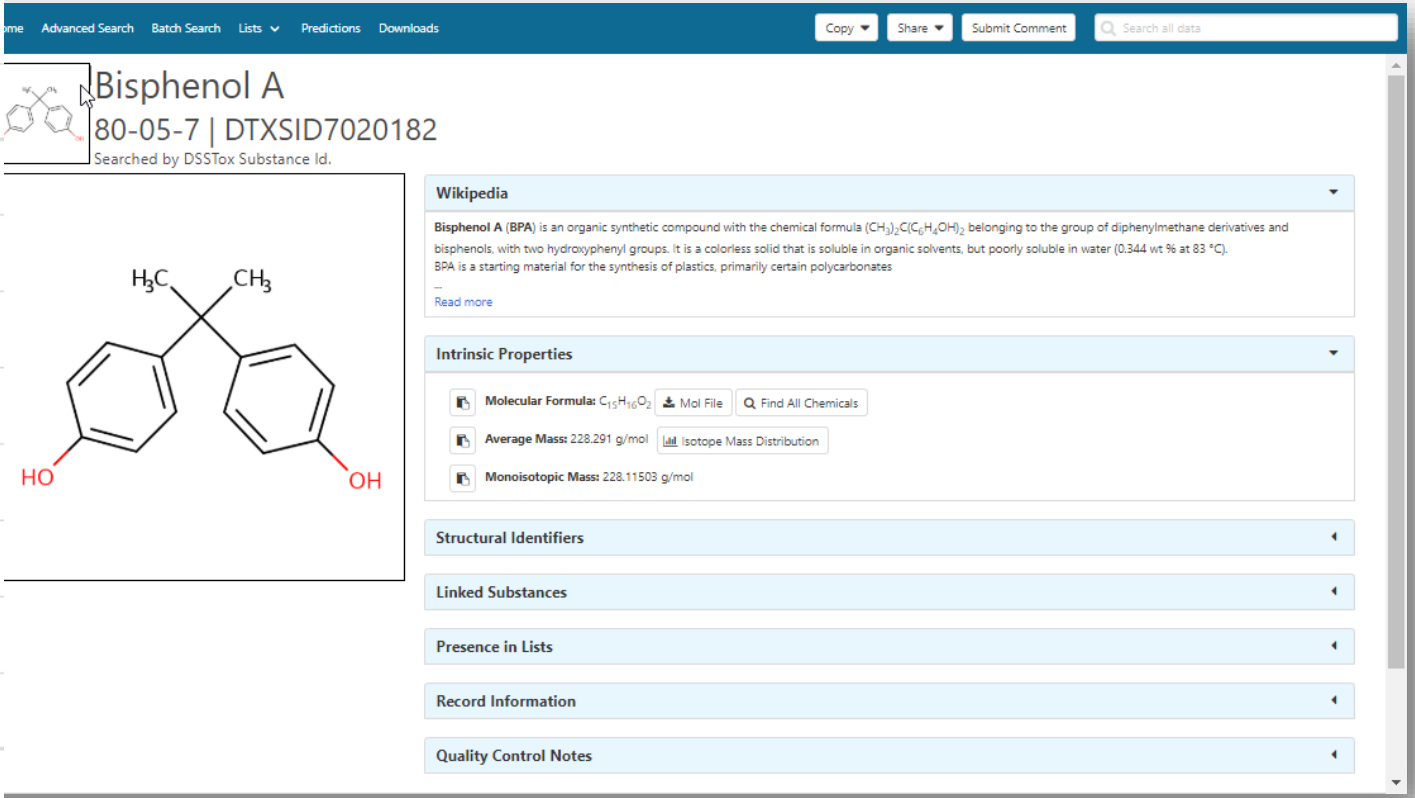
BASIC Search



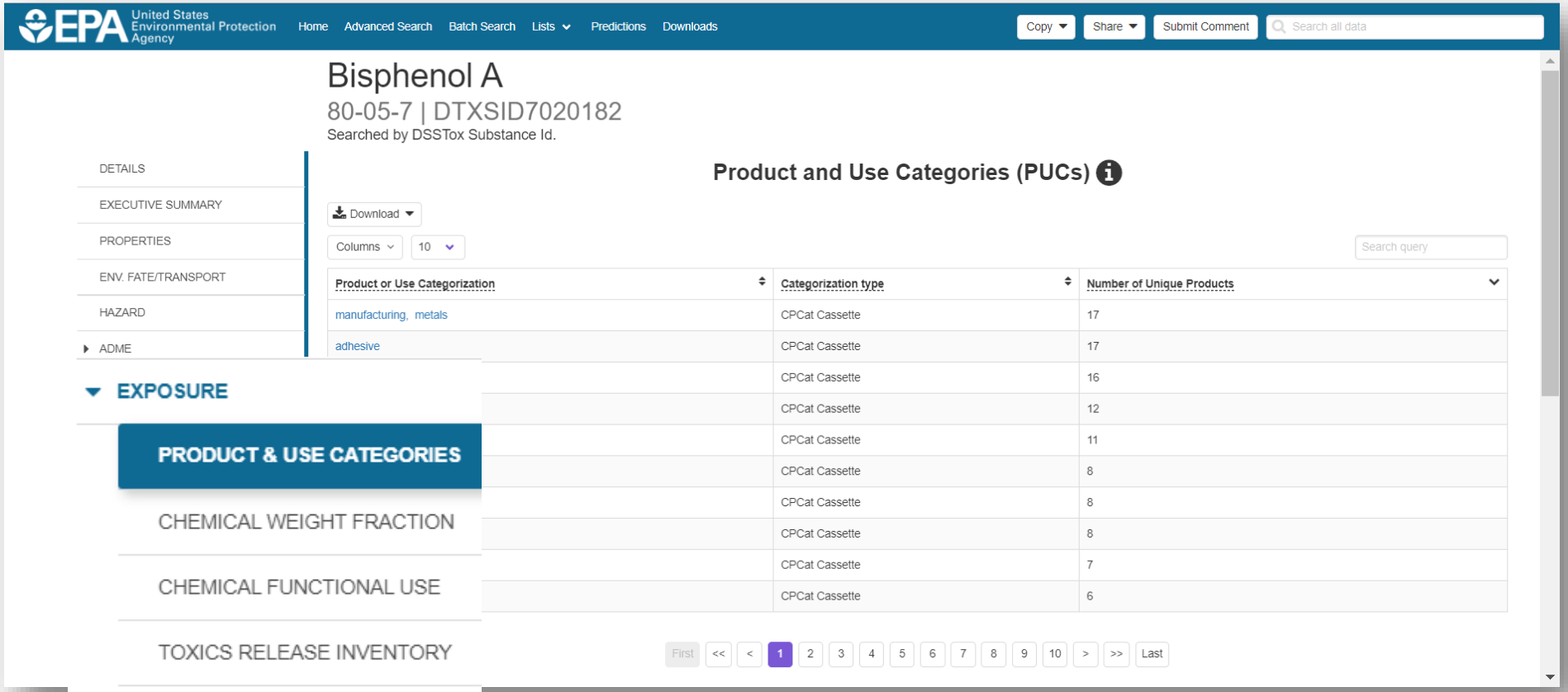
Detailed Chemical Pages

EPA United States Environmental Protection Agency

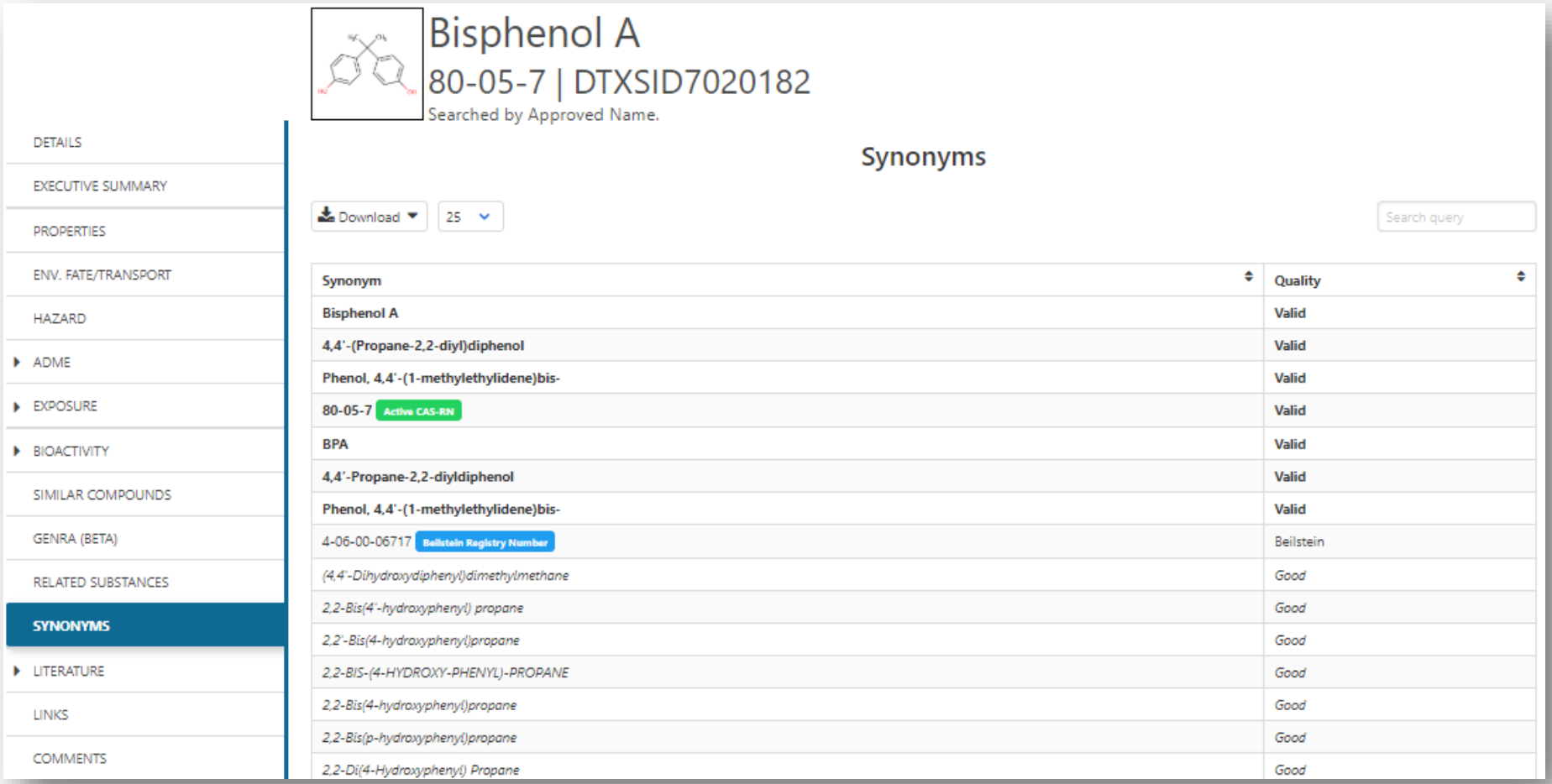
- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



Sources of Exposure to Chemicals



Identifiers to Support Searches



Bisphenol A

80-05-7 | DTXSID7020182

Searched by Approved Name.

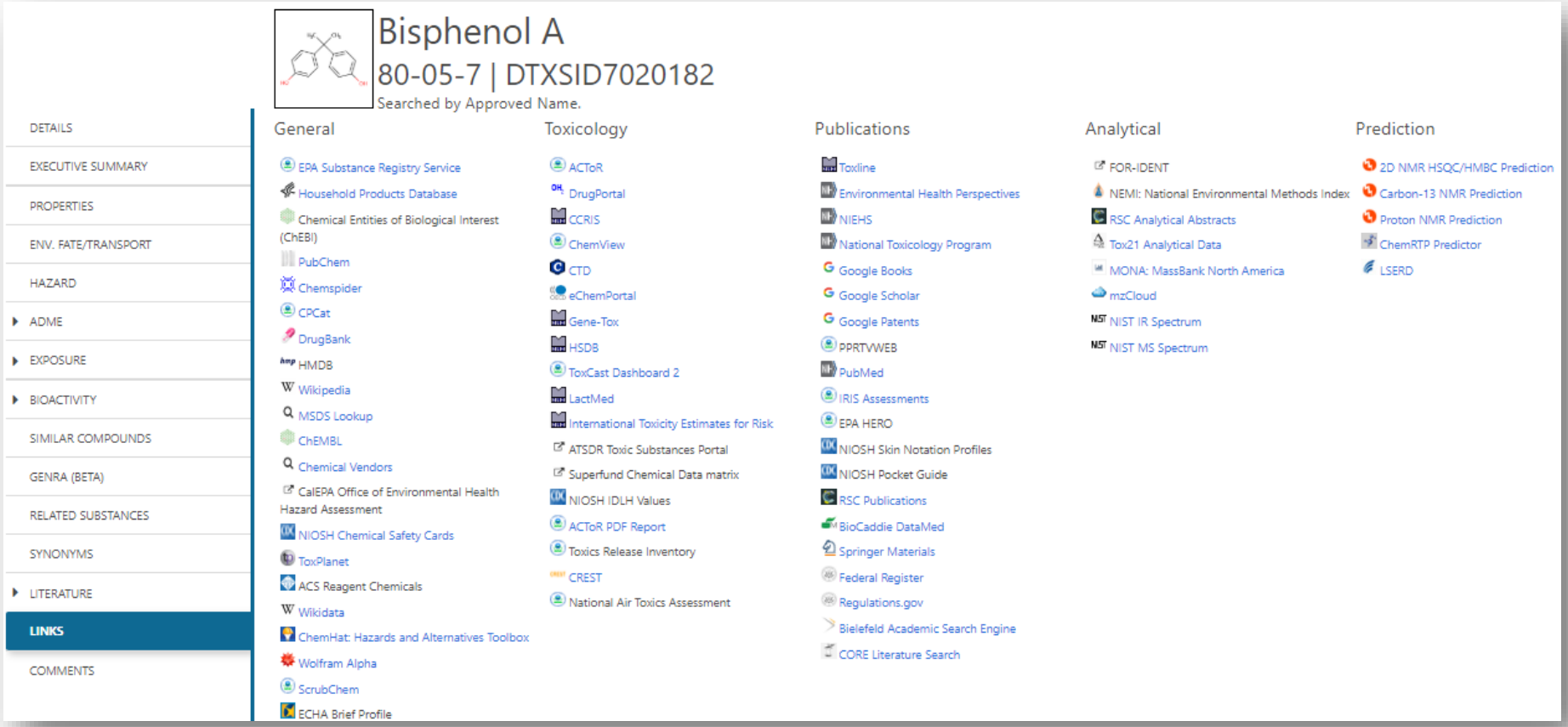
DETAILS
EXECUTIVE SUMMARY
PROPERTIES
ENV. FATE/TRANSPORT
HAZARD
ADME
EXPOSURE
BIOACTIVITY
SIMILAR COMPOUNDS
GENRA (BETA)
RELATED SUBSTANCES
SYNONYMS
LITERATURE
LINKS
COMMENTS

Synonyms

Download | 25 | Search query

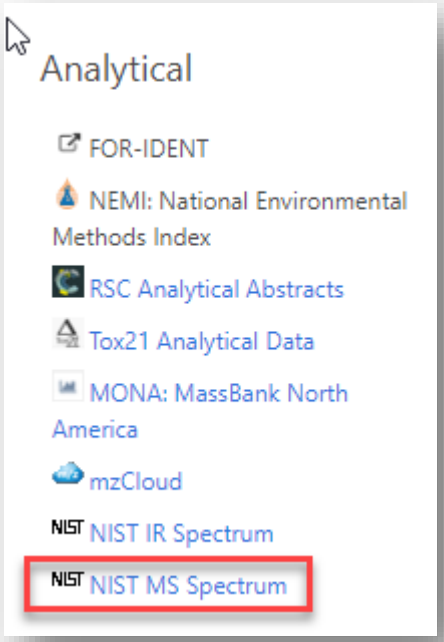
Synonym	Quality
Bisphenol A	**Valid**
4,4'-(Propane-2,2-diyl)diphenol	**Valid**
Phenol, 4,4'-(1-methylethylidene)bis-	**Valid**
80-05-7 Active CAS-RN	**Valid**
BPA	**Valid**
4,4'-Propane-2,2-diyldiphenol	**Valid**
Phenol, 4,4'-(1-methylethylidene)bis-	**Valid**
4-06-00-06717 Beilstein Registry Number	Beilstein
(4,4'-Dihydroxydiphenyl)dimethylmethane	*Good*
2,2-Bis(4'-hydroxyphenyl) propane	*Good*
2,2'-Bis(4-hydroxyphenyl)propane	*Good*
2,2-BIS-(4-HYDROXY-PHENYL)-PROPANE	*Good*
2,2-Bis(4-hydroxyphenyl)propane	*Good*
2,2-Bis(p-hydroxyphenyl)propane	*Good*
2,2-Di(4-Hydroxyphenyl) Propane	*Good*

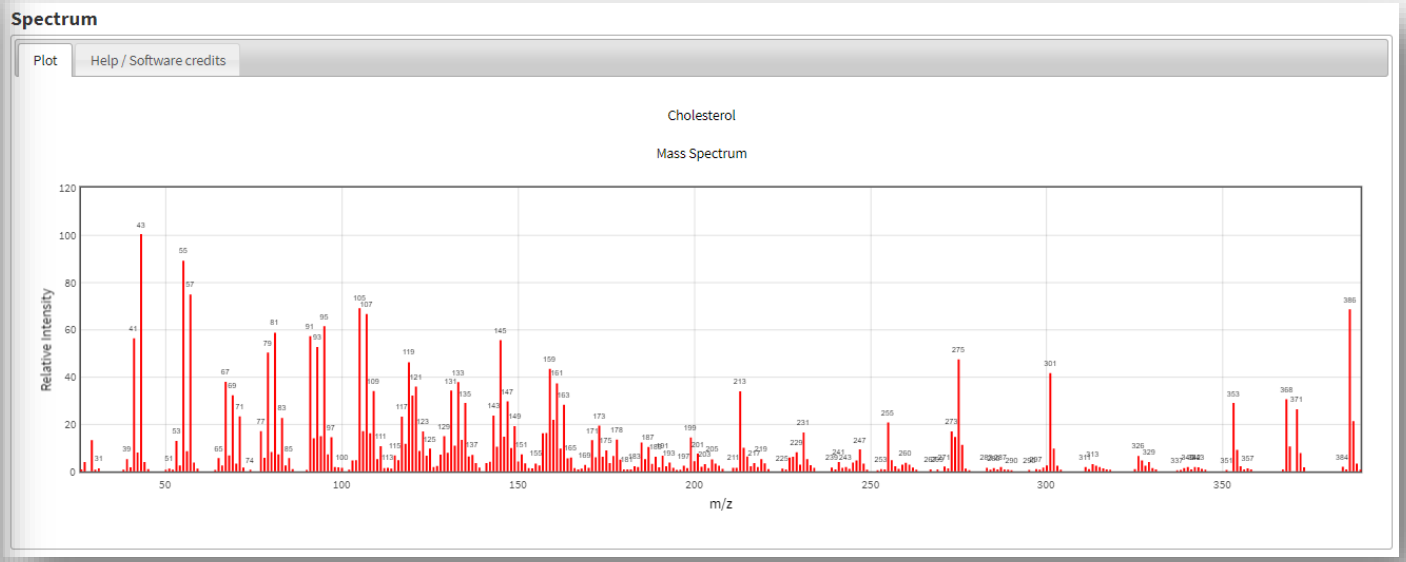
Link Access


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

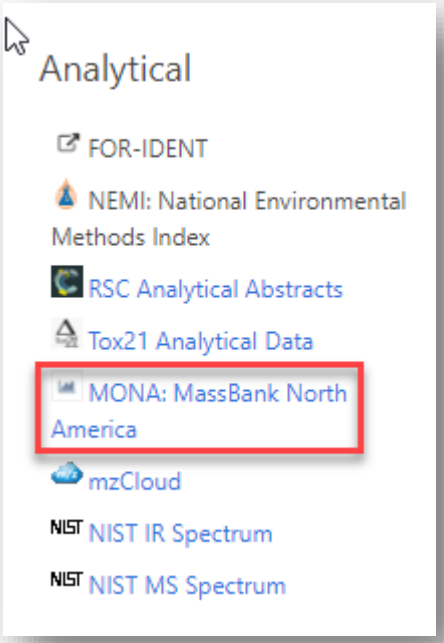
NIST WebBook
https://webbook.nist.gov/chemistry/

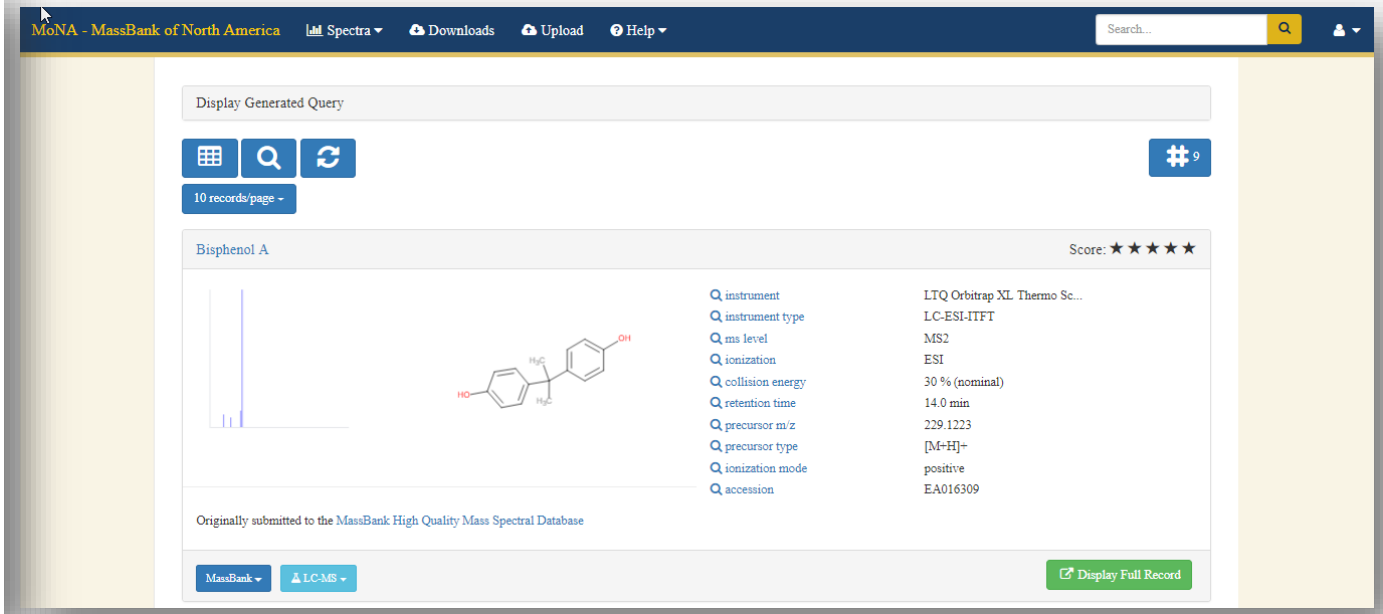
EPA United States Environmental Protection Agency





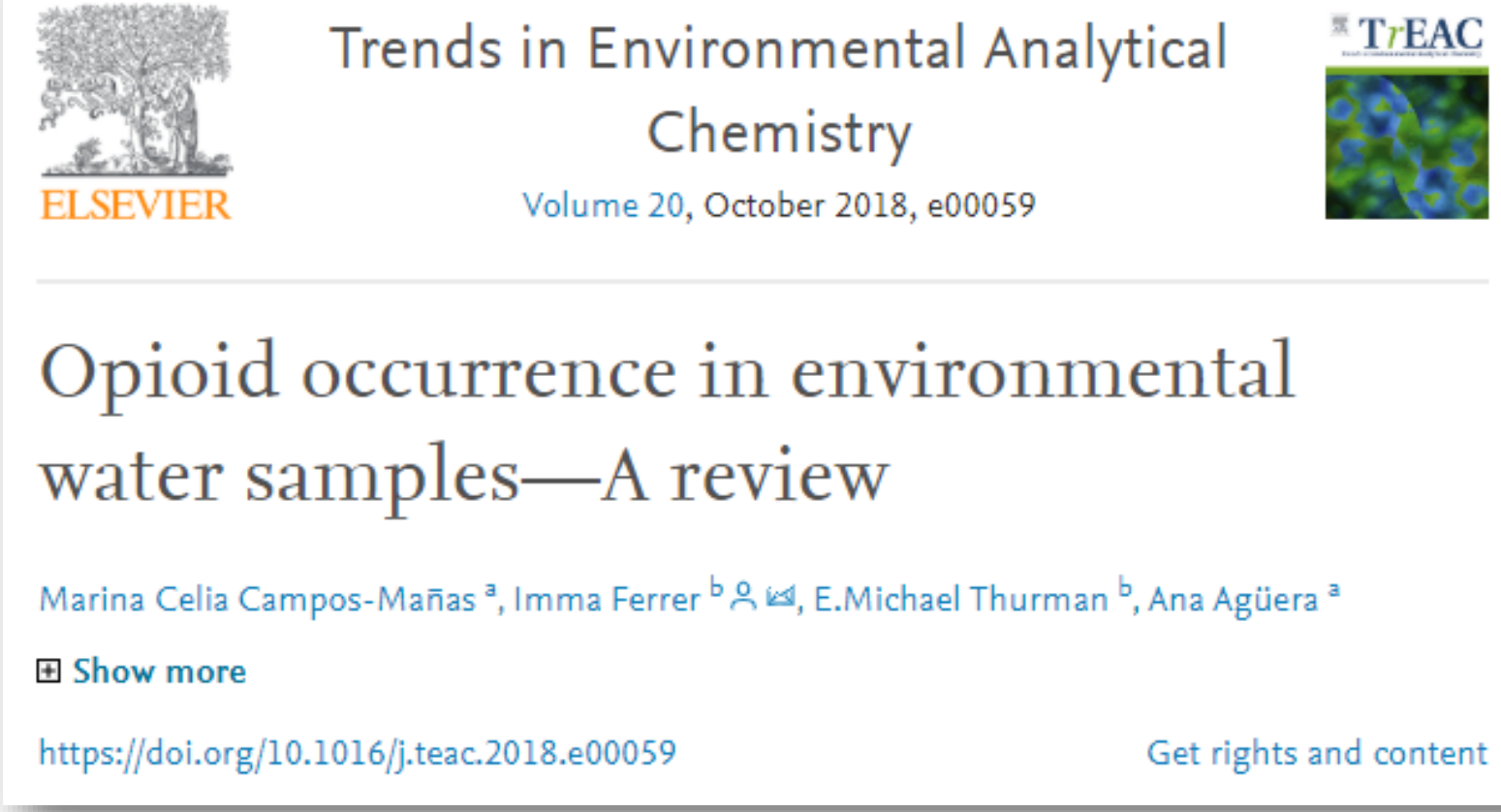
MassBank of North America
https://mona.fiehnlab.ucdavis.edu





Batch Searching

Aggregate data for a list of chemicals



ELSEVIER

Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059

TrEAC

Opioid occurrence in environmental water samples—A review

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

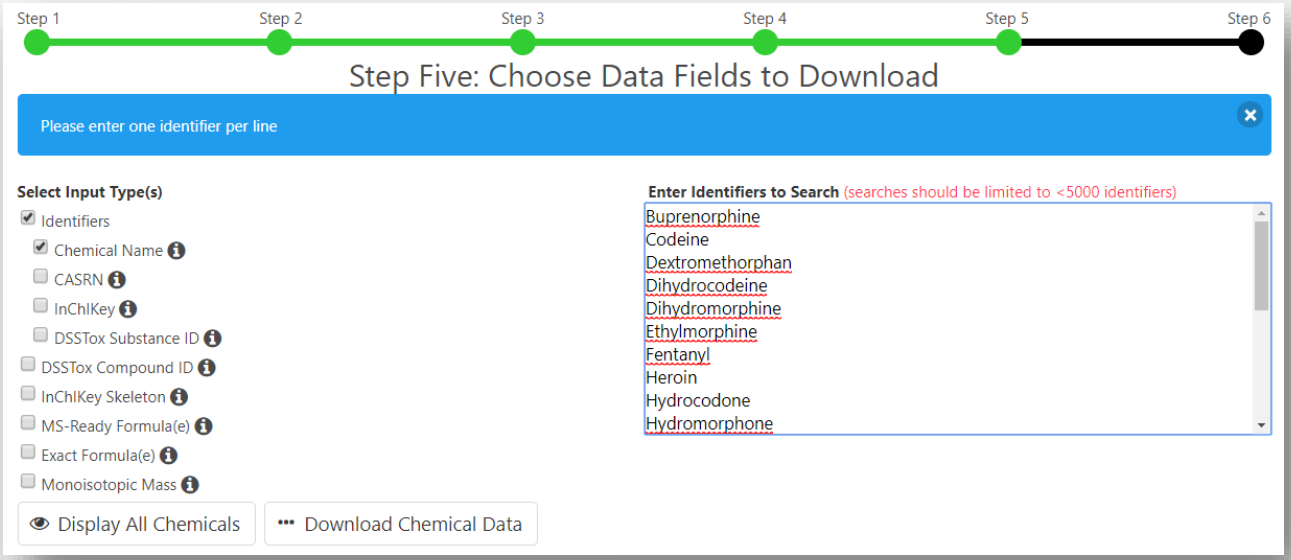
Show more

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

Get rights and content

Batch Search Names

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



Excel Download

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

Add Other Data of Interest

Chemical Identifiers

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

Structures

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

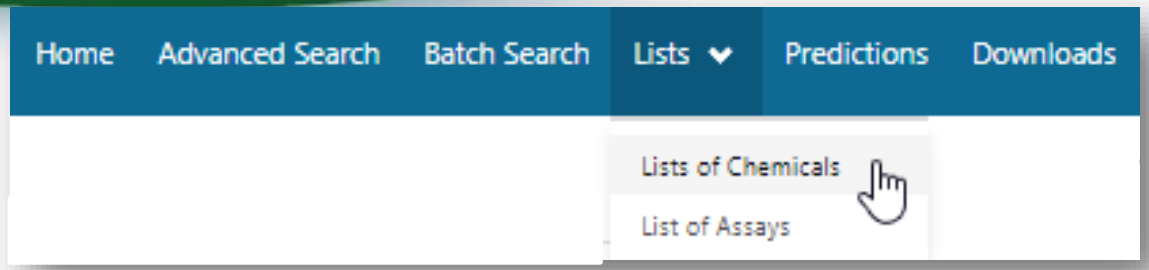
Intrinsic And Predicted Properties

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

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

Chemical Lists of Interest…

Chemical Lists



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

PFAS lists of Chemicals

EPA United States Environmental Protection Agency

Select List

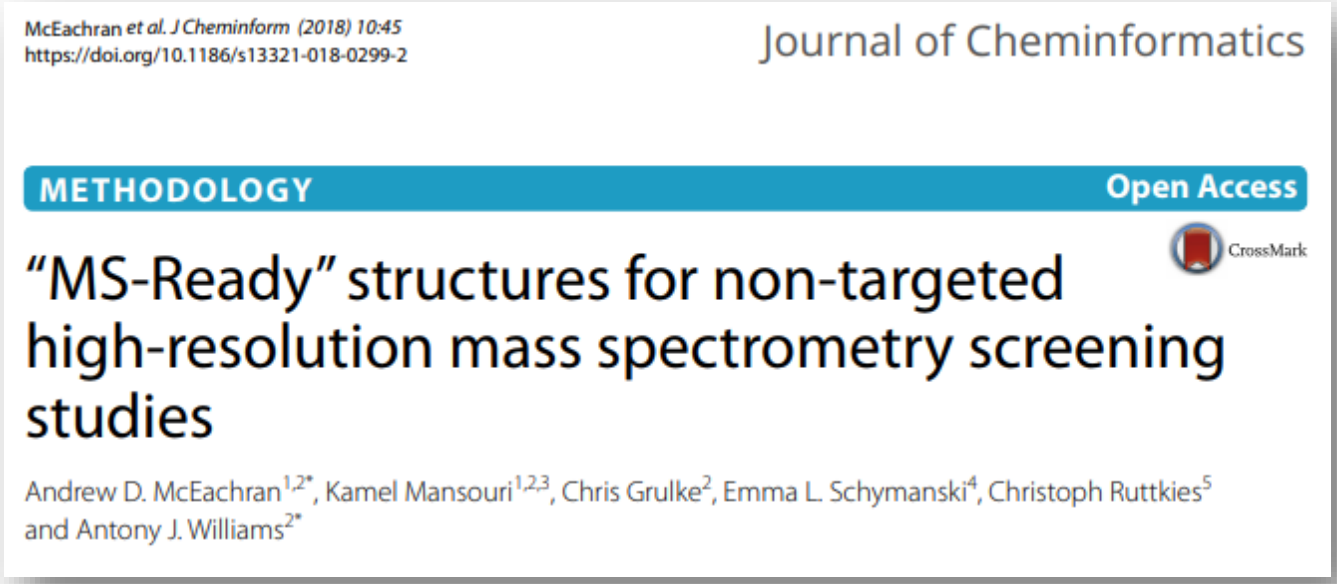
Download | Columns | PFAS | Copy Filtered Lists URL

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

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

"MS-ready" structures



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

Journal of Cheminformatics

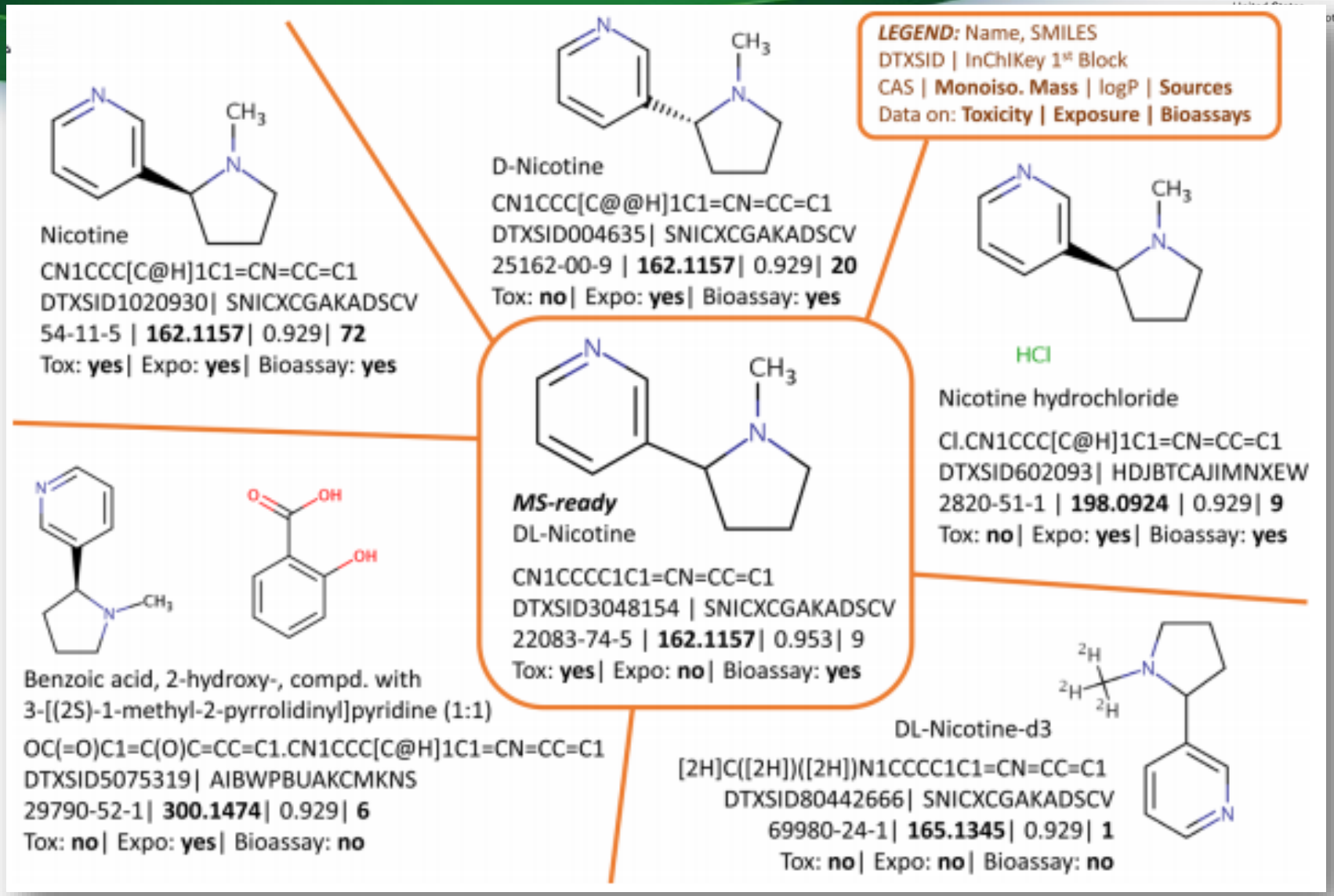
METHODOLOGY Open Access

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

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

Overview of MS-Ready Structures

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

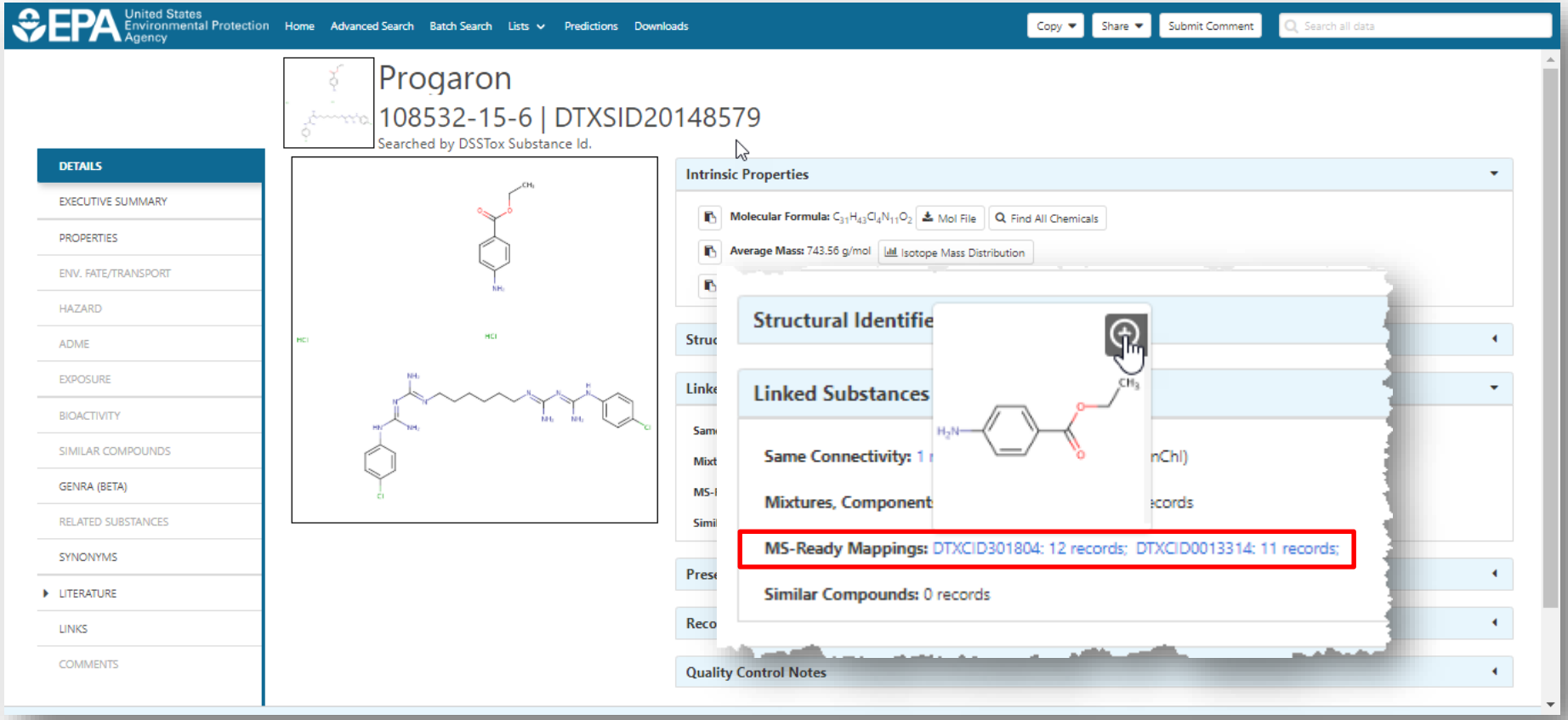


Environmental Science & Technology
Viewpoint
pubs.acs.org/est

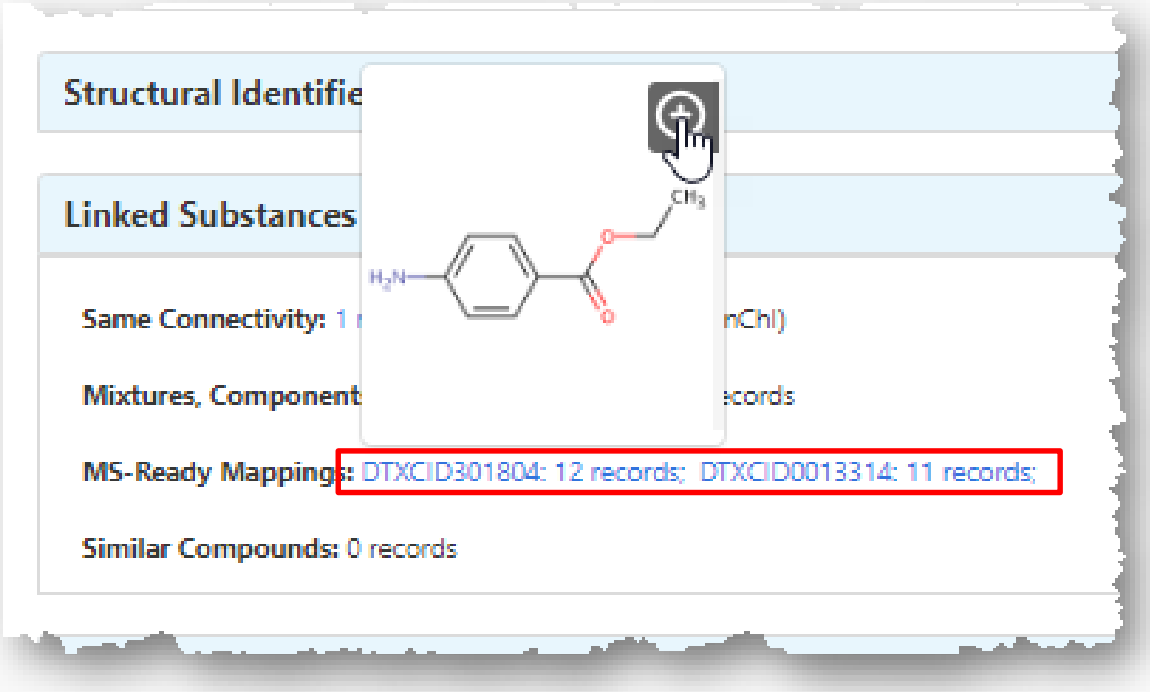
Open Science for Identifying "Known Unknown" Chemicals

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

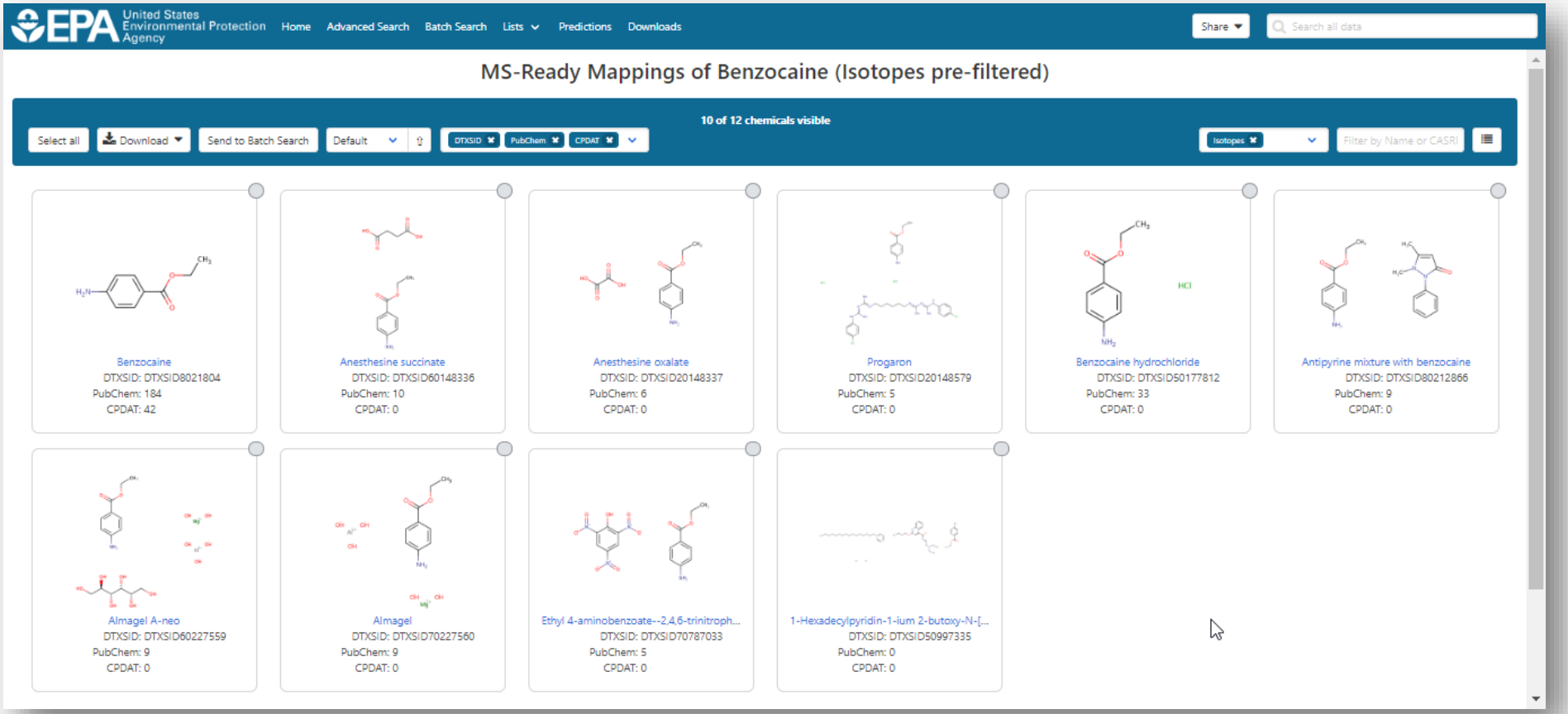
MS-Ready Mappings from Details Page



Two MS-Ready Mappings Set

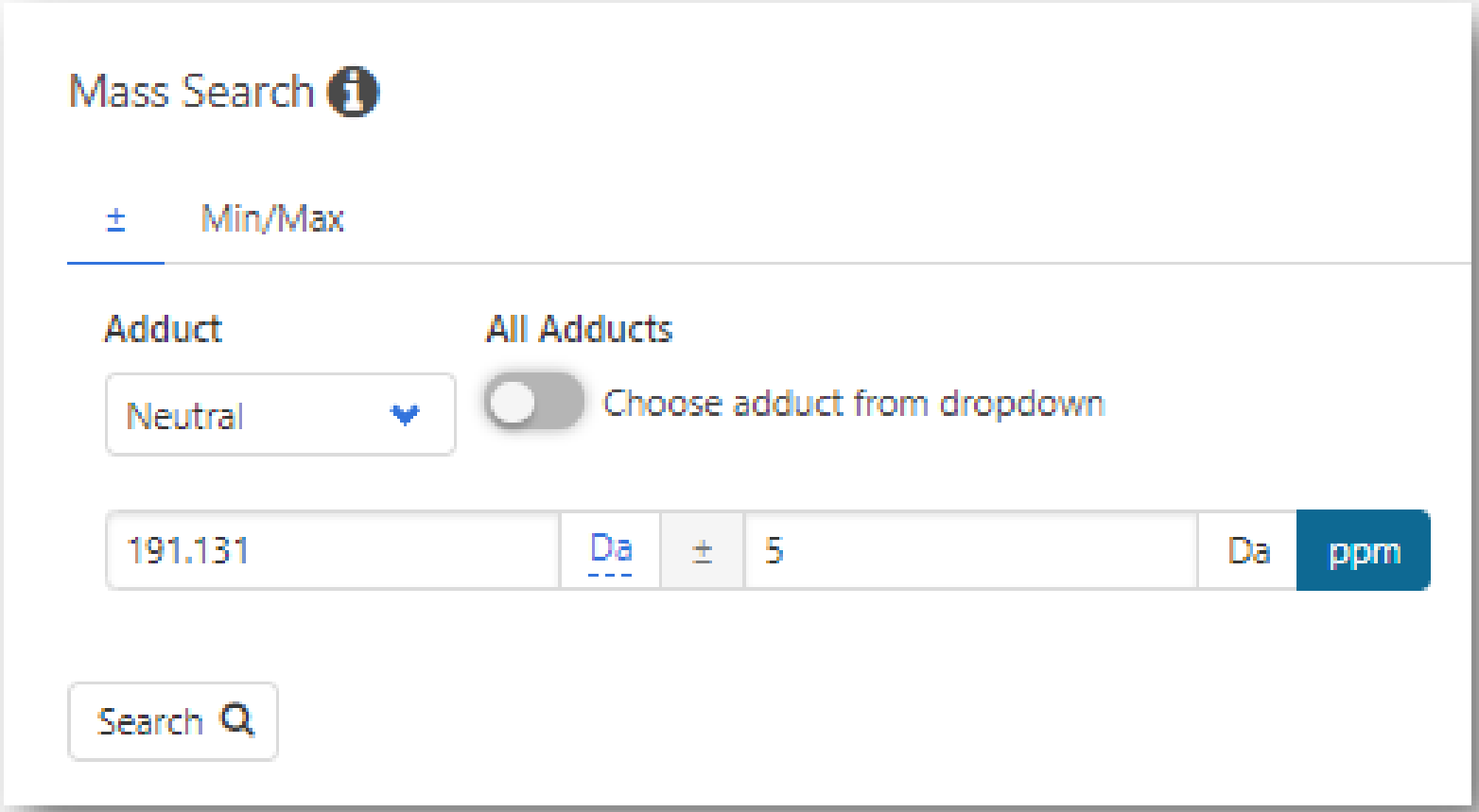


MS-Ready Mappings Set
All substances containing component

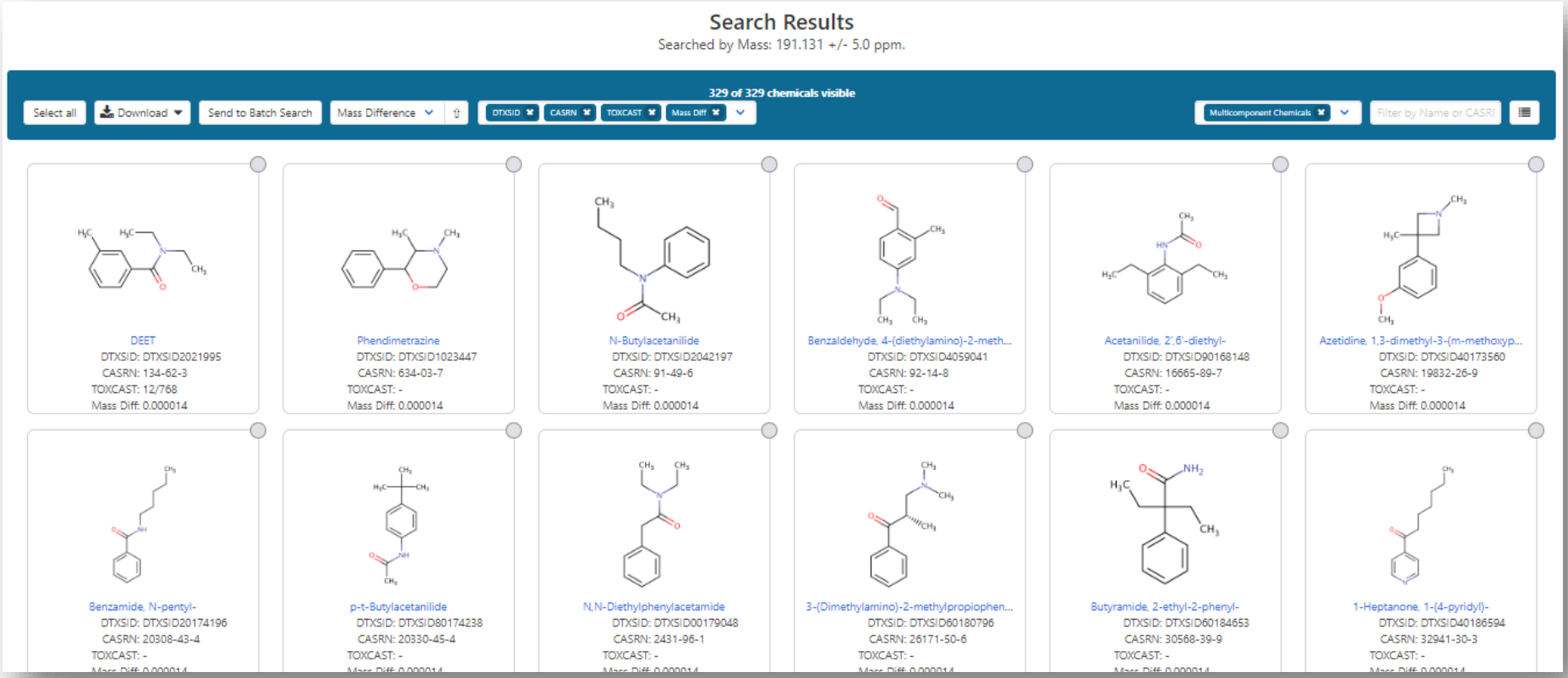


Mass and Formula Searching

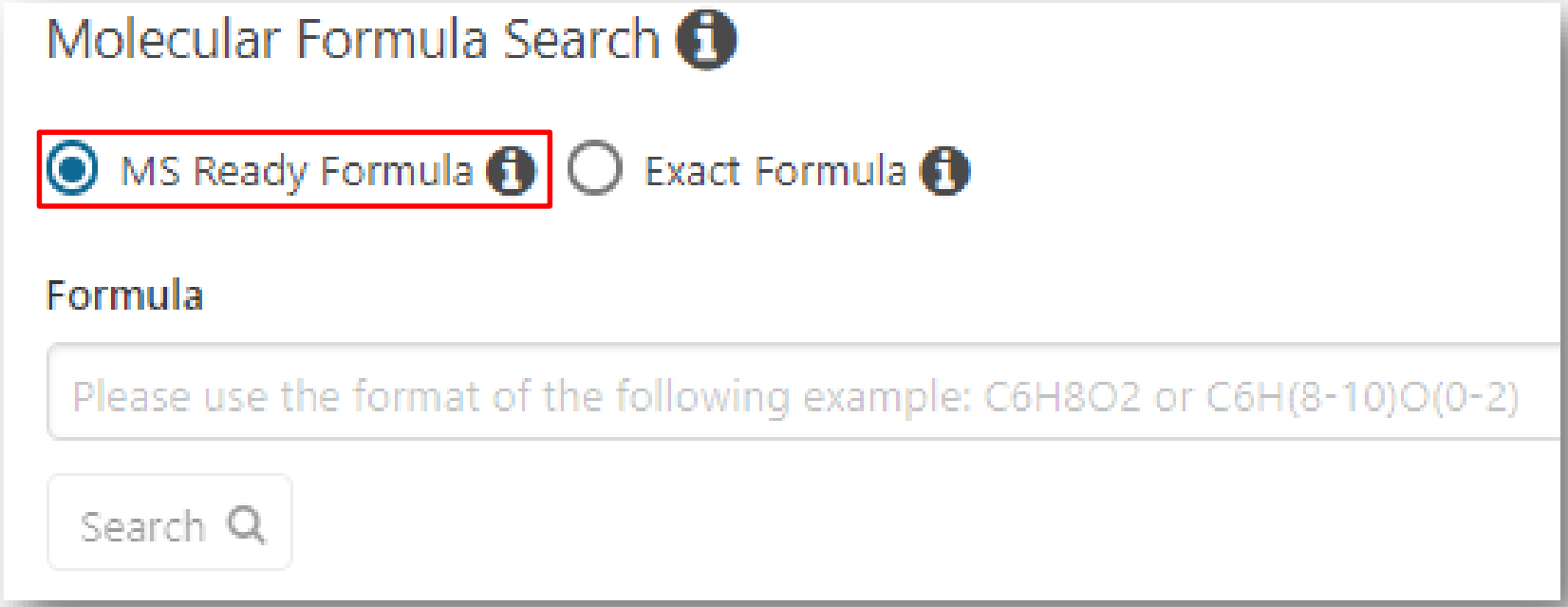
Advanced Searches
Mass Search



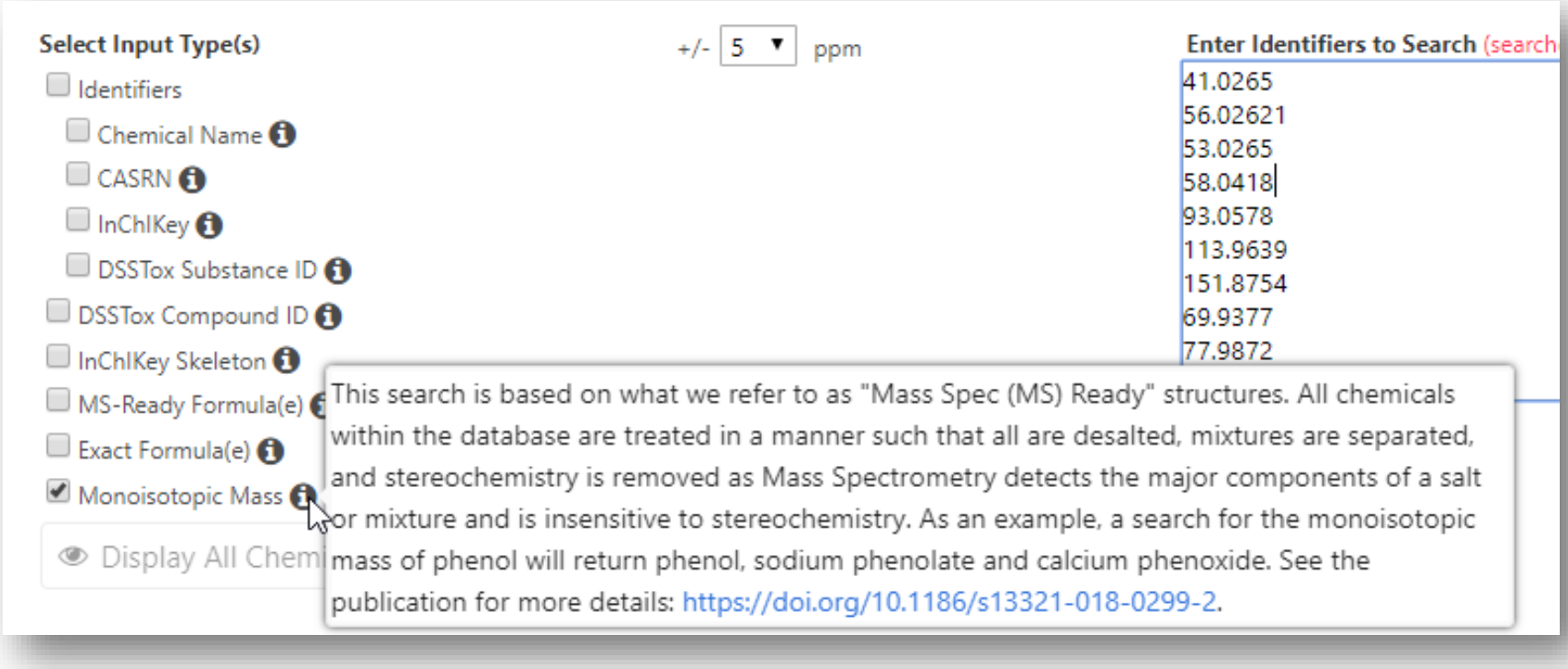
Advanced Searches
Mass Search



MS-Ready Structures for **Formula Search**

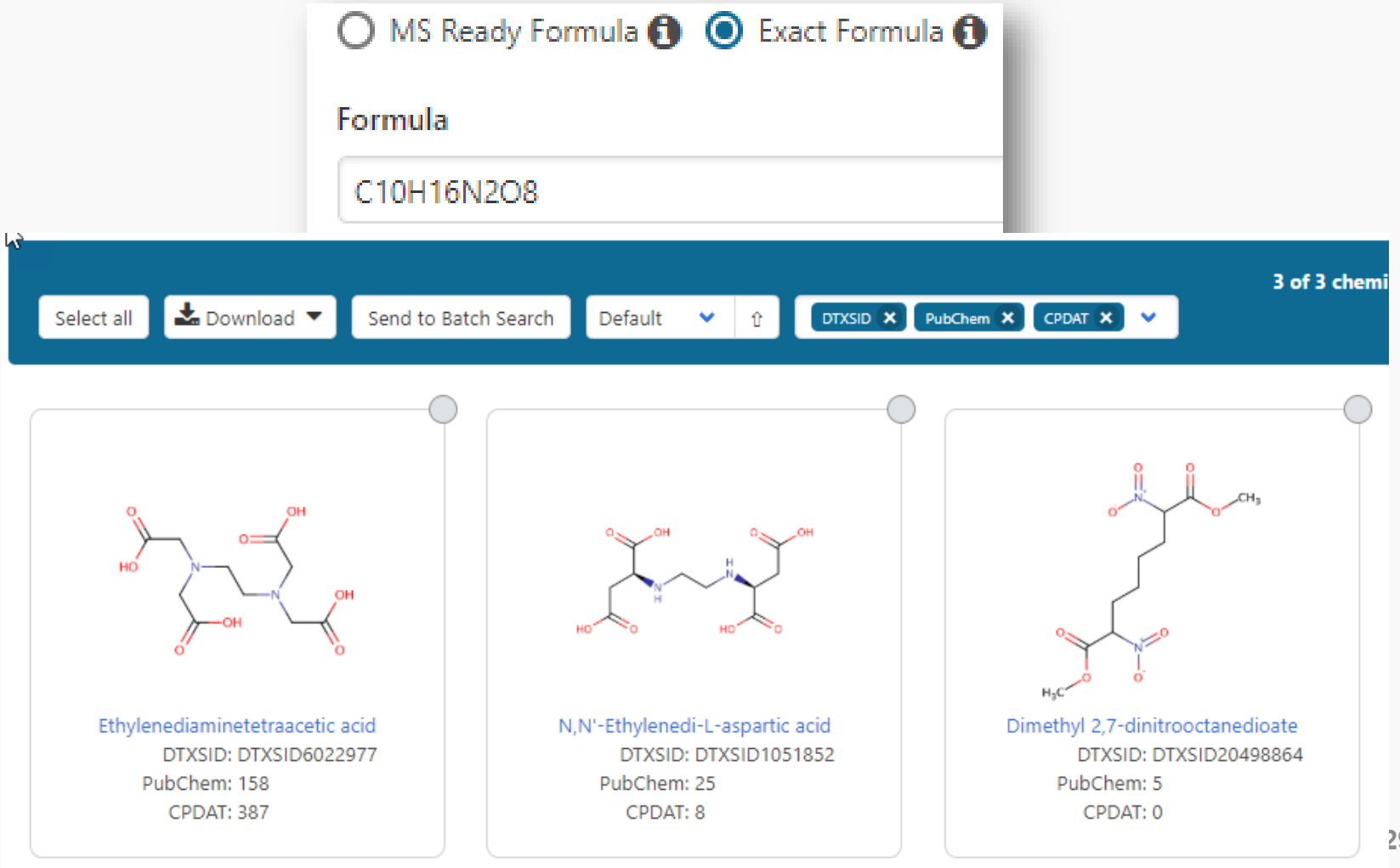


MS-Ready Structures Batch Searches



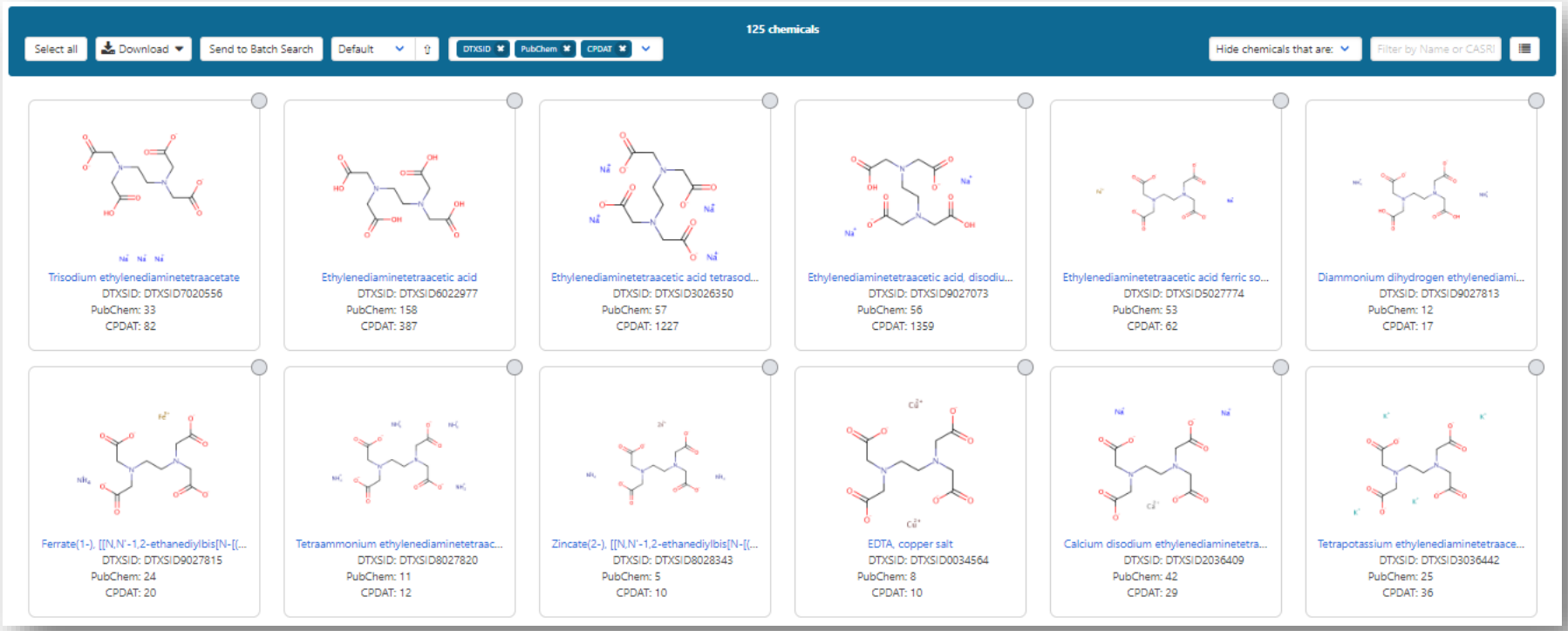
MS-Ready Mappings

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



MS-Ready Mappings

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



MS-Ready Mappings

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

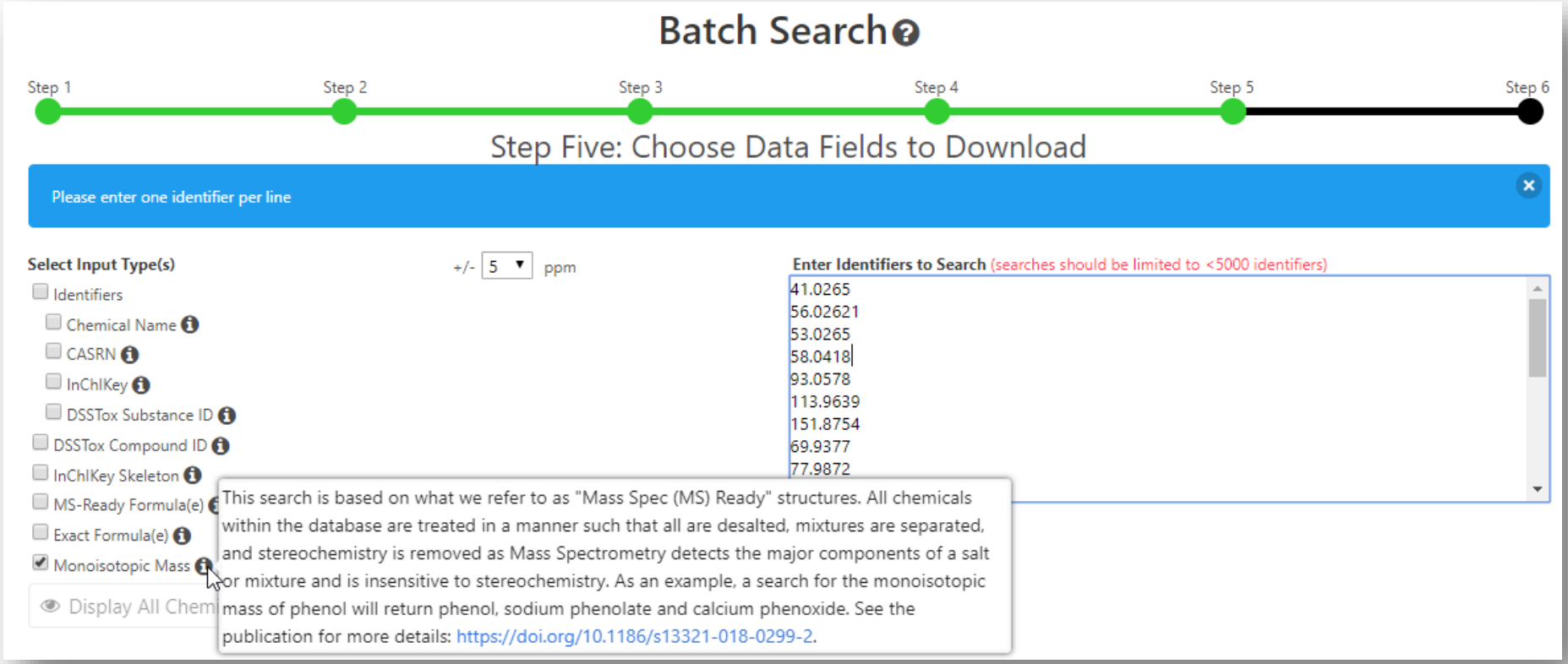
Batch Searching mass and formula

Batch Searching

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

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

Batch Searching Formula/Mass



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

EPA United States Environmental Protection Agency

	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	20
5	C14H22N2O3	DTXSID1045407	13171-25-0	Trimetazidine dihydrochloride	C14H24Cl2N2O3	338.116398	19
6	C14H22N2O3	DTXSID0045753	56715-13-0	R-(+)-Atenolol	C14H22N2O3	266.163042576	19
7	C14H22N2O3	DTXSID2048531	5011-34-7	Trimetazidine	C14H22N2O3	266.163042576	14
8	C14H22N2O3	DTXSID10239405	93379-54-5	Esatenolol	C14H22N2O3	266.163042576	12
9	C14H22N2O3	DTXSID50200634	52662-27-8	N-(2-Diethylaminoethyl)-2-(4-hydroxyphenoxy)acetamide	C14H22N2O3	266.163042576	7
10	C14H22N2O3	DTXSID4020111	51706-40-2	dl-Atenolol hydrochloride	C14H23ClN2O3	302.1397203	6
11	C14H22N2O3	DTXSID1068693	51963-82-7	Benzenamine, 2,5-diethoxy-4-(4-morpholinyl)-	C14H22N2O3	266.163042576	5
12	C18H34N2O6S	DTXSID3023215	154-21-2	Lincomycin	C18H34N2O6S	406.213757997	35
13	C18H34N2O6S	DTXSID7047803	859-18-7	Lincomycin hydrochloride	C18H35ClN2O6S	442.1904357	22
14	C18H34N2O6S	DTXSID20849438	1398534-62-7	PUBCHEM_71432748	C18H35ClN2O6S	442.1904357	1
15	C10H12N2O	DTXSID1047576	486-56-6	Cotinine	C10H12N2O	176.094963014	40
16	C10H12N2O	DTXSID8075330	50-67-9	Serotonin	C10H12N2O	176.094963014	22
17	C10H12N2O	DTXSID8044412	2654-57-1	4-Methyl-1-phenylpyrazolidin-3-one	C10H12N2O	176.094963014	18
18	C10H12N2O	DTXSID80165186	153-98-0	Serotonin hydrochloride	C10H13ClN2O	212.0716407	11
19	C10H12N2O	DTXSID2048870	29493-77-4	(4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine	C10H12N2O	176.094963014	10
20	C10H12N2O	DTXSID10196105	443-31-2	6-Hydroxytryptamine	C10H12N2O	176.094963014	9
21	C10H12N2O	DTXSID90185693	31822-84-1	1,4,5,6-Tetrahydro-5-phenoxypyrimidine	C10H12N2O	176.094963014	7
22	C10H12N2O	DTXSID40178777	2403-66-9	2-Benzimidazolepropanol	C10H12N2O	176.094963014	7
23	C10H12N2O	DTXSID80157026	13140-86-8	N-Cyclopropyl-N'-phenylurea	C10H12N2O	176.094963014	6
24	C10H12N2O	DTXSID30205607	570-14-9	4-Hydroxytryptamine	C10H12N2O	176.094963014	6
25	C14H18N4O3	DTXSID5023900	17804-35-2	Benomyl	C14H18N4O3	290.137890456	68
26	C14H18N4O3	DTXSID3023712	738-70-5	Trimethoprim	C14H18N4O3	290.137890456	51
27	C14H18N4O3	DTXSID40209671	60834-30-2	Trimethoprim hydrochloride	C14H19ClN4O3	326.1145682	8
28	C14H18N4O3	DTXSID70204210	55687-49-5	Benzenemethanol, 4-((2,4-diamino-5-pyrimidinyl)methyl)-2,	C14H18N4O3	290.137890456	5
29	C14H18N4O3	DTXSID20152671	120075-57-2	6-Methoxy-4-(3-(N,N-dimethylamino)propylamino)-5,8-quin	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid--6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6026567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	60

EPA United States Environmental Protection Agency

Candidate ranking using metadata

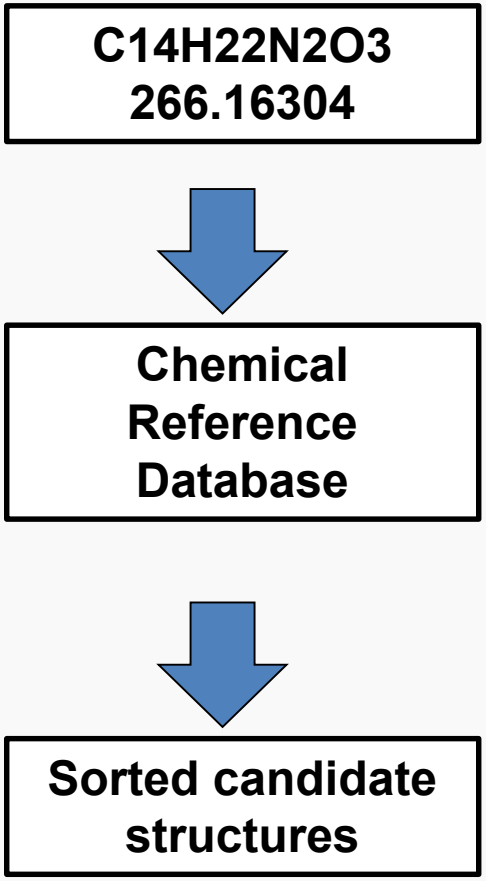
J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

RESEARCH ARTICLE

Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider

Data Source Ranking of "*known unknowns*"

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



The original ChemSpider work

Compound class	Number in class	Average rank	Number of compounds in each position rank-ordered				
			#1	#2	#3	#4	#5+
Pharmaceutical drug	72	1.4	55	9	6	2	
Industrial chemicals	42	5.5	28	6	3		5
Personal care products	8	6.1	3	1			4
Steroid hormones	7	1.0	7				
Perfluorochemicals	6	1.2	5	1			
Pesticides	12	2.3	6	2	3		1
Veterinary drugs	3	1.3	2	1			
Dyes	2	1.0	2				
Food product/natural compounds	4	3.8	2			1	1
Illicit drugs	2	2.0	1		1		
Misc. molecules	3 [a]	1.3	2	1			

Is a bigger database better?

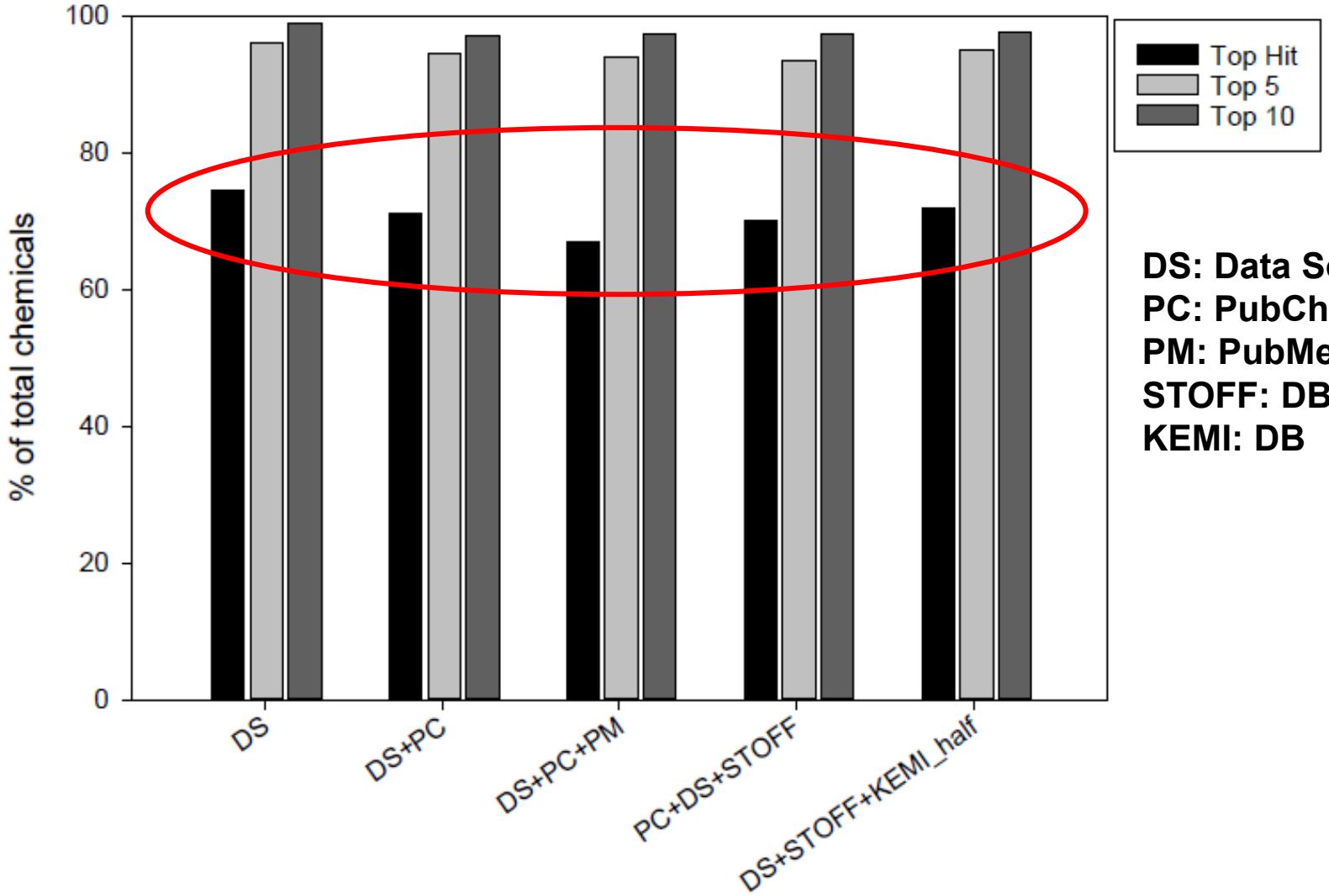
- ChemSpider was 26 million chemicals for the original work
- Much BIGGER today
- Is bigger better??
- Are there other metadata to use for ranking?

Using Metadata for Ranking

- Chosen dashboard metadata to rank candidates
  - Associated data sources
    - Lists in the underlying database (more about lists later)
    - Associated data sources in PubChem
    - Specific source types (e.g. water, surfactants, pesticides)
  - Number of associated literature articles (Pubmed)
  - **Chemicals in the environment** – the number of products/categories containing the chemical is an important source of data (from CPDat database)

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

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



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

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

How did performance compare?

Summary statistics and rank-ordered position in the CompTox Chemistry Dashboard and ChemSpider of the 89 compound subset from the Little et al. [7] study

		Average rank	Number in each position rank-ordered				
		(±SD)	#1	#2	#3	#4	#5+
Mass-based	Dashboard	1.2 ± 0.7	77[a]	5	3	3	
	ChemSpider	2.2 ± 6.1[b]	68	8	7	1	5
Formula-based	Dashboard	1.1 ± 0.4	78[a]	8	2		
	ChemSpider	1.3 ± 1.0	77	8	2	1	2

[a]One chemical (tephrosin) not present in the Dashboard

Data Quality is important

- Data quality in free web-based databases!

Drug Discovery Today

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

ELSEVIER

Review

Keynote

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

Antony J.

Show

https://d

Drug Discovery Today

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

ELSEVIER

Editorial

es

and content

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

Alex M Clark, Antony J Williams and Sean Ekins

Journal of Cheminformatics 2015 **7**:9

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

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

Public Databases require curation

- There is significant bloating in the public databases because of lack of curation
- The number of hits retrieved based on mass or formula searching can explode based on poorly represented chemicals – especially stereochemistry issues
- MS-Ready structures will map back to multiple versions of “the same chemical”.

Will the correct Microcystin LR Stand Up? ChemSpider Skeleton Search

EPA United States Environmental Protection Agency

Matches any text strings used to describe a molecule.

ZYZCGGRZINLQBL

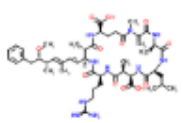
Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID

FILTER

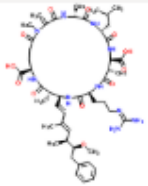
Search Hits Limit: 100

Found 9 results

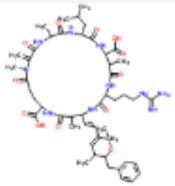
Search term: **ZYZCGGRZINLQBL** (Found by InChIKey (skeleton match))



Cyanoginosin

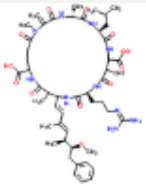


MCYST-LR

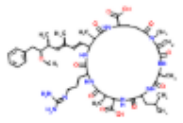


15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-

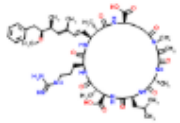
(5R,8S,11R,12S,15S,18R,19S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-3,5-



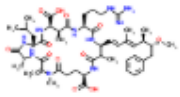
(5R,8S,15S)-15-{3-[(Diaminomethylene)amino]propyl}-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-



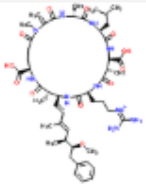
15-{3-[(Diaminomethylene)amino]propyl}-8-isobutyl-18-(6-methoxy-3,5-dimethyl-7-



(5R,8S,11R,12S,15S,18S,19S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,5S,6S)-6-methoxy-3,5-dimethyl-

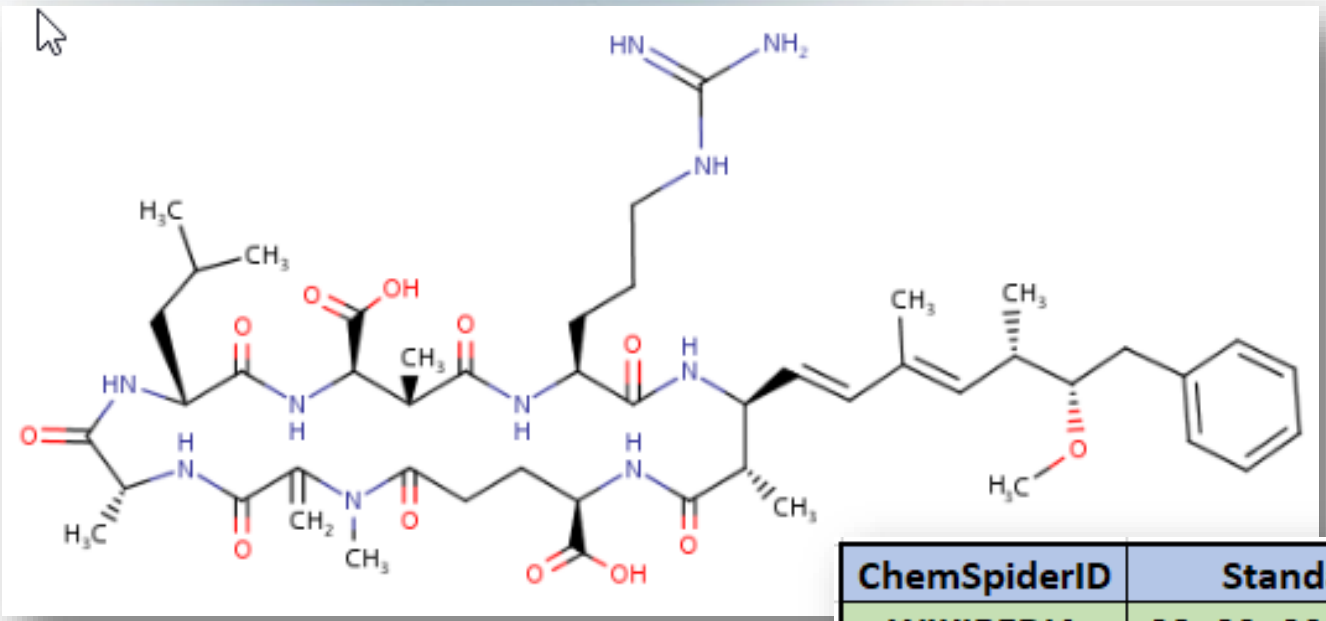


(5R,8R,11R,12S,15S,18S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5R,6R)-6-methoxy-3,5-dimethyl-



Diamino-N-(3-{(5R,8S,11R,12S,15S,18S,19S,22R)-11,22-dicarboxy-8-isobutyl-18-

Comparing ChemSpider Structures

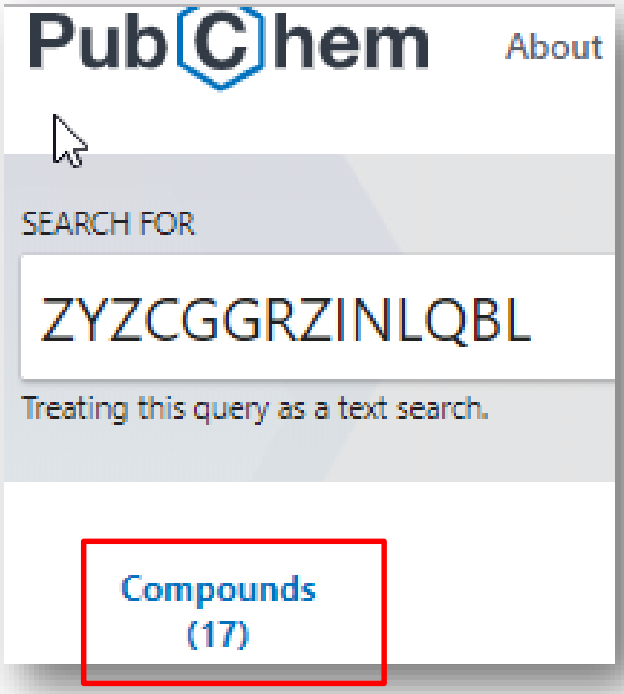


ChemSpiderID	Standard InChIKey Stereolayer
WIKIPEDIA	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
CompTox	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
4941647	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
393078	t28-,29-,30-,31+,34-,35-,36+,**37-**,38-,40+
57618348	t28-,29-,30-,31+,34-,35-,36+,**37-**,38-,40+
29342071	t28-,29-,30-,31+,**34+**,35-,36+,**37-**,38-,40+
7987594	t28-,**29?**,**30?**,31+,**34?**,35-,**36?**,**37-**,38-,**40?**
22900854	t28-,**29?**,**30+**,**31-**,**34+**,**35+**,**36-**,**37-**,38-,**40-**
19692240	NONE
2831283	NONE

Comparing ChemSpider Structures

ChemSpiderID	InChIKey	# Stereocenters	# Different
WIKIPEDIA	ZYZCGGRZINLQBL-JCGNTXOTSA-N	10/10	0
CompTox	ZYZCGGRZINLQBL-JCGNTXOTSA-N	10/10	0
4941647	ZYZCGGRZINLQBL-JCGNTXOTSA-N	10/10	0
393078	ZYZCGGRZINLQBL-GWRQVWKTSA-N	10/10	1
57618348	ZYZCGGRZINLQBL-UPPCHHEJSA-N	10/10	1
29342071	ZYZCGGRZINLQBL-IIJTUTQBSA-N	10/10	2
7987594	ZYZCGGRZINLQBL-BESLYTPASA-N	5/10	6
22900854	ZYZCGGRZINLQBL-QAXSDTKVSA-N	9/10	8
19692240	ZYZCGGRZINLQBL-ORZJCNCZSA-N	0/10	10
2831283	ZYZCGGRZINLQBL-UHFFFAOYSA-N	0/10	10

Other Searches



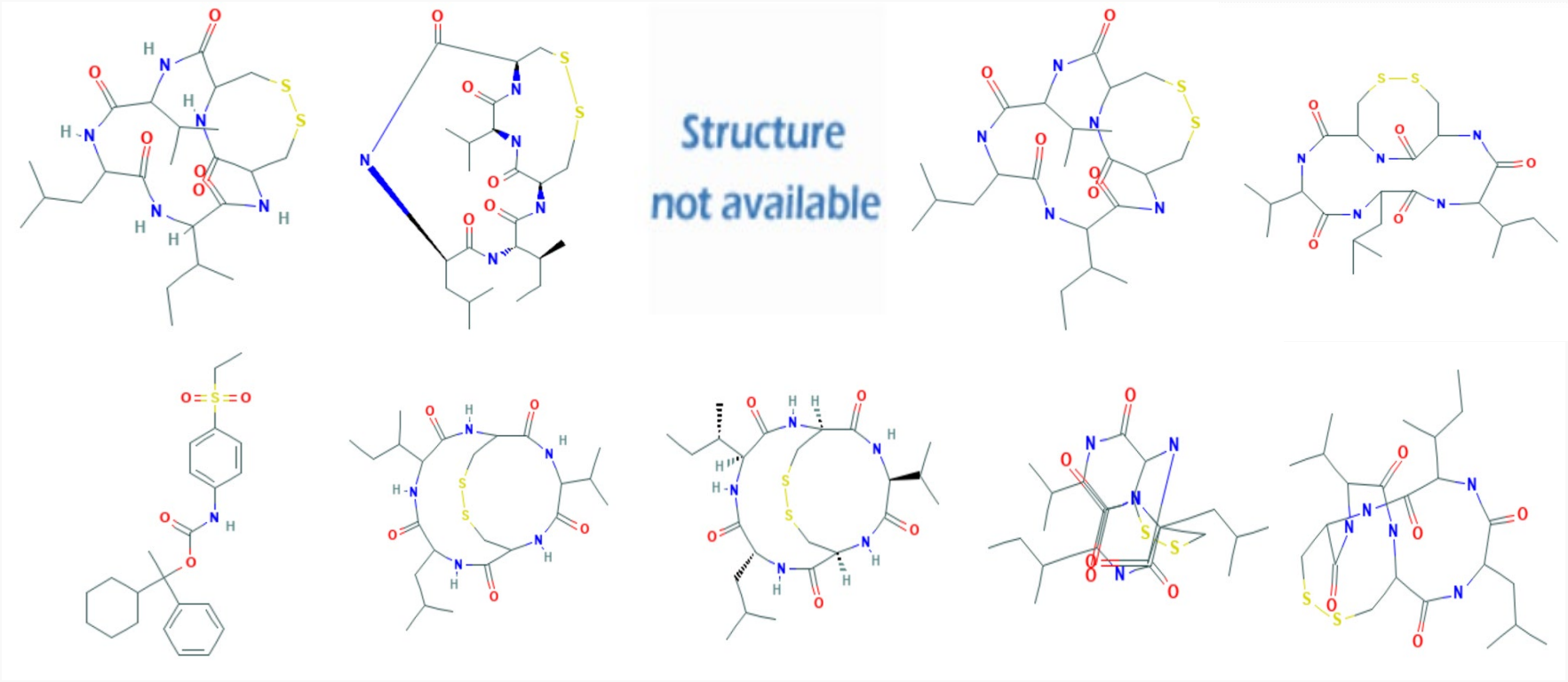
UniChem

Show All entries

CMR. Query InChI...	src_id	Source	src_compound_id
...matches...	1	ChEMBL	CHEMBL444092
...matches...	4	Guide to Pharmacology	4735
...matches...	6	KEGG Ligand	C05371
...matches...	7	ChEBI	6925
...matches...	9	ZINC	ZINC000169715525
...matches...	9	ZINC	ZINC000255288110
...matches...	9	ZINC	ZINC000255288111
...matches...	9	ZINC	ZINC000255288112
...matches...	9	ZINC	ZINC000255288113
...matches...	9	ZINC	ZINC000255288114
...matches...	9	ZINC	ZINC000255288115
...matches...	9	ZINC	ZINC000583653042
...matches...	9	ZINC	ZINC000669680403
...matches...	10	eMolecules	26754757
...matches...	10	eMolecules	31239828
...matches...	11	IBM Patent System	DA3C2F25F29692734272194ED0E2C009
...matches...	14	FDA SRS	EQ8332842Y

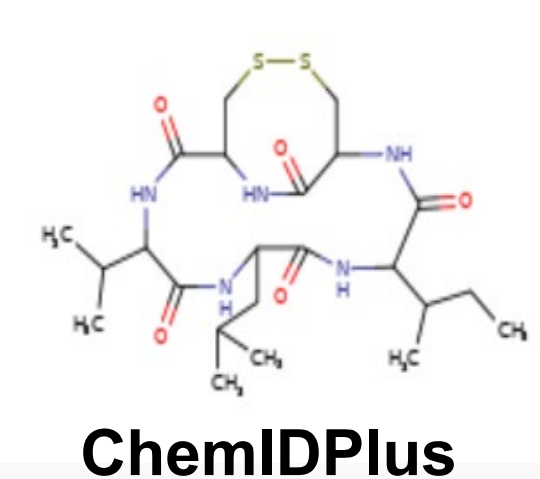
CASRN 3022-92-2 on PubChem

https://pubchem.ncbi.nlm.nih.gov/#query=3022-92-2



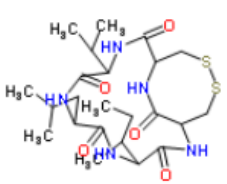
FOUR Different structures, THREE different skeletons

Comparisons…

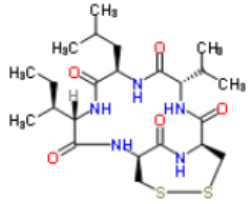


ChemIDPlus

Search term: **3022-92-2** (Found by synonym)



4-sec-Butyl-7-isobutyl-10-isopropyl-15,16-dithia-2,5,8,11,19-



Malformin A1

ChemSpider

CAS Registry Number 3022-92-2

~93 ~14

$C_{23}H_{39}N_5O_5S_2$

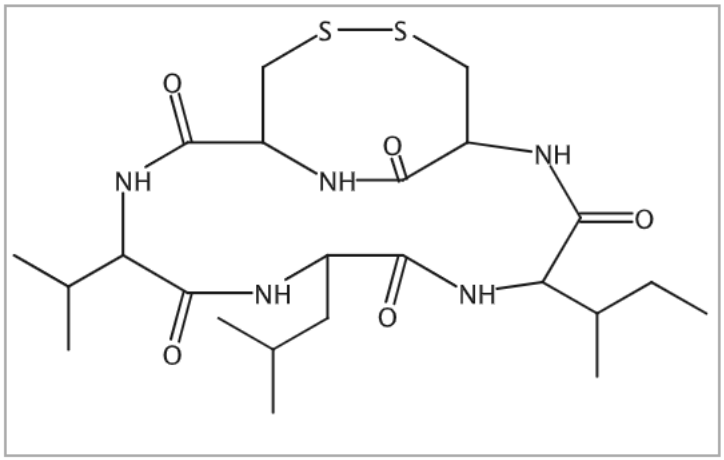
Cyclo(D-cysteinyl-D-cysteinyl-L-valyl-D-leucyl-L-isoleucyl), cyclic (1→2)-disulfide

Molecular Weight
529.72

Melting Point (Experimental)
Value: >300 °C (decomp)

Boiling Point (Predicted)
Value: 921.0±65.0 °C | Condition: Press: 760 Torr

Density (Predicted)



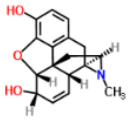
SciFinder

Morphine – 26 Results

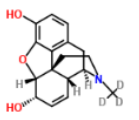
Found 26 results

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

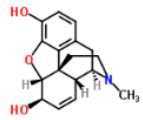
First | < | > | Last



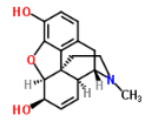
D-(-)-Morphine



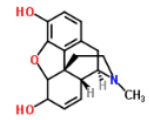
(5alpha,6alpha)-17-(~2~H_3_)Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



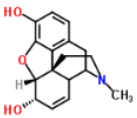
(5alpha,6beta)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



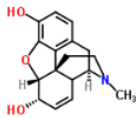
(5beta,6beta,9alpha,13alpha,14alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



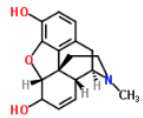
17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



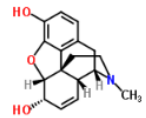
(5alpha,6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



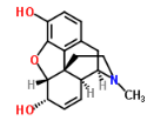
(5alpha,6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



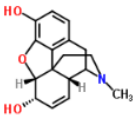
(5alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



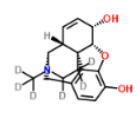
(5alpha,6alpha,9alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



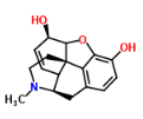
(5alpha,6alpha,14alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



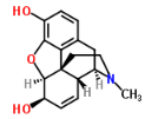
(5alpha,6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



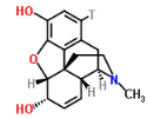
Morphine-D6 solution



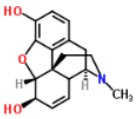
(6beta)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



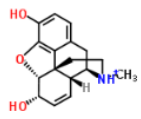
(5beta,6beta)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



(5alpha,6alpha)-17-Methyl(1-~3~H)-7,8-didehydro-4,5-epoxymorphinan-3,6-diol

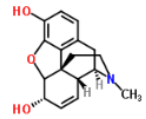


(5alpha,6beta)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol

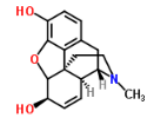


(5alpha,6alpha,17S)-3,6-Dihydroxy-17-methyl-7,8-didehydro-4,5-epoxymorphinan-17-ium

(5beta,6alpha,9alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



(6alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol



(6beta,9alpha,13alpha,14alpha)-17-Methyl-7,8-didehydro-4,5-epoxymorphinan-3,6-diol

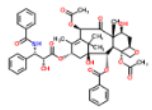
Taxol: 79 Results

EPA United States Environmental Protection Agency

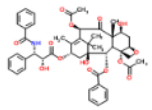
Found 79 results

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

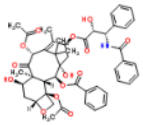
First | < | > | Last



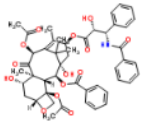
paclitaxel



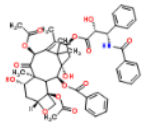
(2alpha,3xi,5beta,7beta,10beta,13alpha)-4 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



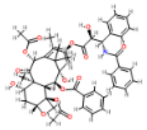
(2alpha,5beta,7alpha,10beta,13alpha)-4,1 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



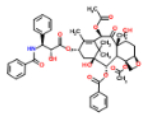
(5beta,7beta,10alpha,13alpha)-4,10-bis(acetyloxy)-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



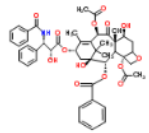
(2alpha,3xi,5beta,7beta,10alpha,13alpha)-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



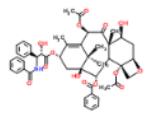
NSC-LSC-1



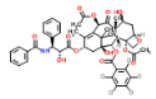
(2alpha,3xi,5beta,7alpha,10beta,13alpha)-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



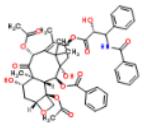
(2alpha,3xi,7beta,10beta,13alpha)-4,10-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



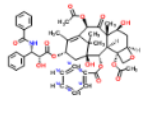
(2alpha,3xi,5beta,7beta,10beta,13alpha)-4 Diacetoxy-13-{[(2S,3S)-3-(benzoylamino)-2-hydroxy-3-



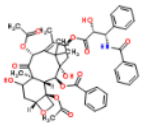
(2alpha,3beta,5beta,7beta,10beta,13alpha Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



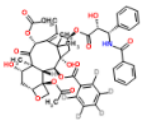
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2R)-3-(benzoylamino)-2-hydroxy-3-phenylpropanoyl]oxy}-1,7-



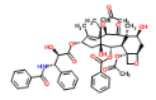
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



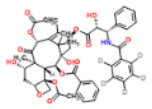
(2alpha,5beta,10beta,13alpha)-4,10-Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



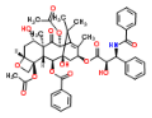
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



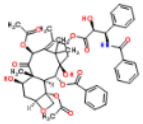
(2alpha,5beta,7beta,8alpha,10beta,13alph Diacetoxy-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



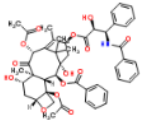
(2alpha,5beta,7beta,10beta,13alpha)-4,10 Diacetoxy-1,7-dihydroxy-13-{[(2R,3S)-2-hydroxy-3-phenyl-3-



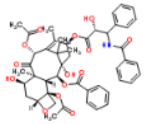
(1b,2a,5b,7b,10b,13a)-4,10-bis(acetyloxy)-13-{[(2R,3S)-3-(benzoylamino)-2-hydroxy-3-



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



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



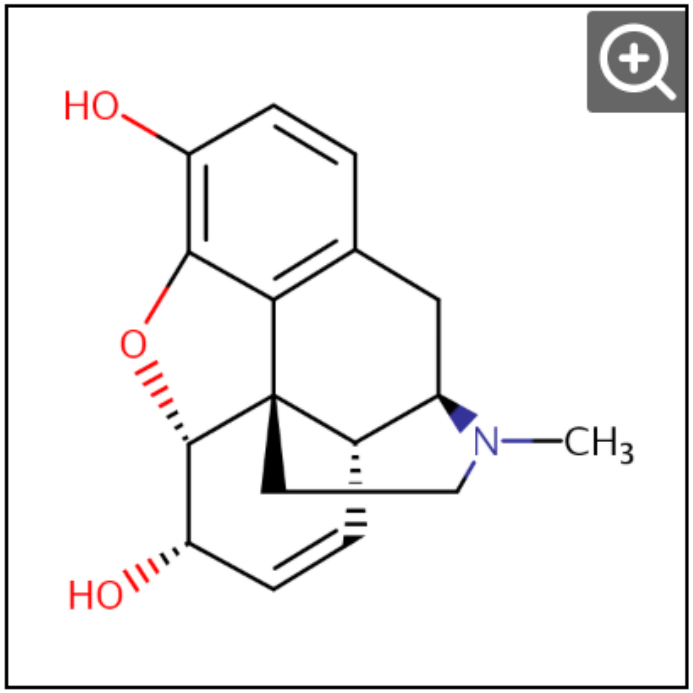
(2alpha,5beta,7alpha,10beta,13alpha)-4,1 Diacetoxy-13-{[(2R)-3-(benzoylamino)-2-hydroxy-3-phenylpropanoyl]oxy}-1,7-

Massive Dilution Effect

- Public databases suffer an enormous dilution effect based on data quality
- Our efforts to curate data provide a smaller but higher quality database – e.g. not registering MOD chemicals from vendors
- Ensuring quality synonym and registry number mappings gives cleaner mappings to public resources (e.g. PubMed)

Additional Mass-Spec Functionality

Formula-Based Search



Wikipedia

Morphine is a pain medication of the opiate family which is found naturally in a number of plants and animals. It acts directly on the central nervous system (CNS) to decrease the feeling of pain. It can be taken for both acute pain and chronic pain. It is frequently used for pain from myocardial infarction and during labor. It can be given by mouth, by injection into a muscle, by injection under the skin, intravenously, injection into the space around the

...

Read more

Intrinsic Properties

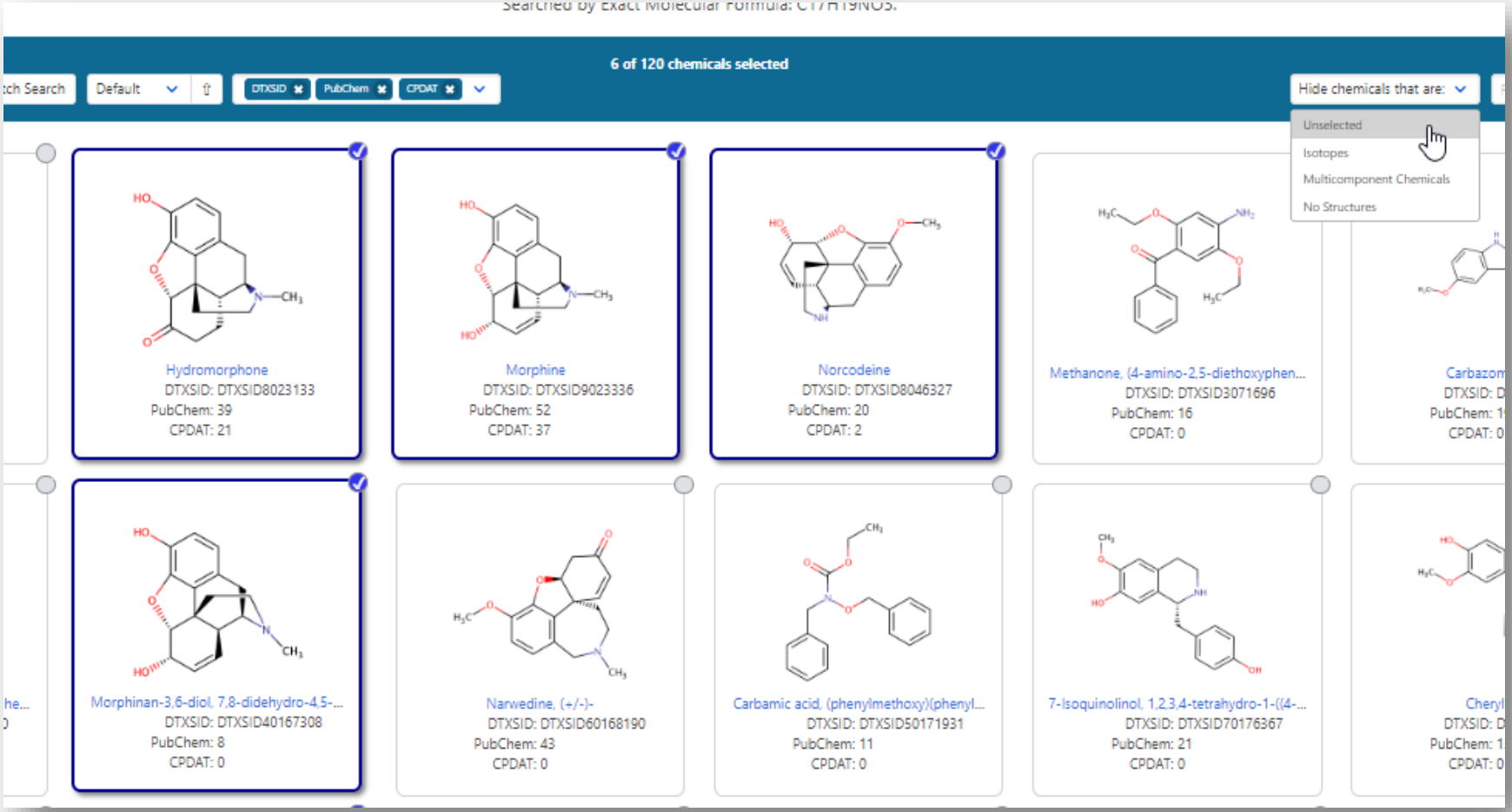
Molecular Formula: $C_{17}H_{19}NO_3$ Mol File Find All Chemicals

Average Mass: 285.343 g/mol Isotope Mass Distribution

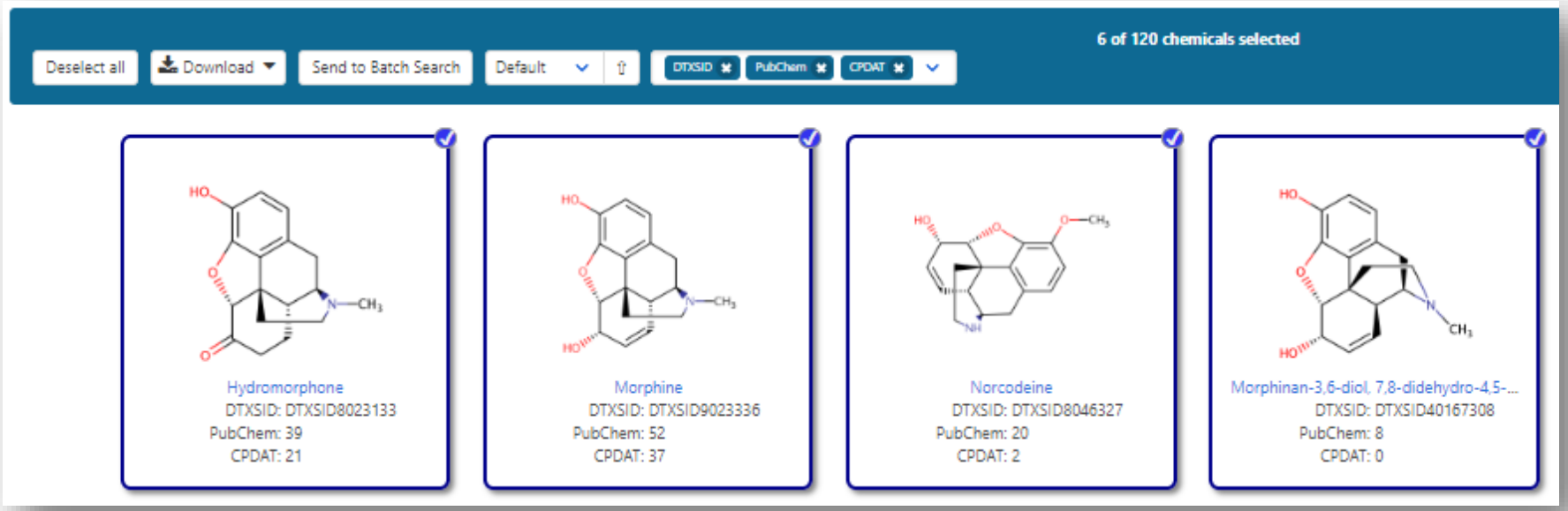
Monoisotopic Mass: 285.136493 g/mol

Structural Identifiers

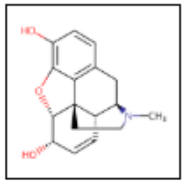
Select Chemicals of Interest



Prune to list of interest



Literature Searching



Morphine

57-27-2 | DTXSID9023336

Searched by Approved Name.

Abstract Sifter

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

Select a Query Term

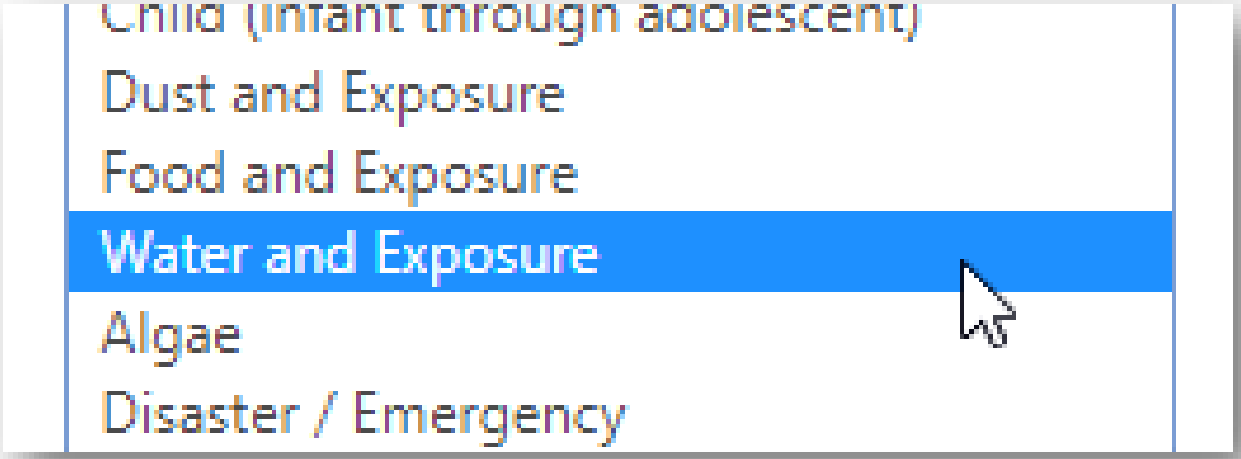
Retrieve Articles

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

Optionally, edit the query before retrieving.

"57-27-2" OR "Morphine"

Literature Searching



Optionally, edit the query before retrieving.

("57-27-2" OR "Morphine") AND ((water OR groundwater OR drinking water) AND Environmental Exposure)

Literature Searching

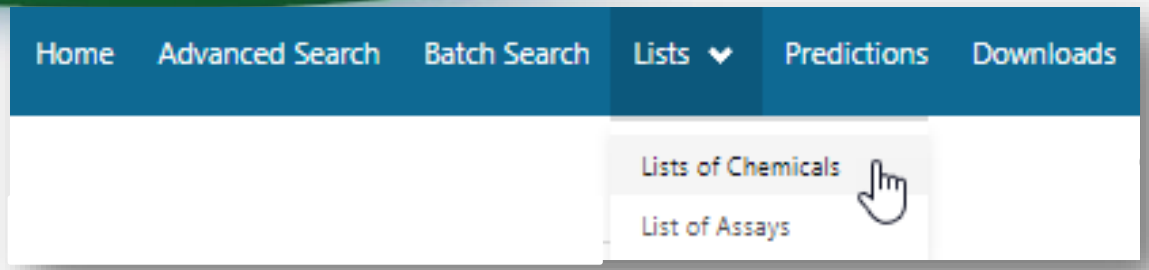
To find articles quickly, enter terms to sift abstracts.

wastewater | Spectrometry | EPA | Clear Terms | Download / Send to... | Download Sifter for Excel

	wastewater	Spectrometry	EPA	Total	PMID	Year	Title	Authors	Journal	Rev
	4	2	0	6	29274731	2017	Simultaneous analysis of opioid analgesics and thei...	Krizman-Matasic; Kostanjevecki; Ahel; Terzic	Journal of chromatography. A	
	0	1	0	1	25768972	2015	Evaluating external contamination of polybrominate...	Poon; Aleksa; Carnevale; Kapur; Goodyer; Koren	Therapeutic drug monitoring	
	0	1	0	1	22544551	2012	Spatial distribution of illicit drugs in surface waters o...	Vazquez-Roig; Andreu; Blasco; Morillas; Picó	Environmental science and pollution research inter...	
	1	1	0	2	20801487	2010	Analysis of llicit and illicit drugs in waste, surface an...	Berset; Brenneisen; Mathieu	Chemosphere	
	1	1	0	2	17935751	2007	Illicit drugs, a novel group of environmental contami...	Zuccato; Castiglioni; Bagnati; Chiabrando; Grassi; ...	Water research	
	2	1	1	4	17607391	2007	Using environmental analytical data to estimate lev...	Bones; Thomas; Paull	Journal of environmental monitoring : JEM	
	3	1	2	6	17180984	2006	Simultaneous determination of psychoactive drugs ...	Hummel; Löffler; Fink; Ternes	Environmental science & technology	
	6	0	0	6	30583189	2018	Assessment of drugs of abuse in a wastewater trea...	Kumar; Tscharke; O'Brien; Mueller; Wilkins; Padhye	The Science of the total environment	
	0	0	3	3	30488421	2018	Effect of enriched environment during adolescence ...	Mohammadian; Najafi; Miladi-Gorji	Developmental psychobiology	
	3	0	0	3	29574368	2018	Estimation of the consumption of illicit drugs during ...	Foppe; Hammond-Weinberger; Subedi	The Science of the total environment	
	1	0	0	1	28787791	2017	Evaluation of in-sewer transformation of selected illi...	Gao; Banks; Li; Jiang; Lai; Mueller; Thai	The Science of the total environment	
	9	0	0	9	28472697	2017	Occurrence and fate of illicit drugs and pharmaceuti...	Causanilles; Ruepert; Ibáñez; Emke; Hernández; d...	The Science of the total environment	
	0	0	0	0	28010888	2016	Dose-dependent effects of morphine on lipopolysac...	Mottaz; Schönenberger; Fischer; Eggen; Schirmer; ...	Environmental pollution (Barking, Essex : 1987)	
	0	0	0	0	27746311	2016	Effects of voluntary exercise on the viability, prolifer...	Haydari; Safari; Zarbakhsh; Bandegi; Miladi-Gorji	Neuroscience letters	
	0	0	0	0	27261879	2016	Genotoxic effects induced by the exposure to an en...	Parolini; Magni; Castiglioni; Binelli	Ecotoxicology and environmental safety	
	3	0	0	3	27179320	2016	Temporal trends in drug use in Adelaide, South Aus...	Tscharke; Chen; Gerber; White	The Science of the total environment	

Chemical Lists

Chemical Lists



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

EPAHFR: Hydraulic Fracturing

WATER|EPA; Chemicals associated with hydraulic fracturing

Search EPAHFR Chemicals

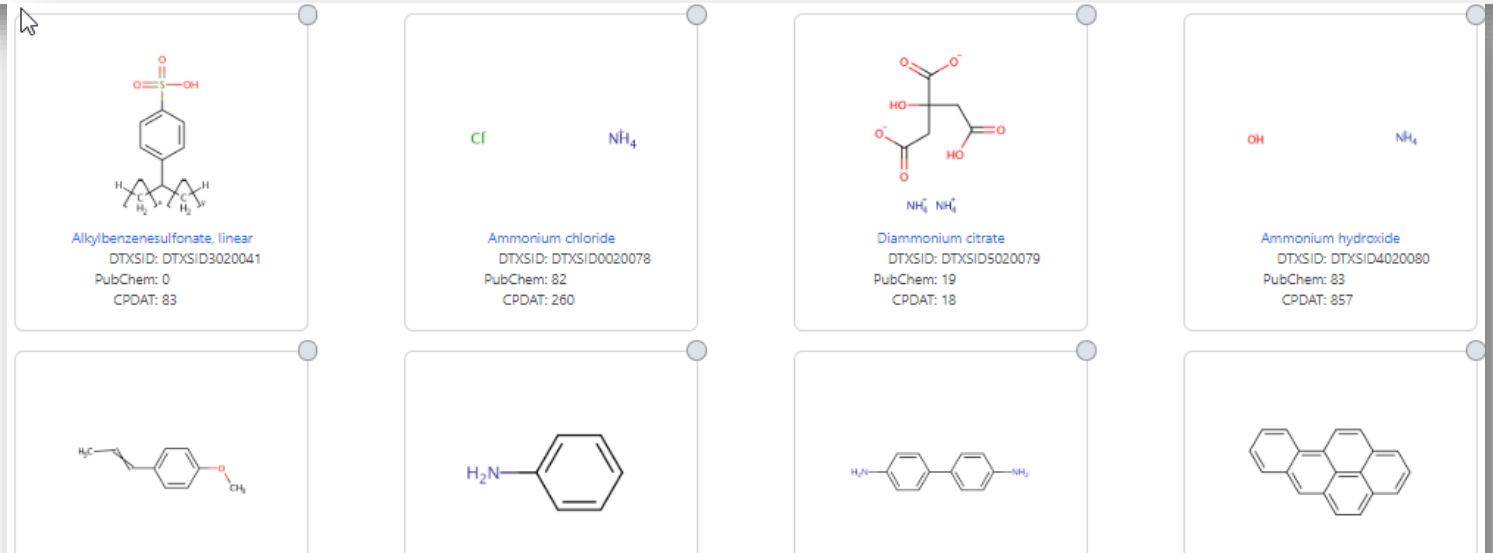
Identifier substring search

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



PFAS lists of Chemicals

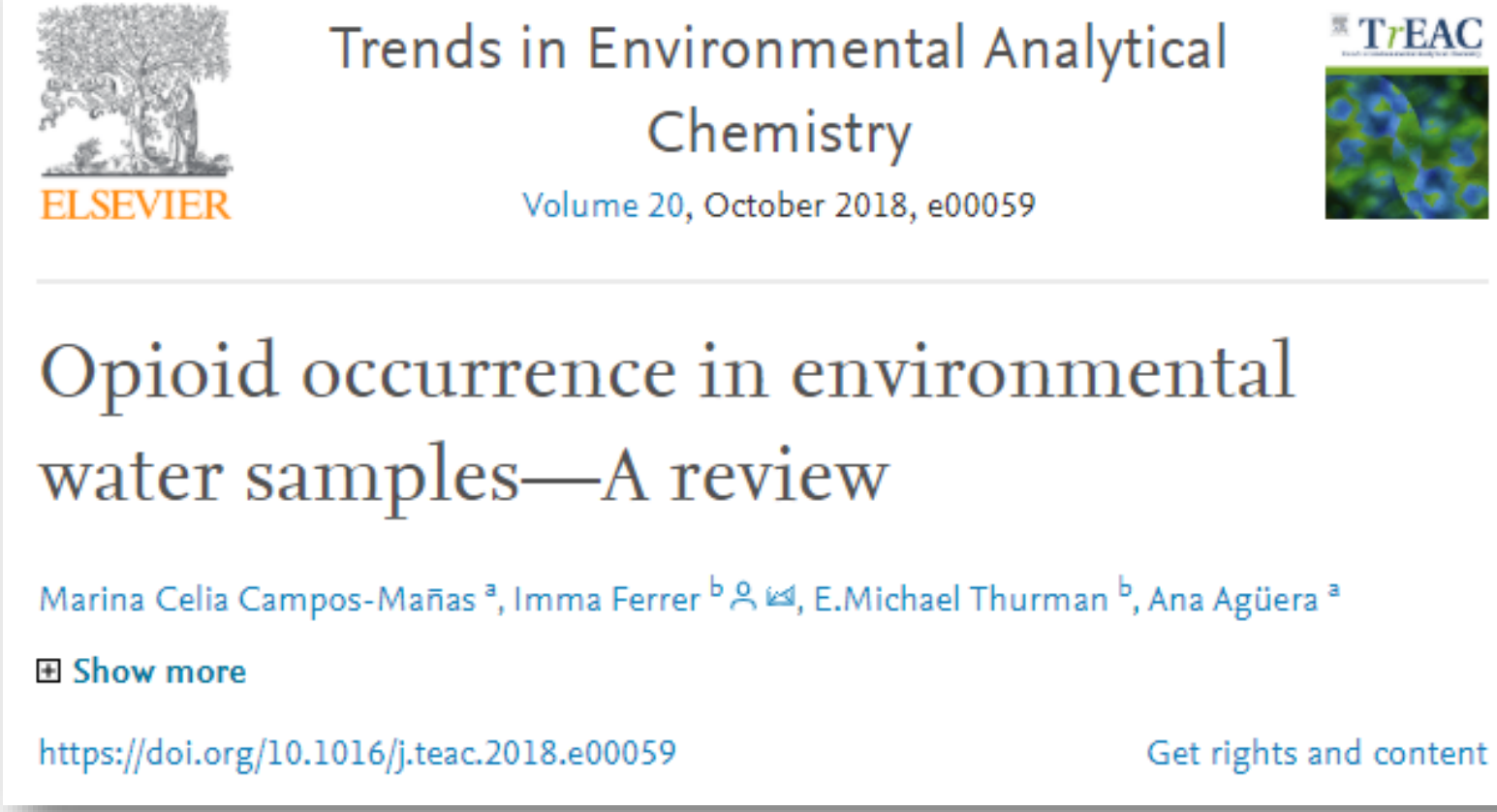
EPA United States Environmental Protection Agency

Select List

Download | Columns | PFAS | Copy Filtered Lists URL

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

List of Opioids – Presence in Lists?



ELSEVIER

Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059

TrEAC

Opioid occurrence in environmental water samples—A review

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

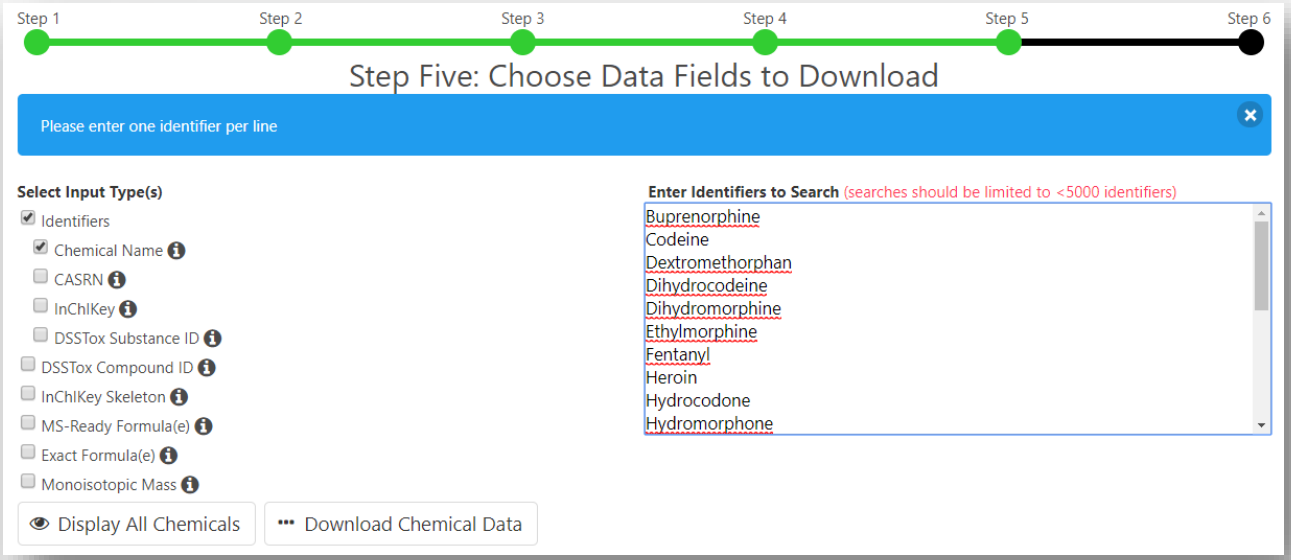
Show more

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

Get rights and content

Batch Search Names

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



Excel Download

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

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

EPA United States Environmental Protection Agency

Work in Progress

List Registration Activities

- Registering and curating numerous lists
  - NIST library of chemicals – lots of clean up to do here…
  - United States Geological Survey chemicals in water
  - Scientific Working Group for the Analysis of Seized Drugs
  - Synthetic Cannabinoids
  - Mycotoxins

Blood Exposome Curation

- Blood exposome data collection from Barupal and Fiehn…

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

…and comparing with our own “Abstract Sifter” approach

Disinfection By-products

- DBP list assembly starting with 2011 review…

> **Disinfection By-Products: Formation and Occurrence in Drinking Water** ☆
>
> **SD Richardson,** National Exposure Research Laboratory, US Environmental Protection Agency, Athens, GA, USA

…plus a number identified since then

- Anyone interested in DBPs or want to contribute?

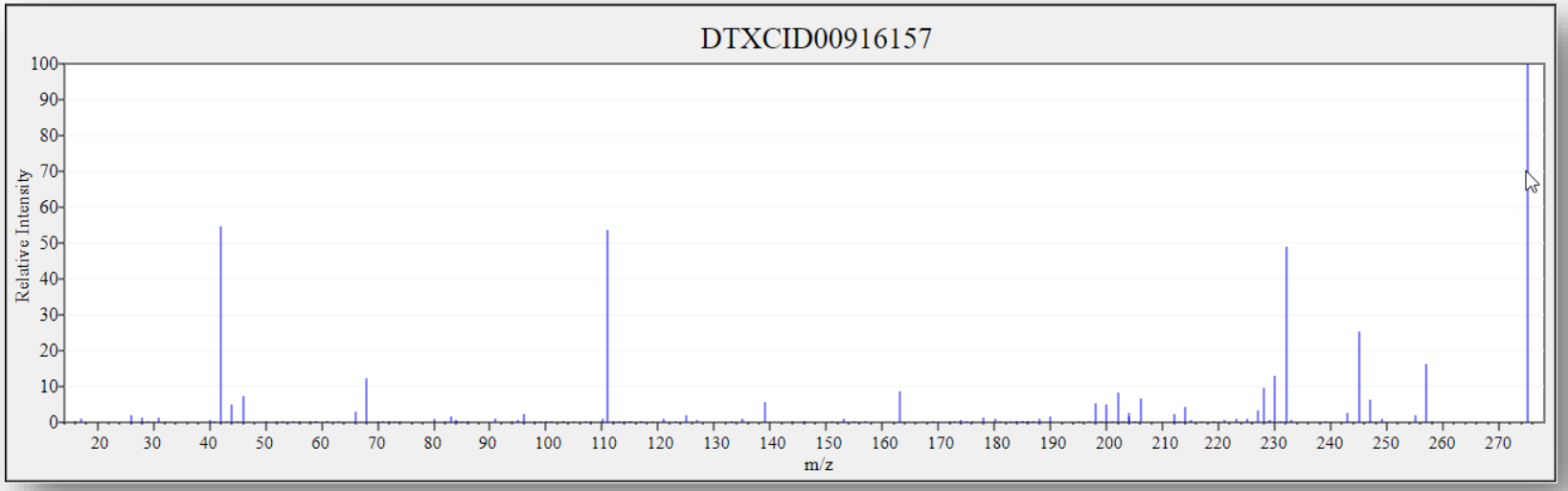
Prototype Work in Progress

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

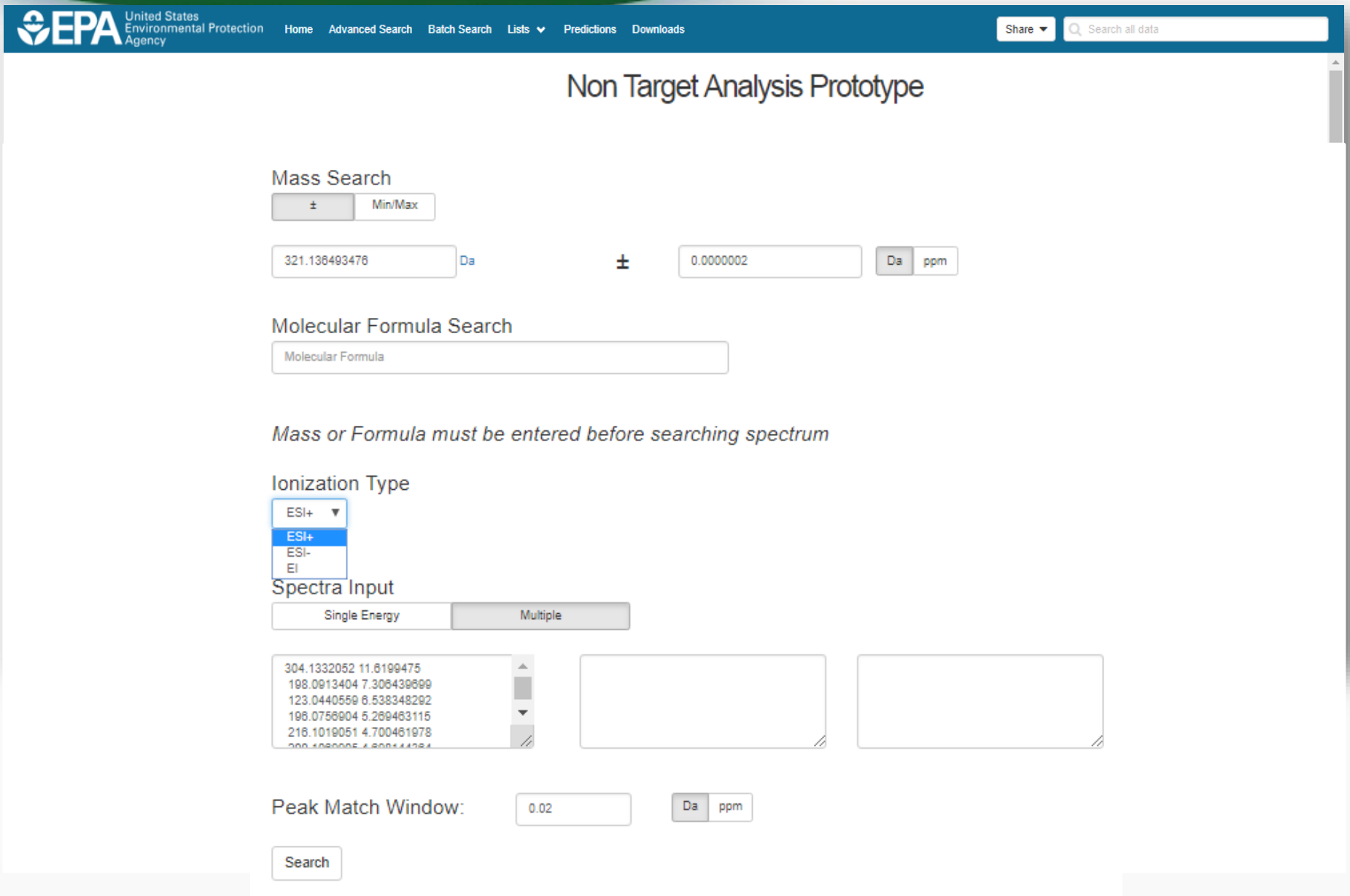
Predicted Mass Spectra

http://cfmid.wishartlab.com/

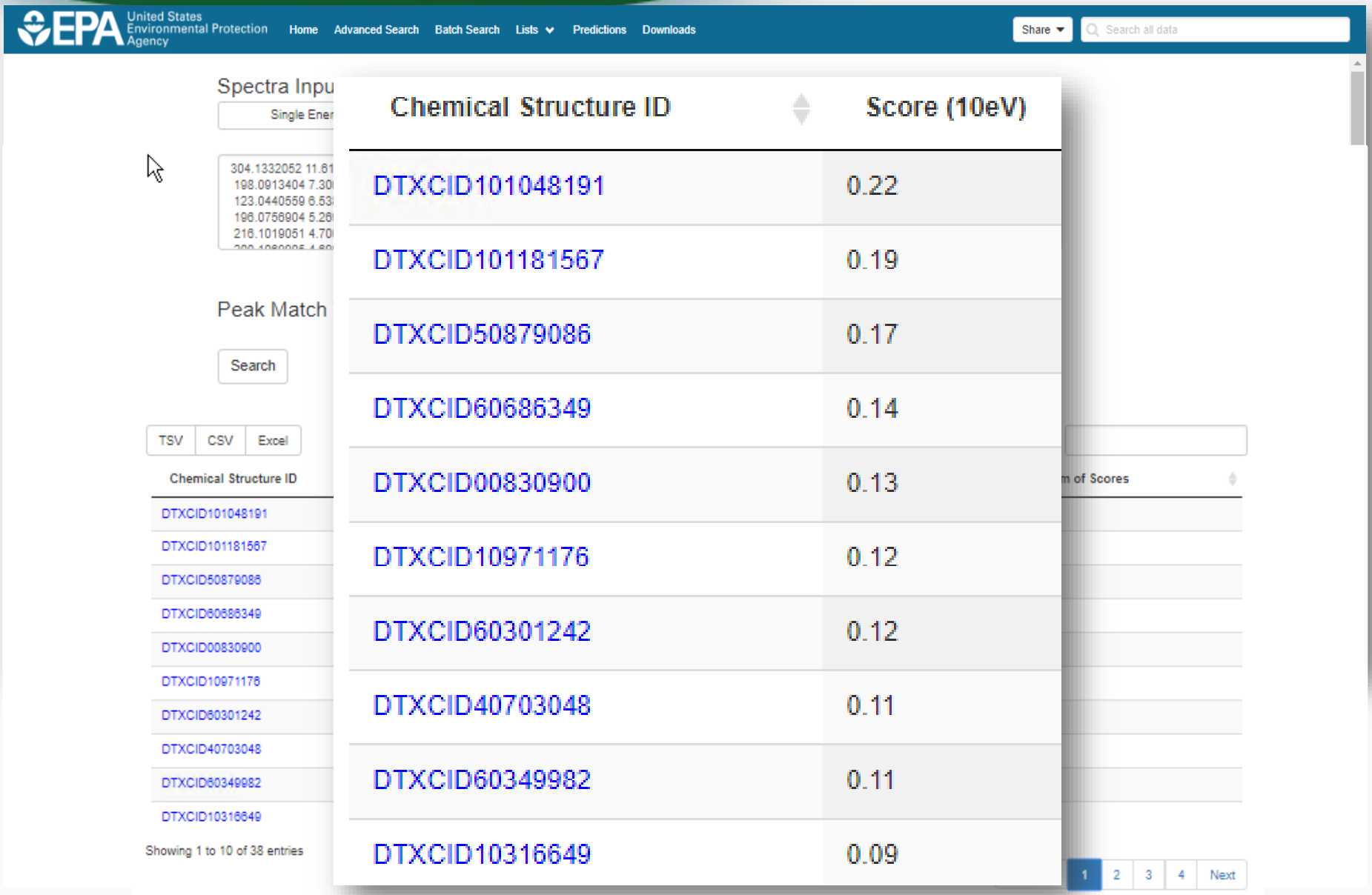
- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



Search Expt. vs. Predicted Spectra

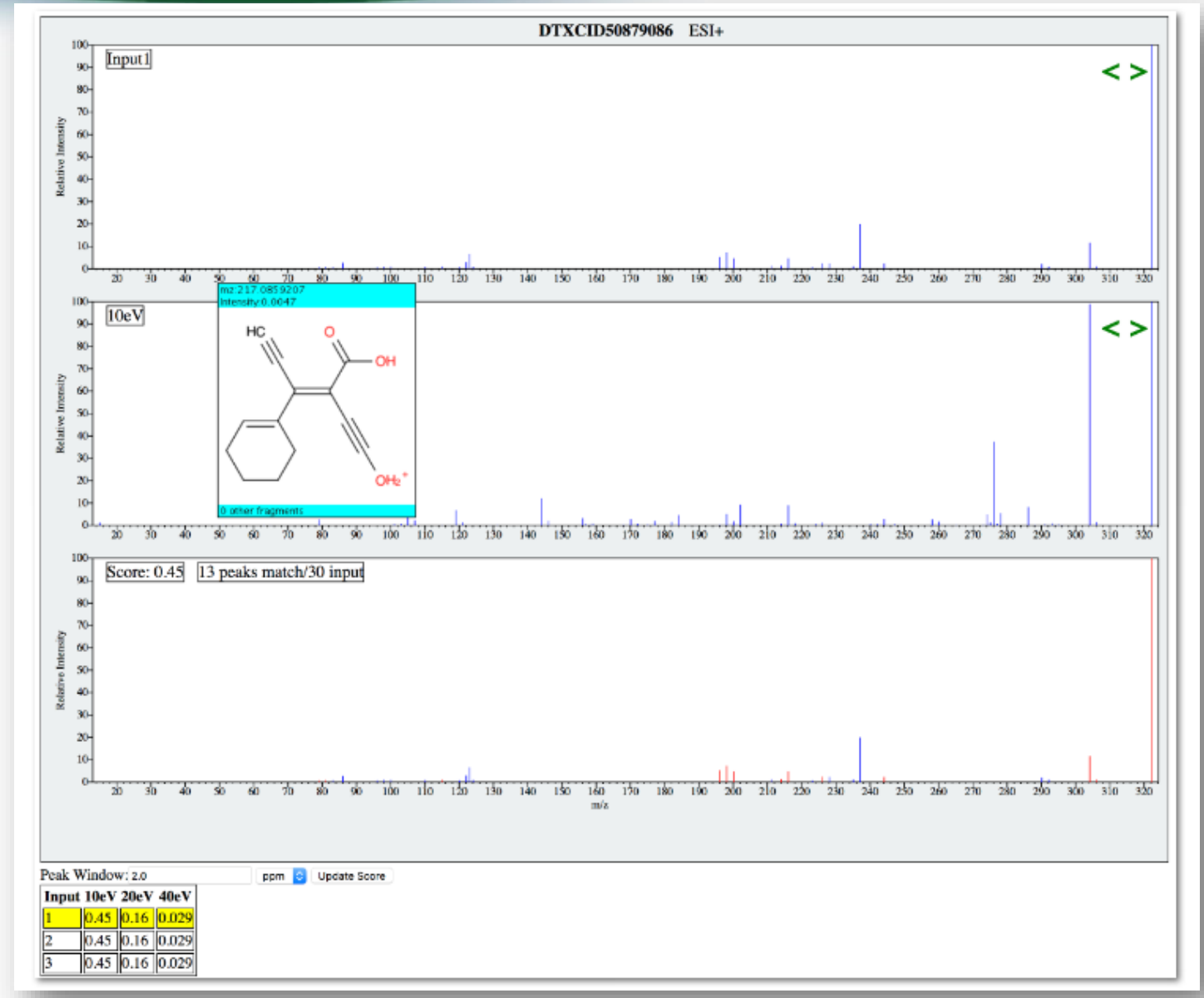


Search Expt. vs. Predicted Spectra



Chemical Structure ID	Score (10eV)
DTXCID101048191	0.22
DTXCID101181567	0.19
DTXCID50879086	0.17
DTXCID60686349	0.14
DTXCID00830900	0.13
DTXCID10971176	0.12
DTXCID60301242	0.12
DTXCID40703048	0.11
DTXCID60349982	0.11
DTXCID10316649	0.09

Spectral Viewer Comparison



Presented at ACS Fall 2019

EPA United States Environmental Protection Agency

Non-Targeted Screening of Wastewater for Water Reuse using Mass Spectrometry

Prototype development using the US-EPA CompTox Chemicals Dashboard data and CFM-ID Fragment Prediction algorithms

Jerry Zweigenbaum, Agilent, Wilmington DE, **Andrew McEachran**, Agilent, Santa Clara CA, **Alex Chao**, Oak Ridge Associated Universities (ORAU), US EPA, National Exposure Research Laboratory, **and Antony J. Williams**, National Center for Computational Toxicology, US EPA Research Triangle Park NC

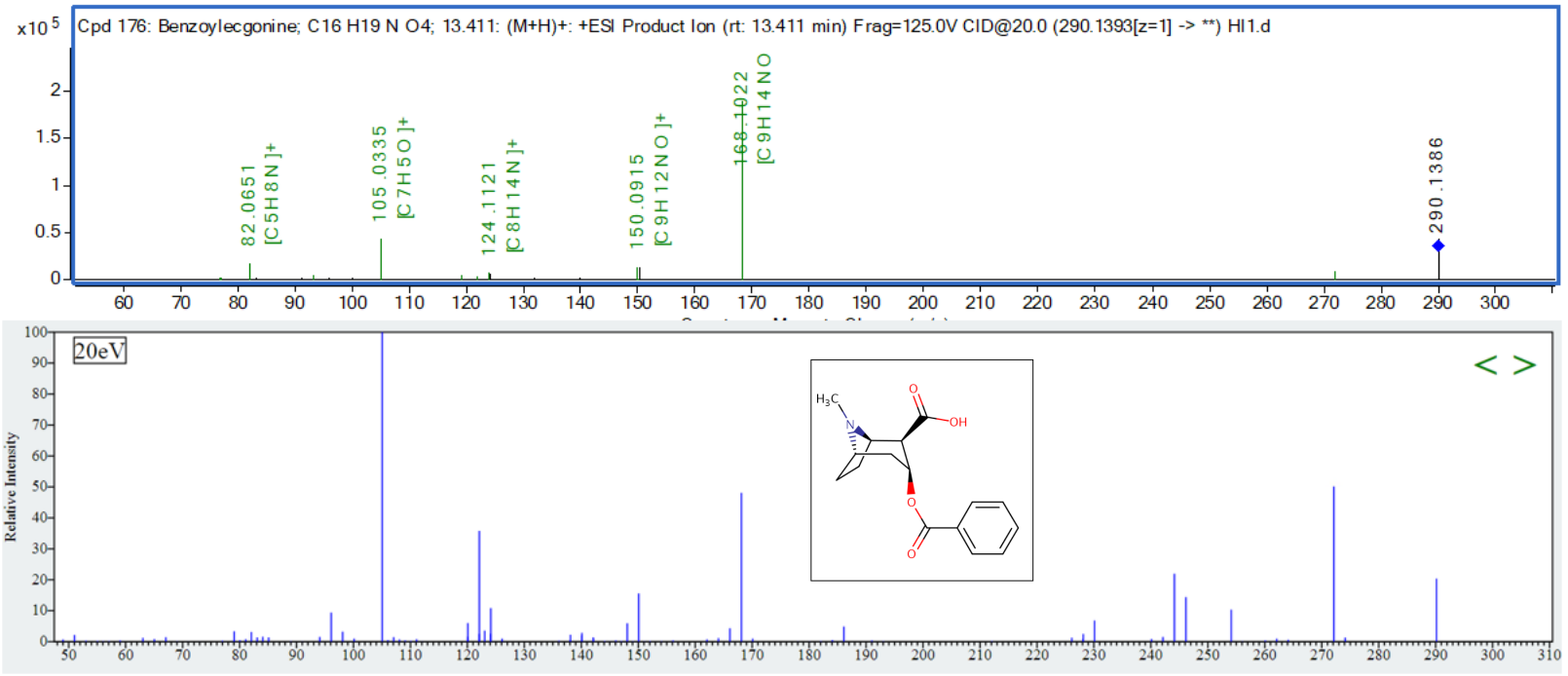
ACS Fall 2019 National Meeting & Exposition in San Diego

Agilent Trusted Answers

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

Example match

MS/MS of benzoylecgonine



Predicted Data Already Public
Publication and Data Files

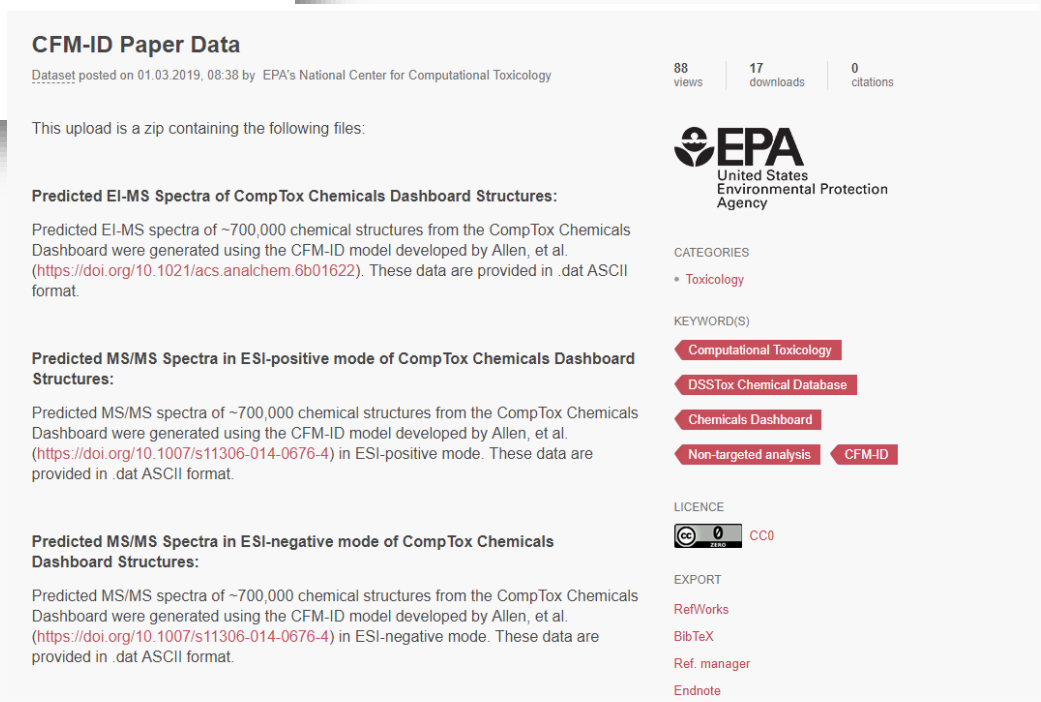
EPA United States Environmental Protection Agency

Data Descriptor | OPEN | Published: 02 August 2019

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

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

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



CFM-ID Paper Data

Dataset posted on 01.03.2019, 08:38 by EPA's National Center for Computational Toxicology

88 views | 17 downloads | 0 citations

This upload is a zip containing the following files:

Predicted EI-MS Spectra of CompTox Chemicals Dashboard Structures:

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

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

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

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

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

EPA United States Environmental Protection Agency

CATEGORIES
- Toxicology

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

LICENCE
CC0

EXPORT
RefWorks
BibTeX
Ref. manager
Endnote

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

In Press:

Analytical and Bioanalytical Chemistry

EPA United States Environmental Protection Agency

Analytical and Bioanalytical Chemistry

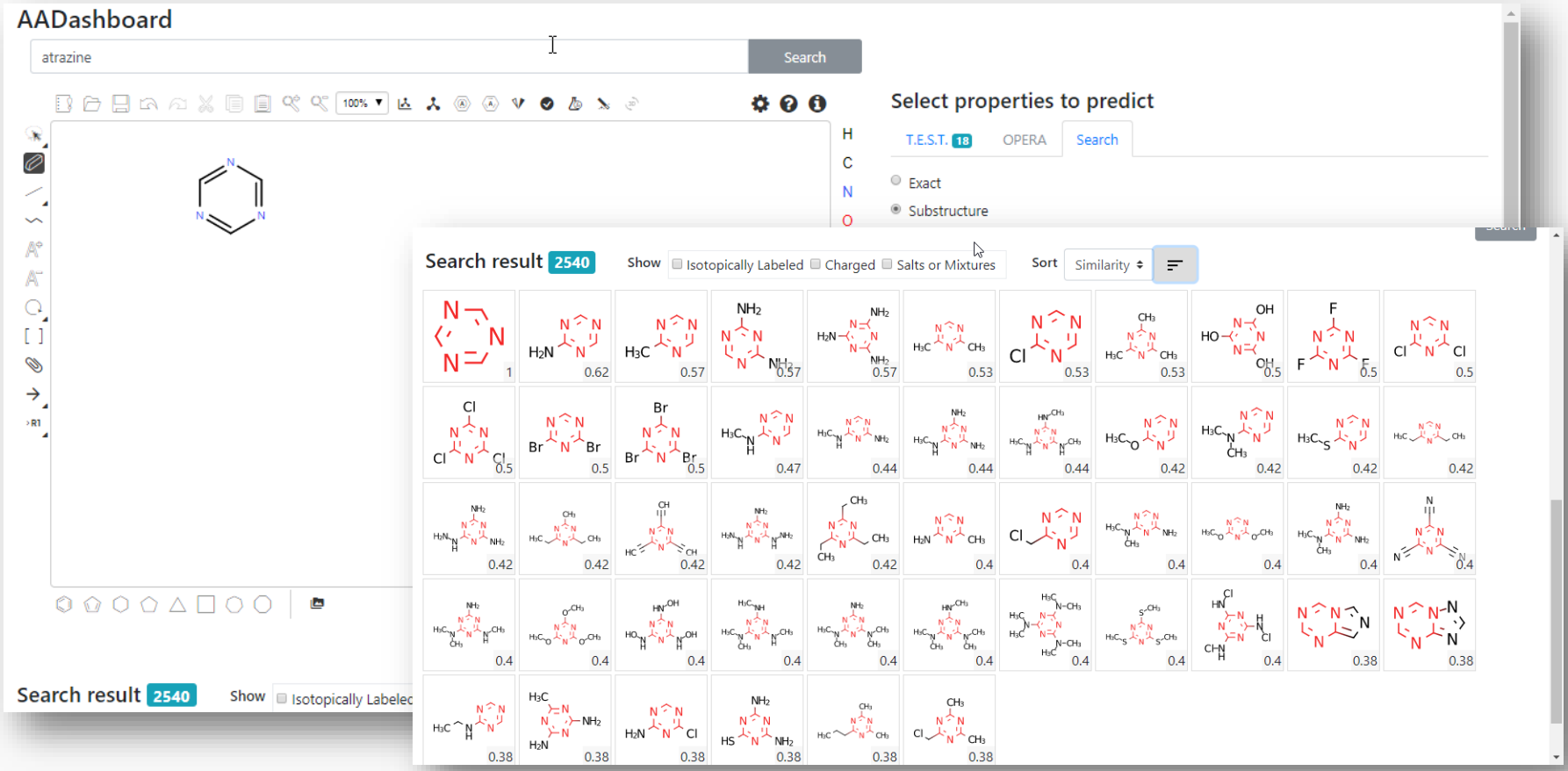
RESEARCH PAPER

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

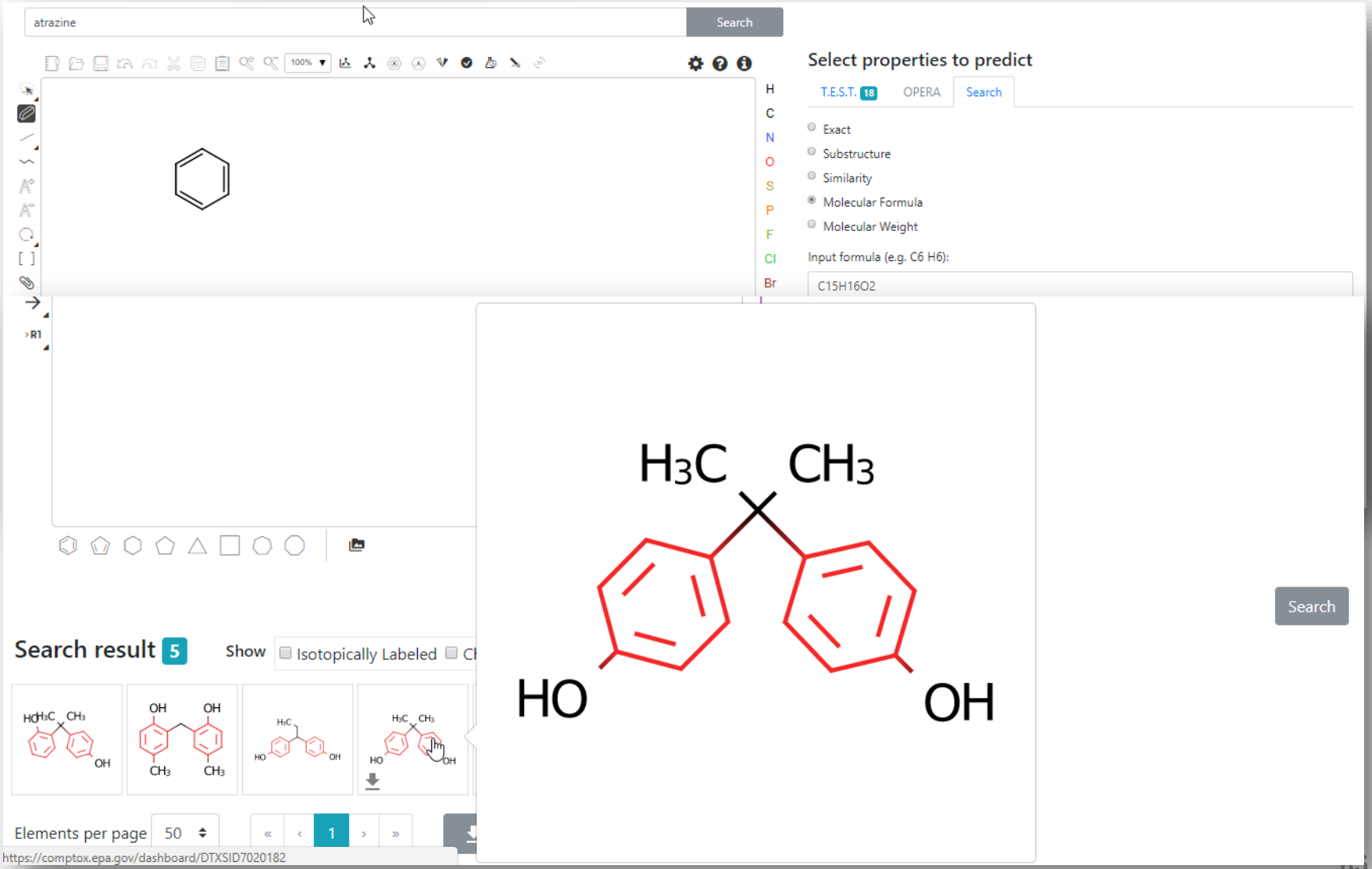
Alex Chao [1,2] · Hussein Al-Ghoul [1,2] · Andrew D. McEachran [1,3] · Ilya Balabin [4] · Tom Transue [4] · Tommy Cathey [4] · Jarod N. Grossman [2,3] · Randolph Singh [1,5] · Elin M. Ulrich [2] · Antony J. Williams [6] · Jon R. Sobus [2]

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

Prototype Development

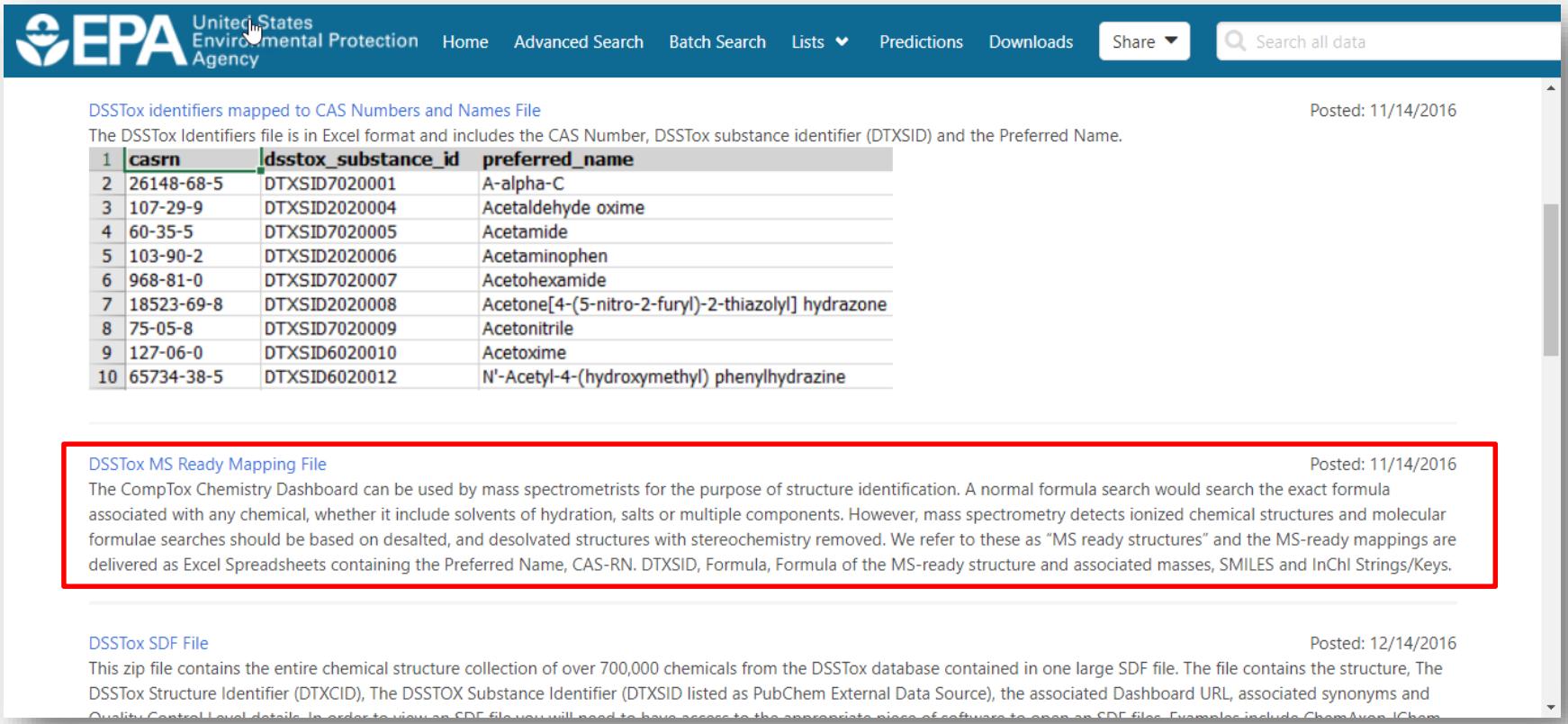


Prototype Development



API services and Open Data

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



	casrn	dsstox_substance_id	preferred_name
2	26148-68-5	DTXSID7020001	A-alpha-C
3	107-29-9	DTXSID2020004	Acetaldehyde oxime
4	60-35-5	DTXSID7020005	Acetamide
5	103-90-2	DTXSID2020006	Acetaminophen
6	968-81-0	DTXSID7020007	Acetohexamide
7	18523-69-8	DTXSID2020008	Acetone[4-(5-nitro-2-furyl)-2-thiazolyl] hydrazone
8	75-05-8	DTXSID7020009	Acetonitrile
9	127-06-0	DTXSID6020010	Acetoxime
10	65734-38-5	DTXSID6020012	N'-Acetyl-4-(hydroxymethyl) phenylhydrazine

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

- Data in UI, JSON and XML format

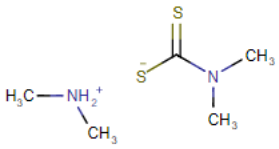
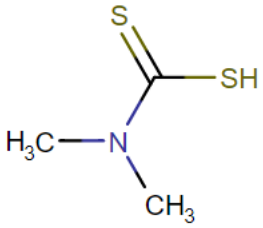
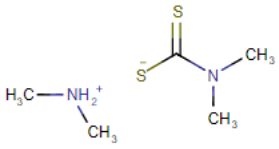
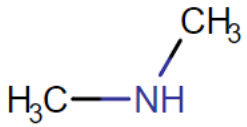
https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=80-05-7
https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=80-05-7
https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=80-05-7

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

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

InChIKey to DTXCIDs

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

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

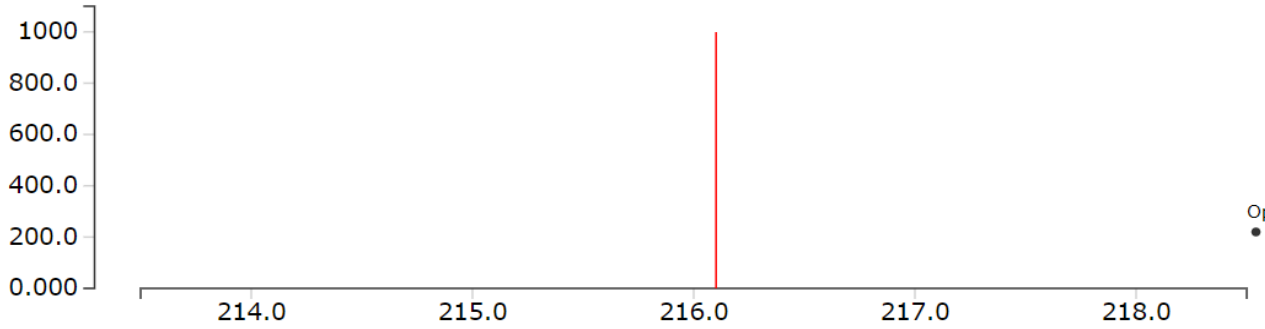
MassBank mapping to Dashboard

MassBank Record: EA028808

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

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

Mass Spectrum



Options
● Labels

Chemical Structure

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

Future Work

Collision Cross Section Prediction

EPA United States Environmental Protection Agency

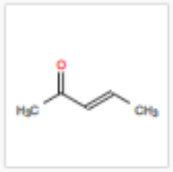
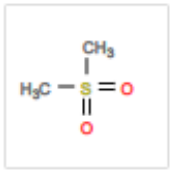
PNNL Collision Cross Section Database

Pacific Northwest NATIONAL LABORATORY

Search

Showing 1 to 25 of 1000 rows 25 rows per page

‹ 1 2 3 4 5 ... 40 ›

Chemical	SMILES	InChI	Formula	Mass	CCS (Å^2)
 (3E)-pent-3-en-2-one	*	*	C_5H_8O	84.0575	[M-H]$^-$ 112.1 ISiCLE Lite v0.1.0 [M+Na]$^+$ 112.6 ISiCLE Lite v0.1.0 [M+H]$^+$ 113.1 ISiCLE Lite v0.1.0
 Dimethyl sulfone	*	*	$C_2H_6O_2S$	94.0089	[M-H]$^-$ 106.9 ISiCLE Lite v0.1.0 [M+Na]$^+$ 107.3 ISiCLE Lite v0.1.0 [M+H]$^+$ 108.1 ISiCLE Lite v0.1.0
 isothiocyanatocyclopropane	*	*	C_4H_5NS	99.0143	[M-H]$^-$ 111.9 ISiCLE Lite v0.1.0 [M+Na]$^+$ 112.1 ISiCLE Lite v0.1.0 [M+H]$^+$ 110.0 ISiCLE Lite v0.1.0

Data and Services used by the Community

NORMAN Suspect List Exchange

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

EPA United States Environmental Protection Agency

NORMAN

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

Wastewater Suspect List based on Swedish Product Data	Wastewater Suspect List **Original File with Mapped DTXSIDs** (12/02/2019)	KEMIWWSUS **InChIKeys** (12/02/2019)	A prioritized list of 1,123 substances relevant for wastewater based on Swedish product registry data, including scores. Provided by Stellan Fischer, KEMI.
Algal toxins list from CompTox	ALGALTOX **XLSX**, **CSV** (14/02/2019) CompTox **ALGALTOX List**	ALGALTOX **InChIKeys** (14/02/2019)	List of algal toxins (generated during blooms) from the CompTox Chemicals Dashboard.
CCL 4 Chemical Candidate List	CCL4 **XLSX**, **CSV** (14/02/2019) CompTox **CCL4 List**	CCL4 **InChIKeys** (14/02/2019)	Contaminants that are not (yet) regulated in the USA but are known or anticipated to occur in public water systems; from CompTox.
Hydrogen Deuterium Exchange (HDX) Standard Set	HDXNOEX **XLSX**, **CSV** (14/02/2019) CompTox **HDXNOEX List** CompTox **HDXEXCH List**	HDXNOEX **InChIKeys** (14/02/2019)	Environmental standard set used to investigate hydrogen deuterium exchange in small molecule HRMS (Ruttkies et al. submitted). **HDXEXCH** list also contains observed deuterated species.
Neurotoxicants Collection from Public Resources	NEUROTOXINS **XLSX**, **CSV** (14/02/2019) CompTox **NEUROTOXINS List**	NEUROTOXINS **InChIKeys** (14/02/2019)	A list of neurotoxicants compiled from public resources, details on CompTox and Schymanski *et al.* (submitted).
Statins Collection from Public Resources	STATINS **XLSX**, **CSV** (14/02/2019) CompTox **STATINS List**	STATINS **InChIKeys** (14/02/2019)	A list of statins (lipid-lowering medications) compiled from public resources, details on CompTox.
Synthetic Cannabinoids and Psychoactive Compounds	SYNTHCANNAB **XLSX**, **CSV** (14/02/2019) CompTox **SYNTHCANNAB List**	SYNTHCANNAB **InChIKeys** (14/02/2019)	A list of synthetic cannabinoids and psychoactive compounds assembled from public resources, from CompTox.

Integration to MetFrag in place

https://jcheminf.biomedcentral.com/articles/10.1186/s13321-018-0299-2

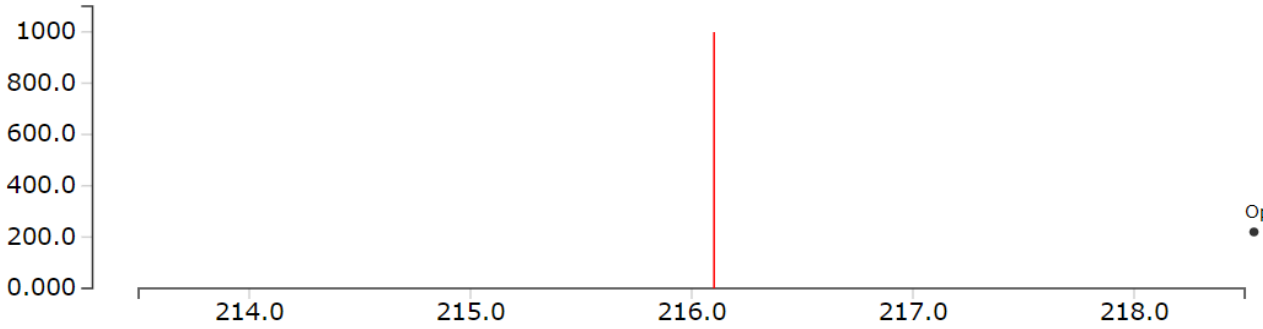
MassBank mapping to Dashboard Based on Web Service lookup

MassBank Record: EA028808

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

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

Mass Spectrum



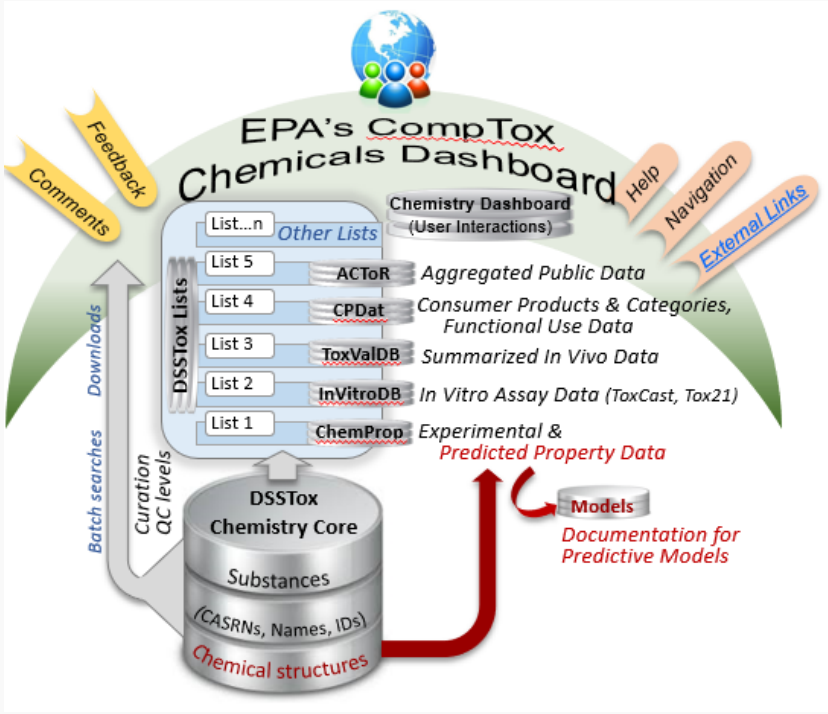
Options
● Labels

Chemical Structure

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

Conclusion

- Dashboard access to data for ~875,000 chemicals (~895k in the Spring Release)
- MS-Ready data facilitates structure identification
- Related metadata facilitates candidate ranking
- Relationship mappings and chemical lists of great utility
- Curation and mutual sharing of chemical lists is important (e.g. NORMAN)

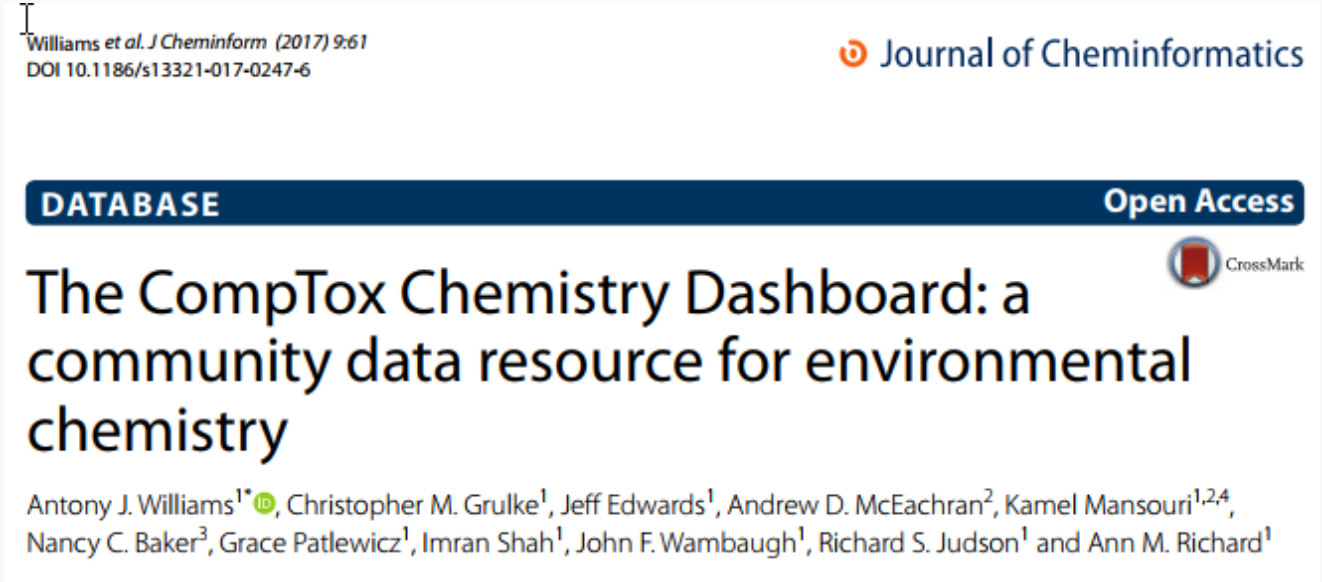


Acknowledgements

- IT Development team – especially Jeff Edwards and Jeremy Dunne
- Exposure Modeling and CPDat – John Wambaugh, Kristin Isaacs, Katharine Phillips and Kathie Dionisio
- EPA-MS colleagues – Jon Sobus, Elin Ulrich, Mark Strynar, Seth Newton, Alex Chao
- Chris Grulke (ChemReg system)
- Charlie Lowe (postdoc)
- Andrew McEachran (ex-postdoc, MS-Ready/CFM-ID)
- Kamel Mansouri (ex-postdoc, MS-Ready)
- Valery Tkachenko (working on new MS-Ready)
- Emma Schymanski, LCSB, Luxembourg
- NORMAN Network and all contributors

Contact

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



Williams et al. J Cheminform (2017) 9:61
DOI 10.1186/s13321-017-0247-6

Journal of Cheminformatics

DATABASE Open Access

The CompTox Chemistry Dashboard: a community data resource for environmental chemistry

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

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