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

US-EPA Comptox Chemicals Dashboard to support mass spectrometry targeted and non-targeted analysis

Antony Williams[1], *Alex Chao*[2], *Tom Transue*[3], *Tommy Cathey*[3], *Elin Ulrich*[4] *and Jon Sobus*[4]

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

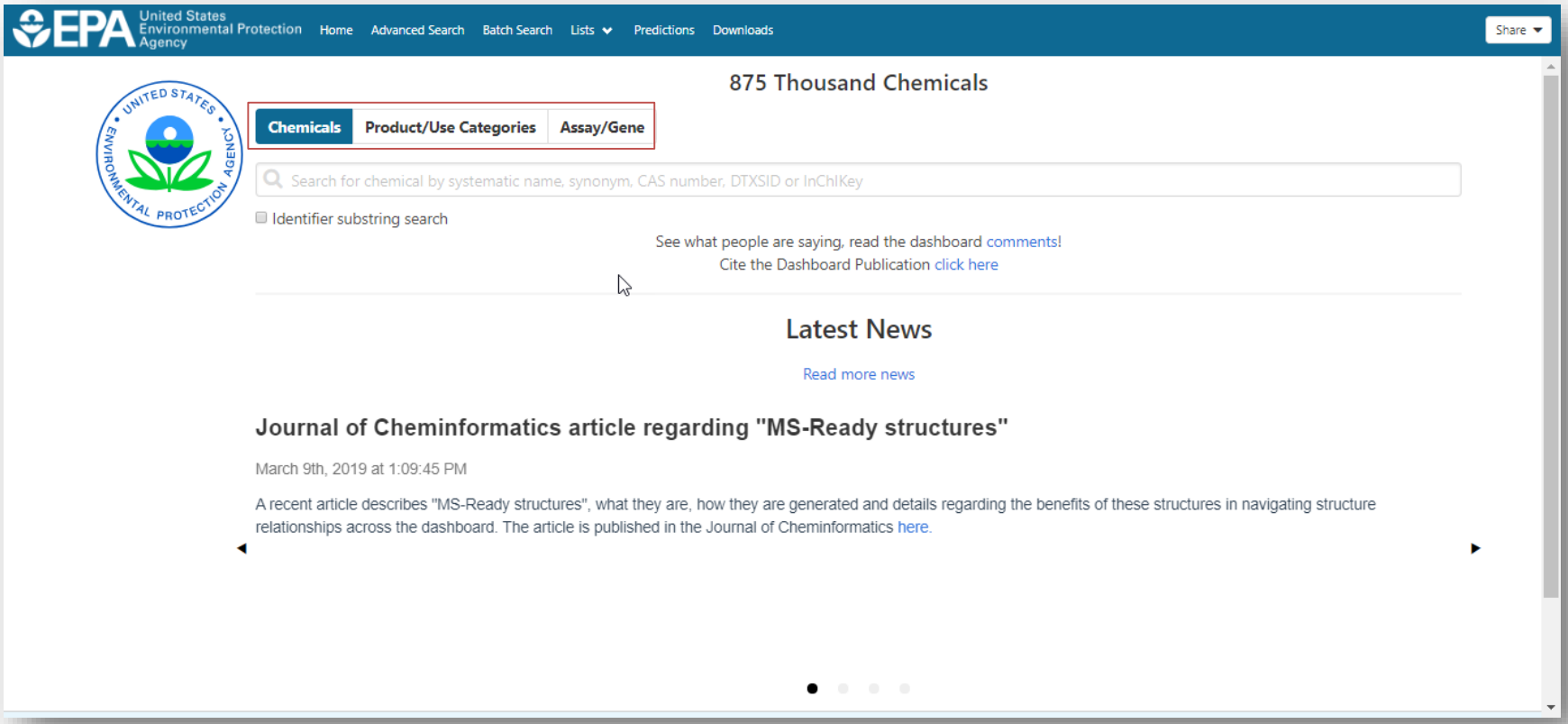
August 2019
ACS Fall Meeting, San Diego

CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard

EPA United States Environmental Protection Agency

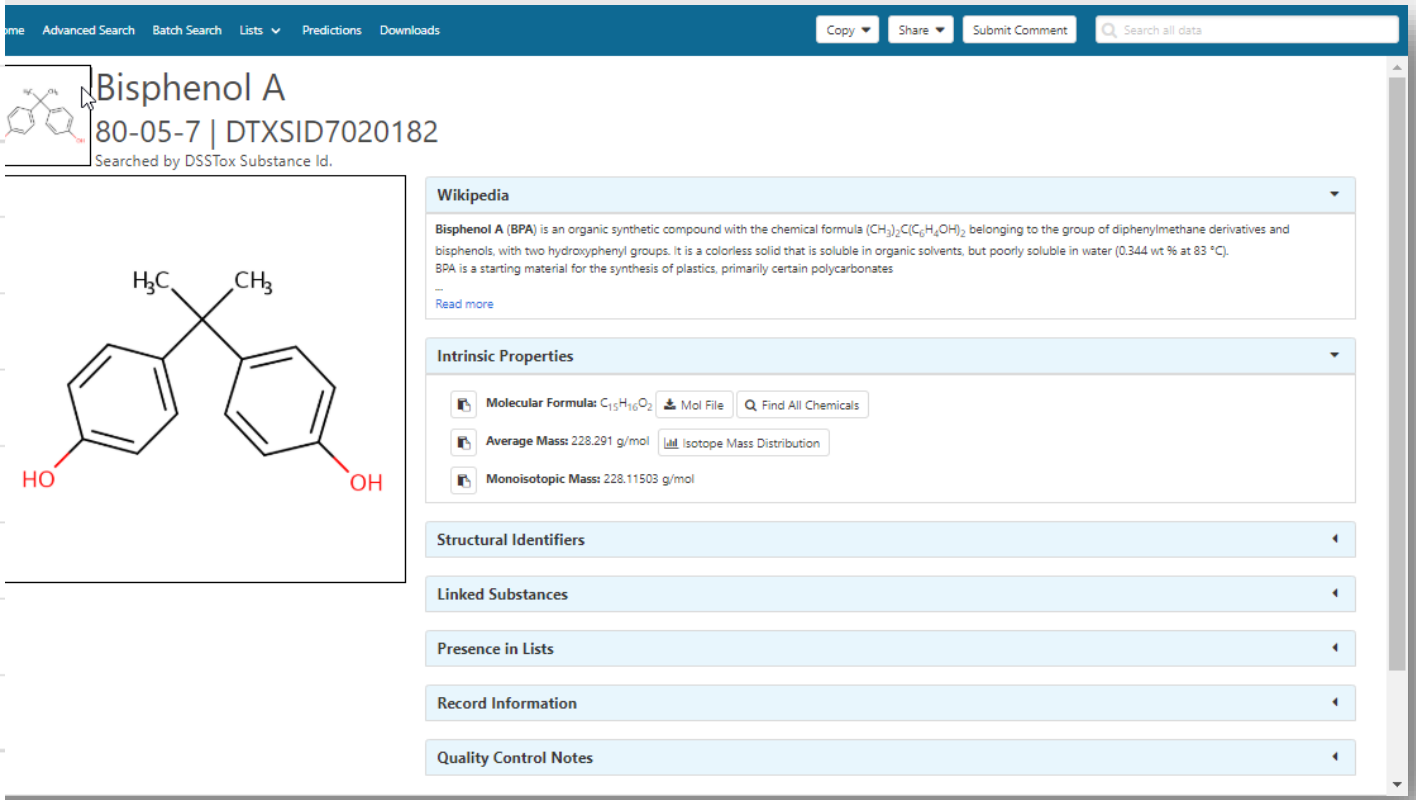
875k Chemical Substances



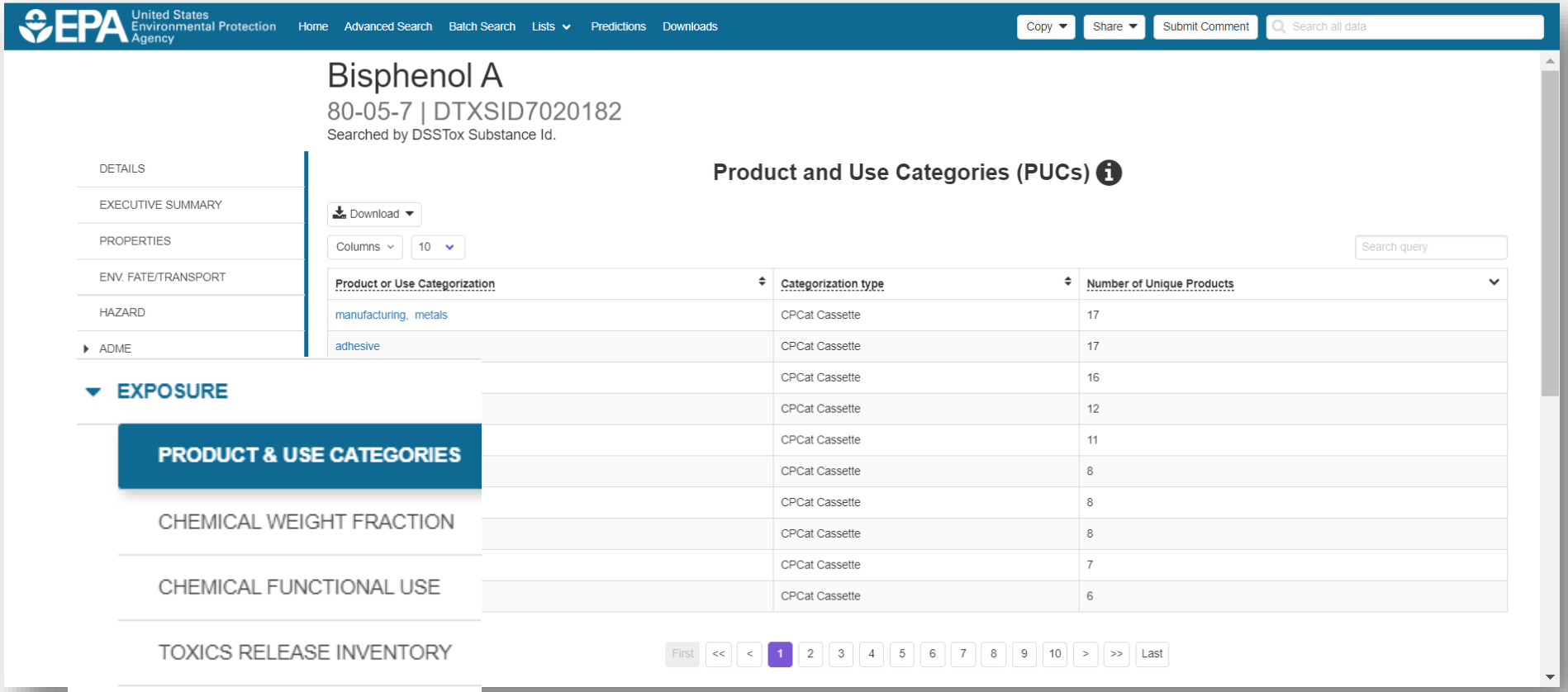
Detailed Chemical Pages

EPA United States Environmental Protection Agency

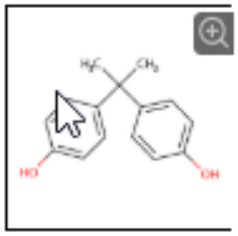
- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



Sources of Exposure to Chemicals



Physicochemical properties and environmental fate and transport



Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

Property

Summary

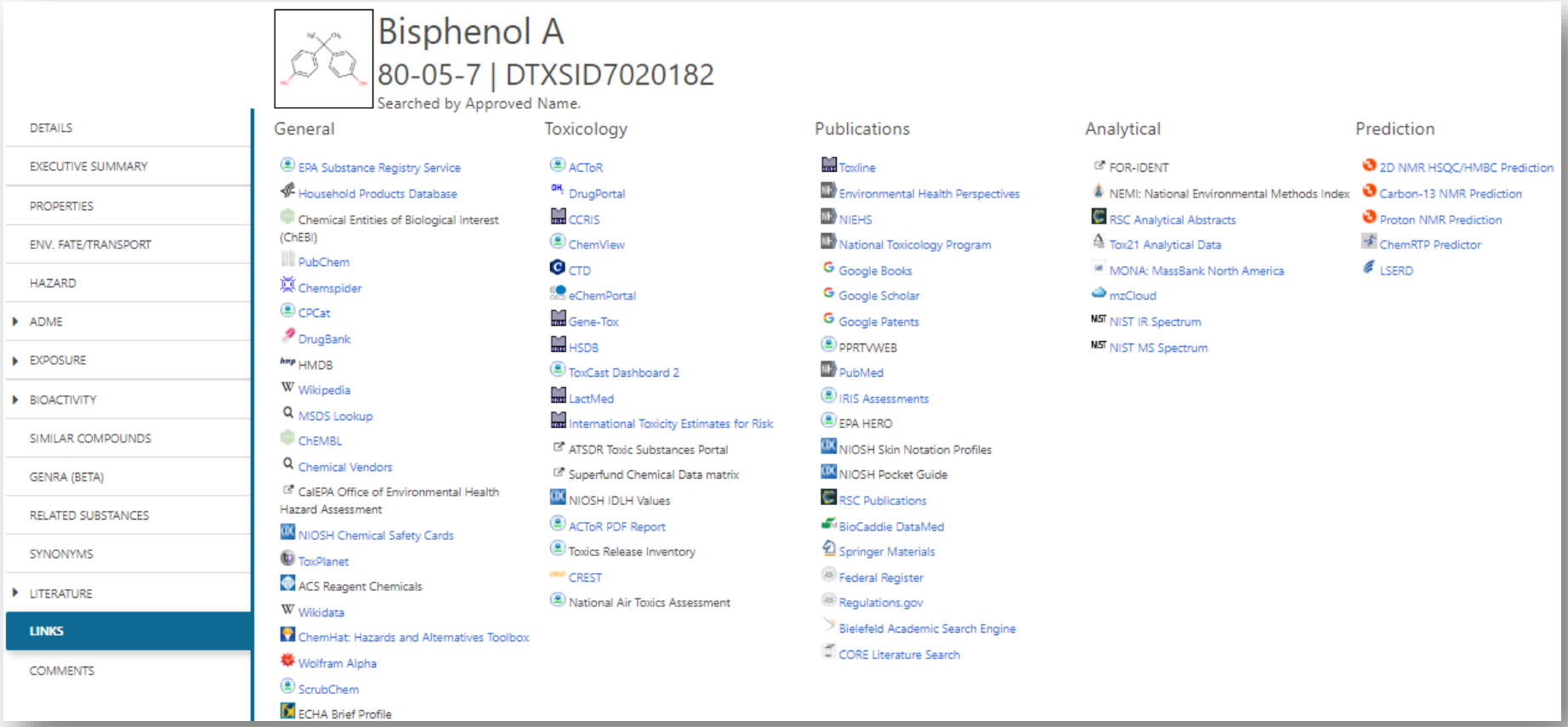
Download Columns

Property	Experimental average	Predicted average
LogP: Octanol-Water	3.32 (1)	3.29
Melting Point	155 (7)	139
Boiling Point	200 (1)	363
Water Solubility	5.26e-4 (1)	9.62e-4
Vapor Pressure	-	8.37e-7
Flash Point	-	190

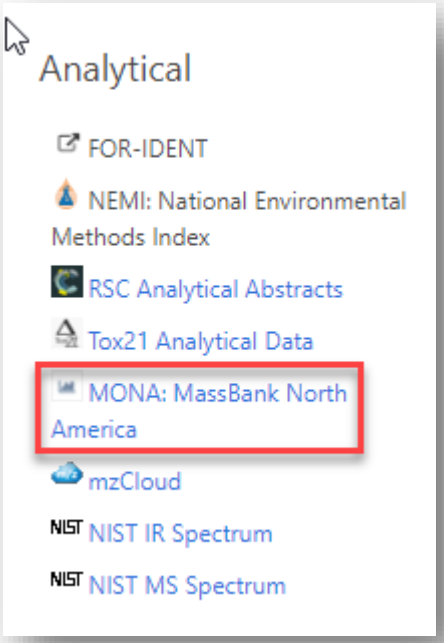
CompTox Chemicals Dashboard

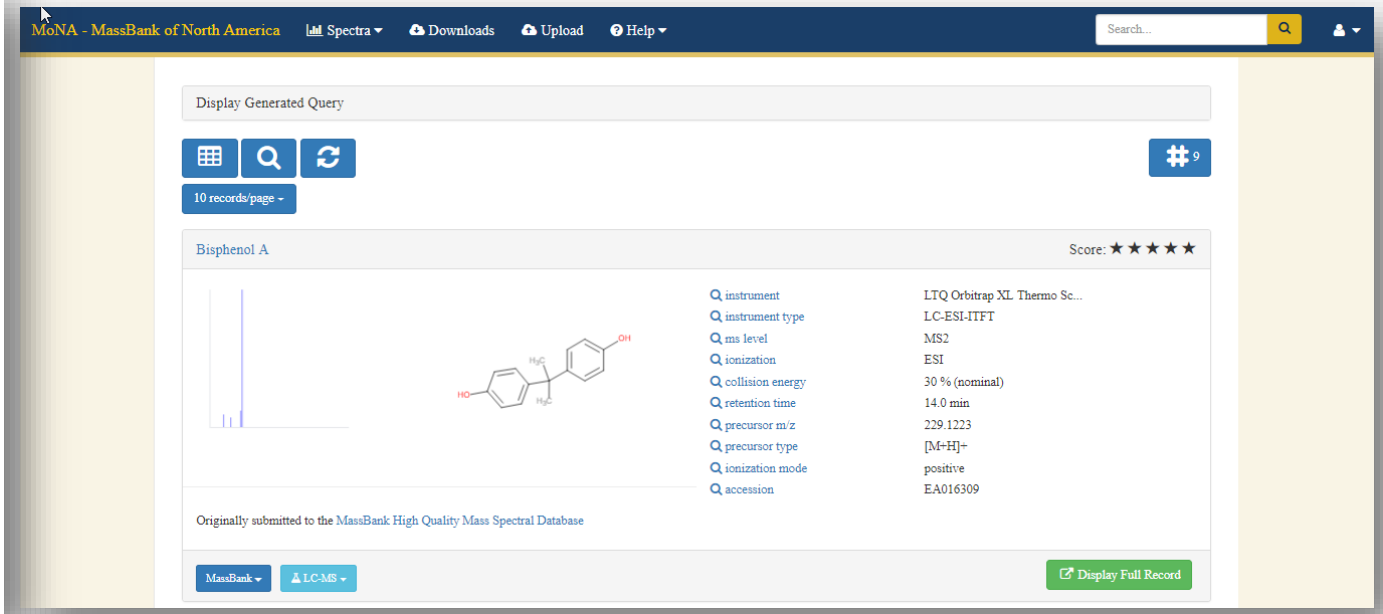
- Can provide access to toxicity, environmental fate and transport and metabolism data
- Individual chemicals can map to degradation products and metabolites
-
- Advanced searches support mass and formula searches

Link farm to public resources

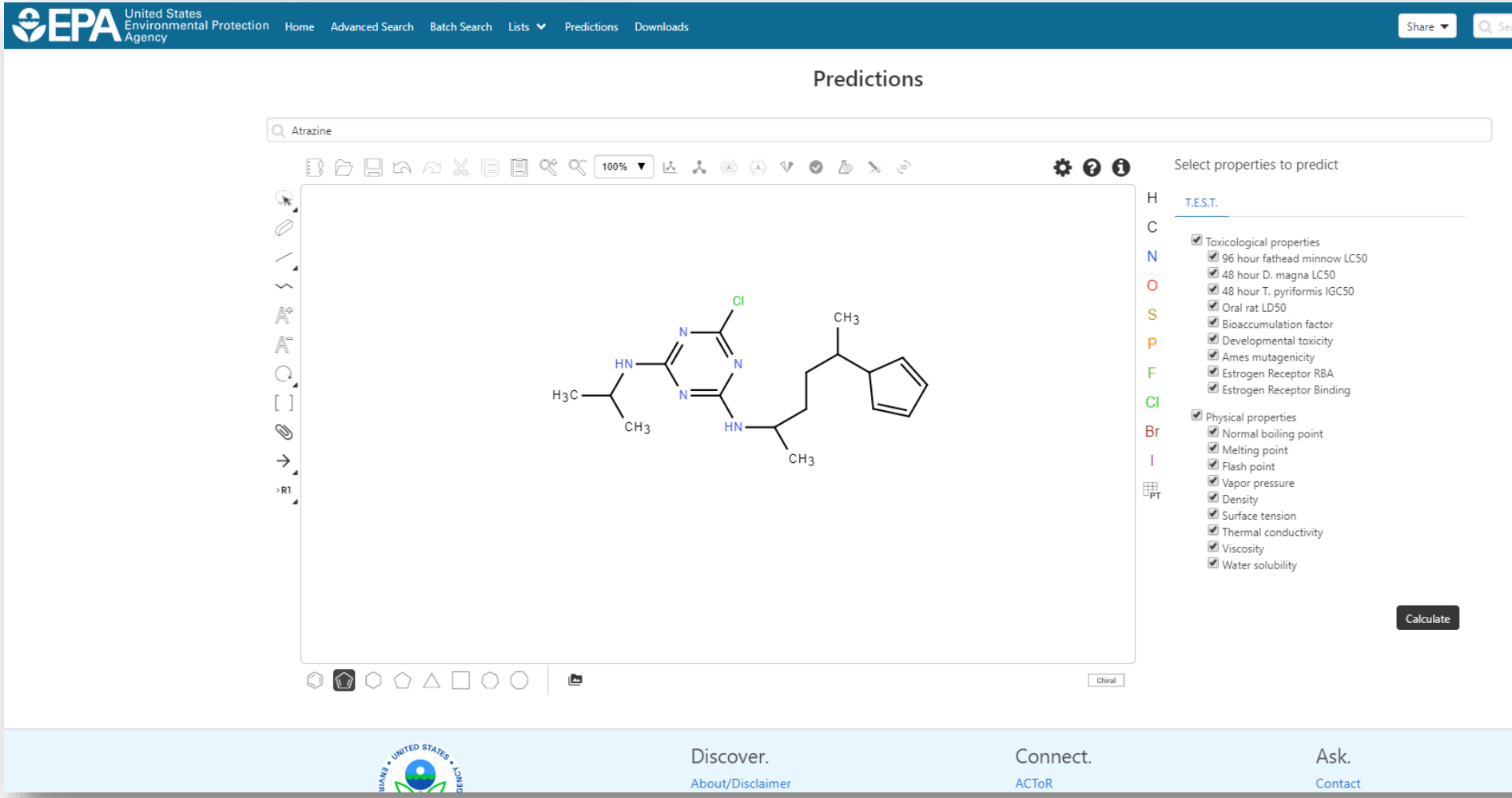


MassBank of North America
https://mona.fiehnlab.ucdavis.edu



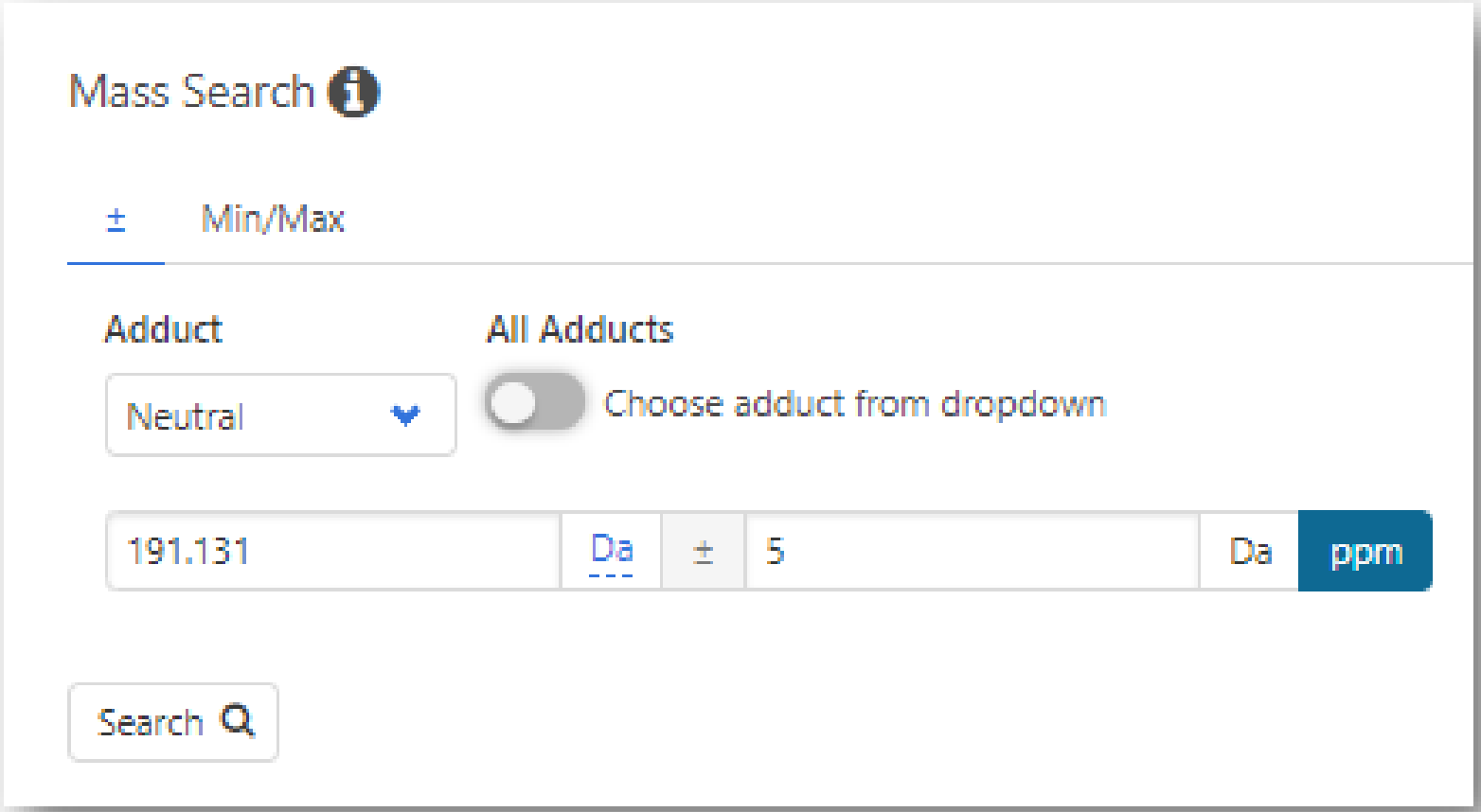


Toxicity Estimation Software Tool (TEST) Real Time Predictions

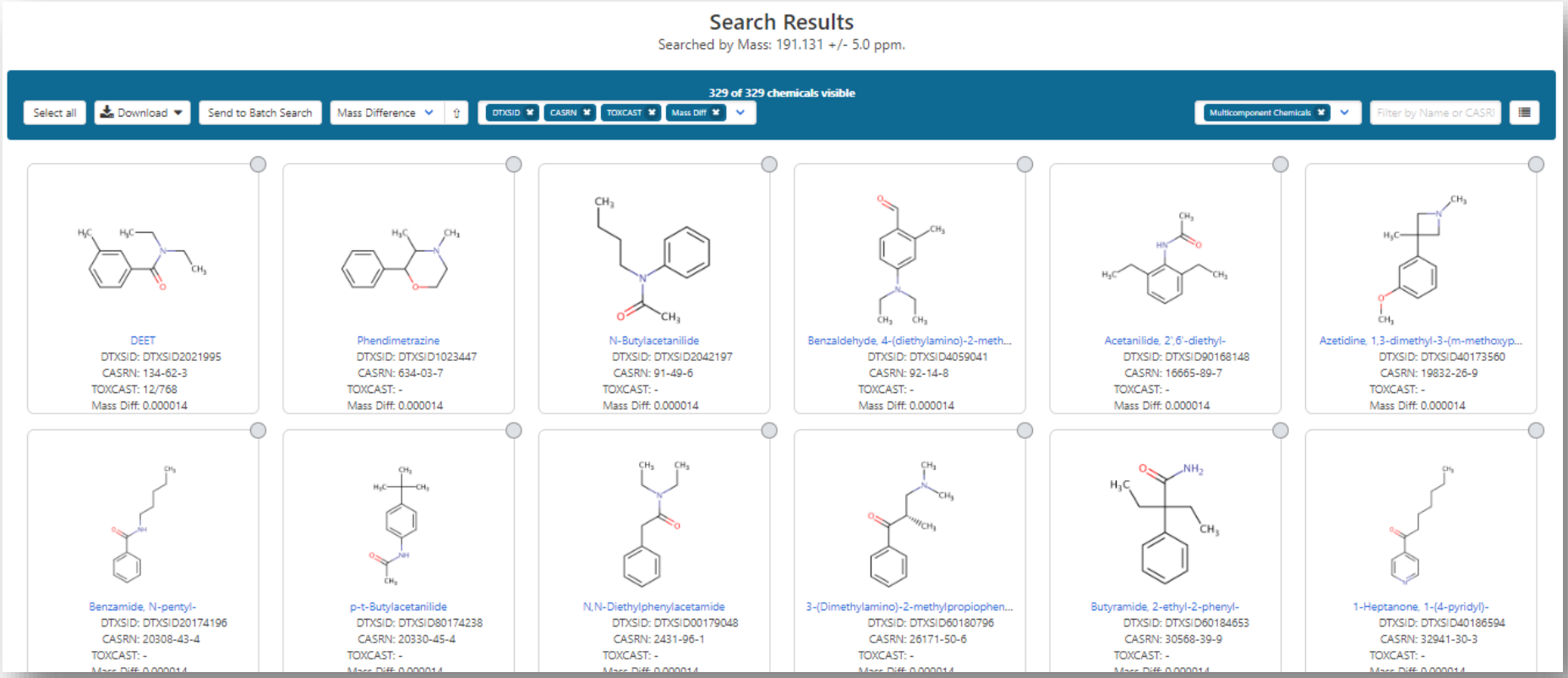


Mass & Formula Searching

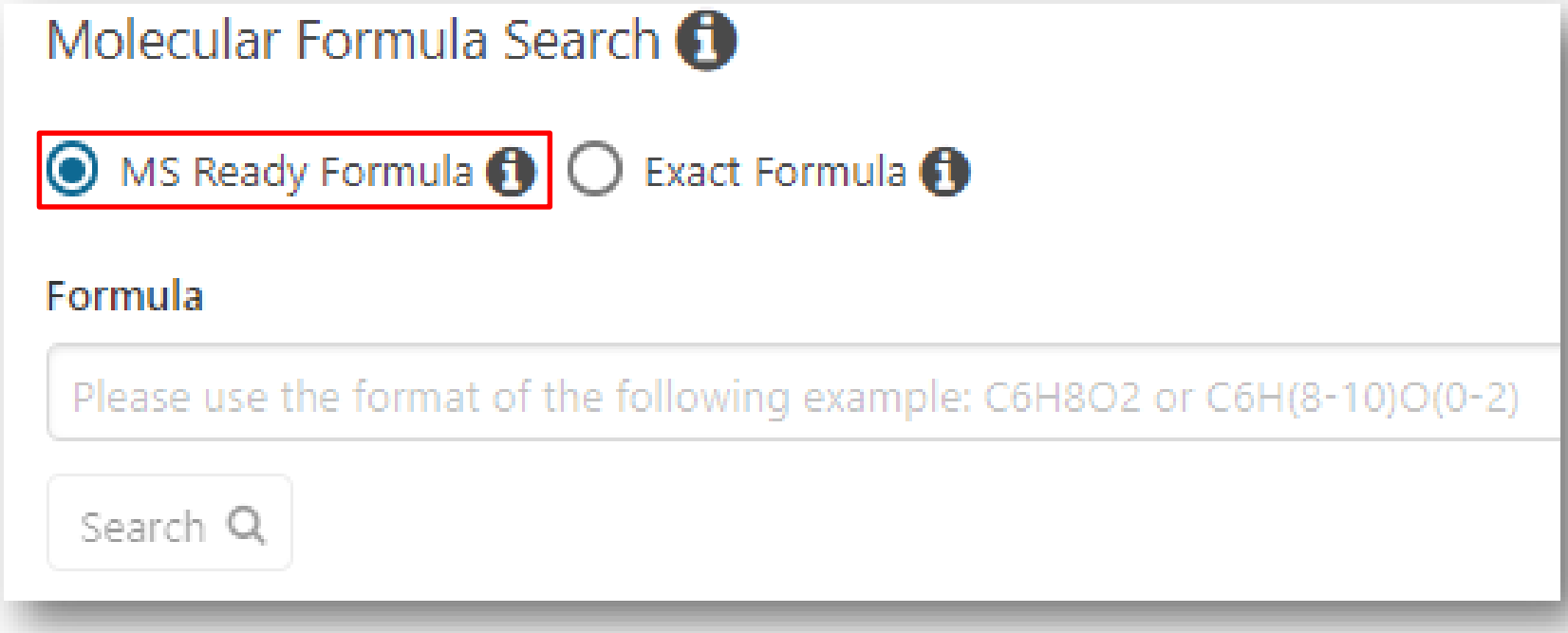
Advanced Searches
Mass Search



Advanced Searches
Mass Search



MS-Ready Structures for **Formula Search**



“MS-Ready Structures”

https://doi.org/10.1186/s13321-018-0299-2

EPA United States Environmental Protection Agency

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

Journal of Cheminformatics

METHODOLOGY Open Access

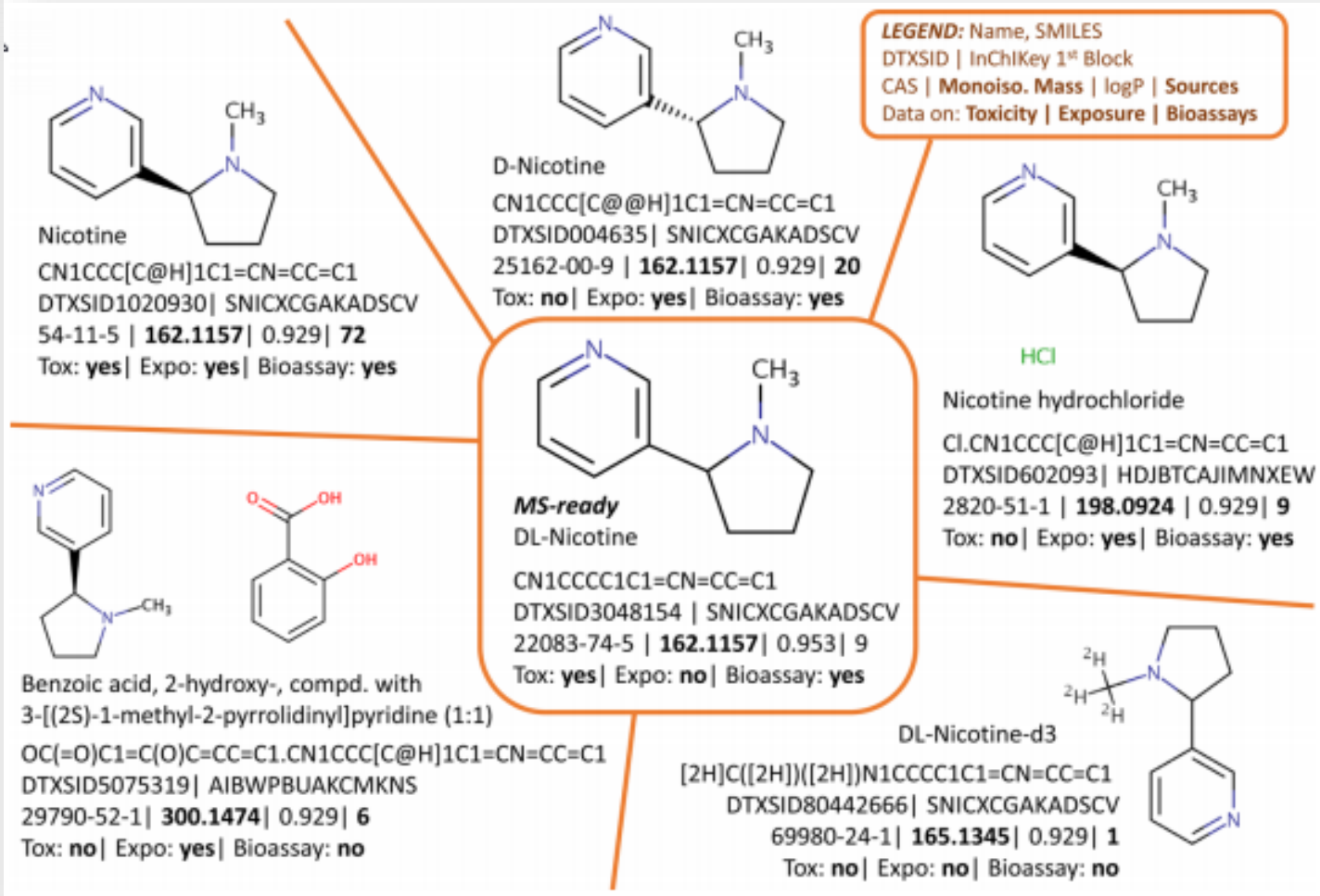
“MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies

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

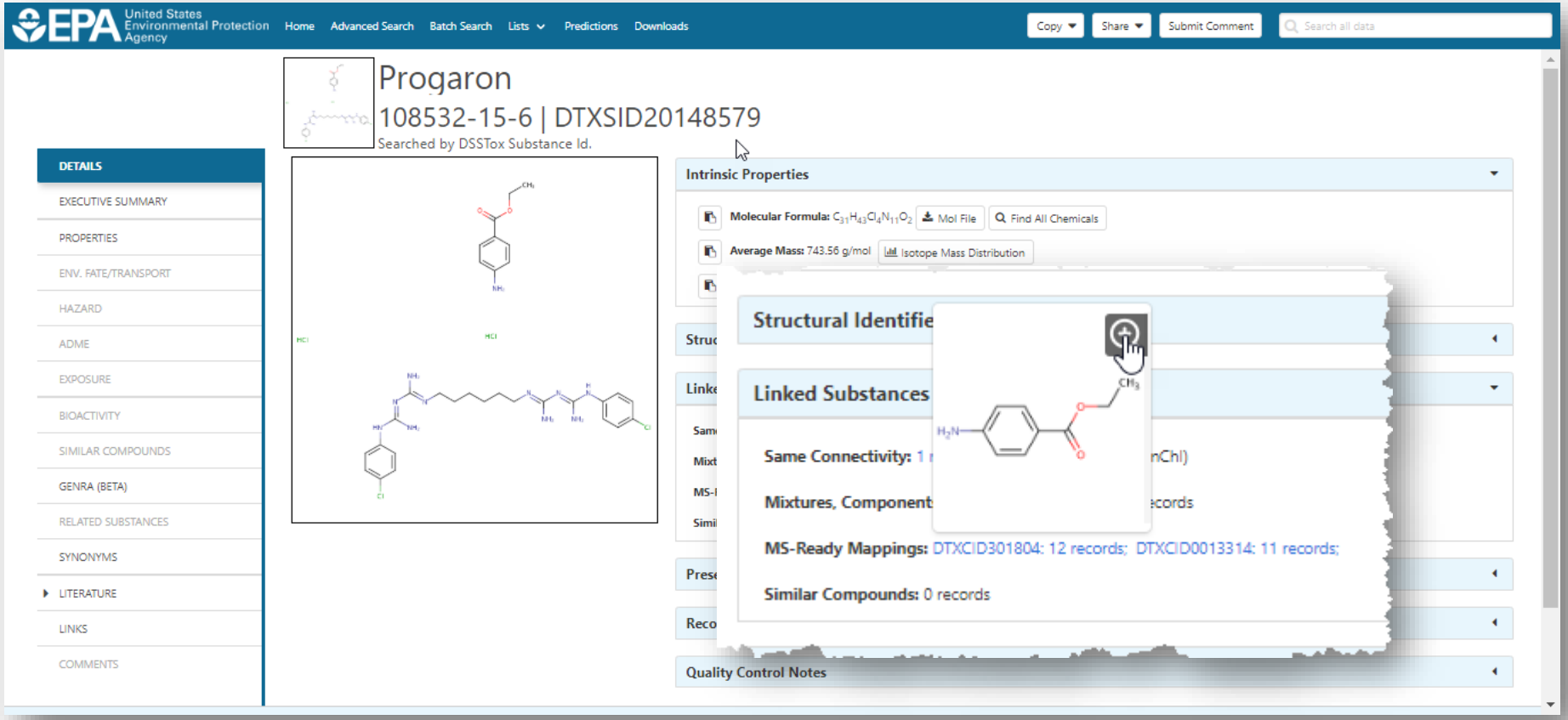
Environmental Science & Technology
Viewpoint
pubs.acs.org/est

Open Science for Identifying "Known Unknown" Chemicals

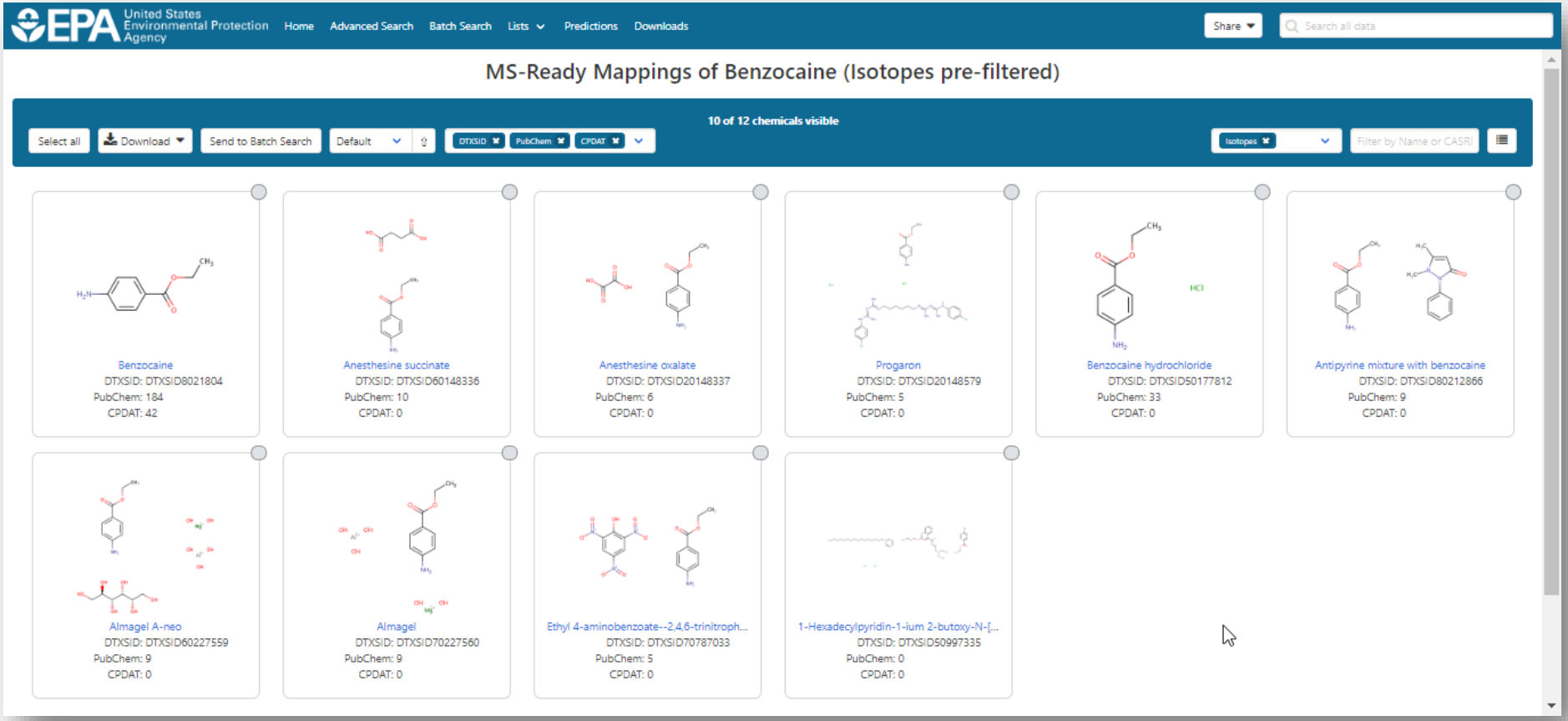
Emma L. Schymanski*,† and Antony J. Williams*,‡



MS-Ready Mappings

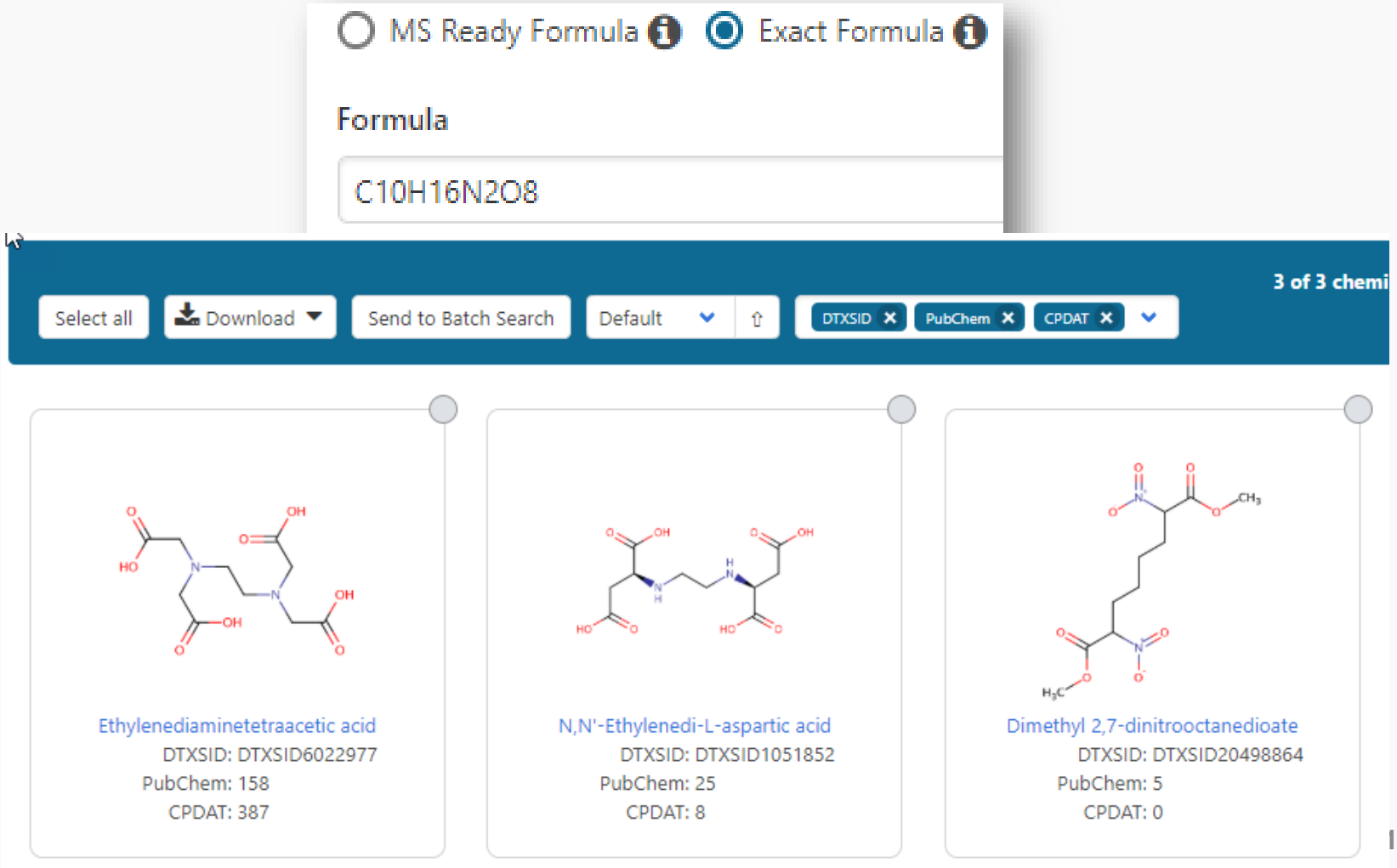


MS-Ready Mappings Set



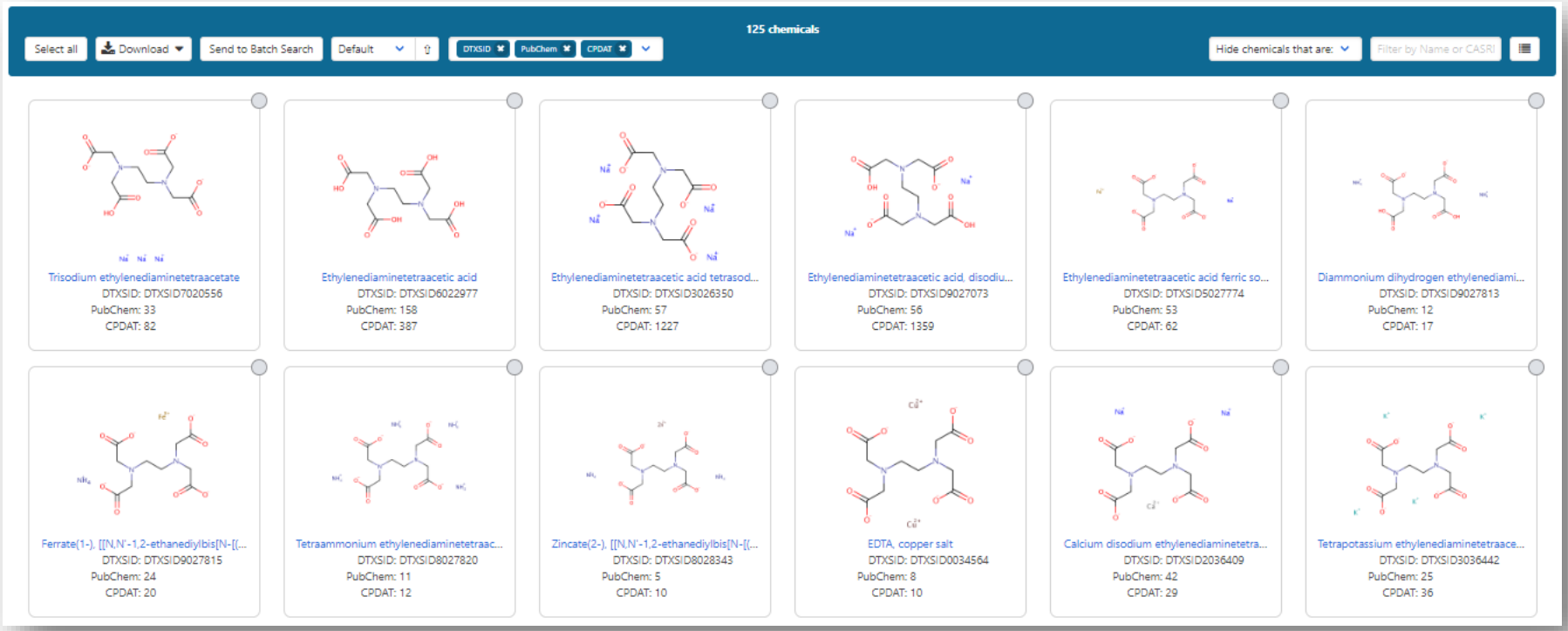
MS-Ready Mappings

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



MS-Ready Mappings

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



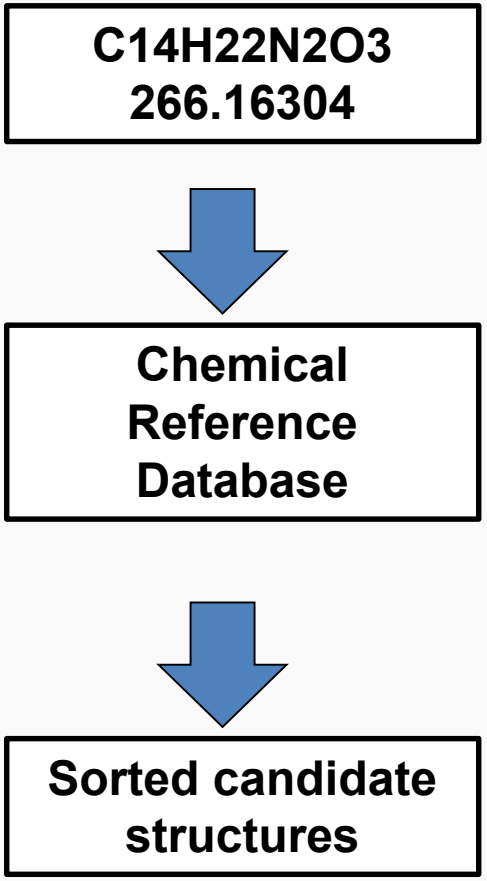
MS-Ready Mappings

- 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**

Candidate ranking

Data Source Ranking of "known unknowns"

- Mass and/or formula is for an unknown chemical but contained within a reference database
- Most likely candidate chemicals have the most associated data sources, most associated lit. articles or both



Is a bigger database better?

© American Society for Mass Spectrometry, 2011

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

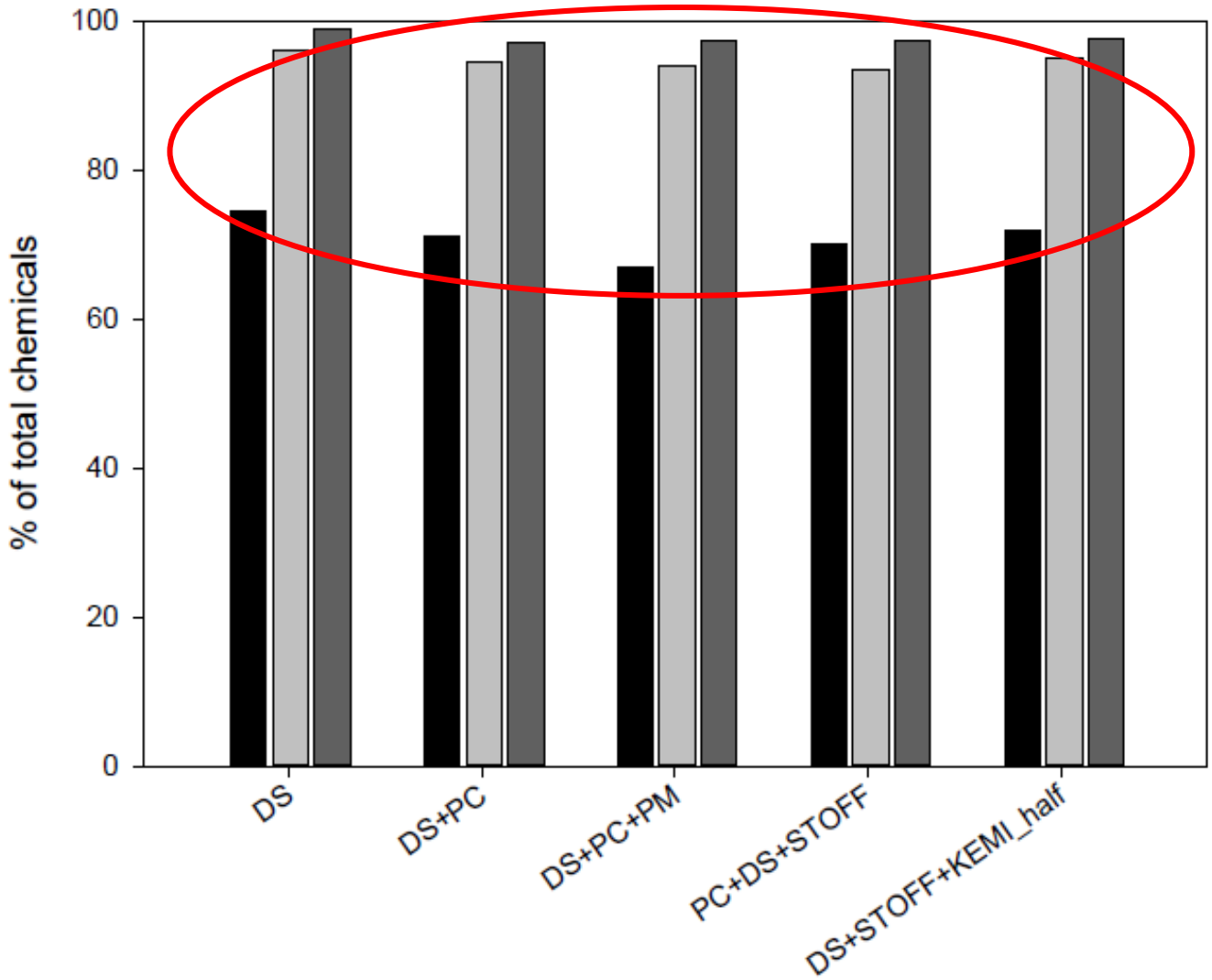
- ChemSpider was 26 million chemicals then
- Much BIGGER today
- Is bigger better??

71 Million chemical structures

Using Metadata for Ranking

- Use available metadata to rank candidates
  - Associated data sources
    - Associated lists in the underlying database
    - Associated data sources in PubChem
    - Specific types (e.g. water, surfactants, pesticides etc.)
  - Number of associated literature articles (Pubmed)
  - **Chemicals in the environment** – the number of products/categories containing the chemical is a very important source of data

Identification ranks for 1783 chemicals using multiple data streams



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]

- Dashboard content *was* 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			

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

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

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

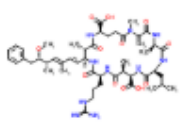
Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID

FILTER

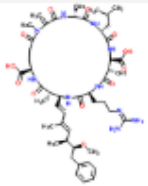
Search Hits Limit: 100

Found 9 results

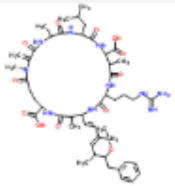
Search term: **ZYZCGGRZINLQBL** (Found by InChIKey (skeleton match))



Cyanoginosin

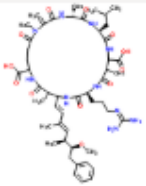


MCYST-LR

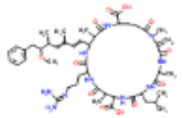


15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-

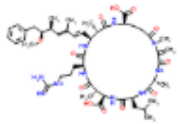
(5R,8S,11R,12S,15S,18R,19S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-3,5-



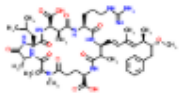
(5R,8S,15S)-15-{3-[(Diaminomethylene)amino]propyl}-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-



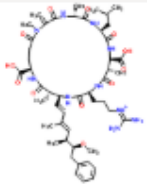
15-{3-[(Diaminomethylene)amino]propyl}-8-isobutyl-18-(6-methoxy-3,5-dimethyl-7-



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

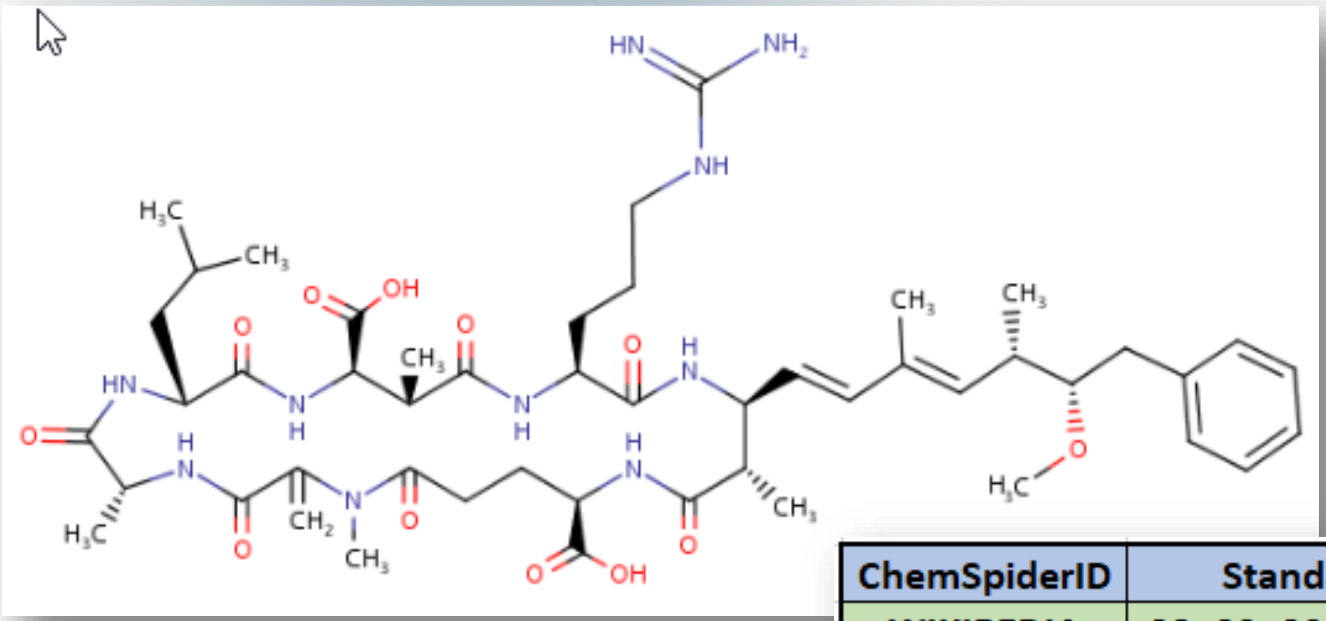


(5R,8R,11R,12S,15S,18S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5R,6R)-6-methoxy-3,5-dimethyl-



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

Comparing ChemSpider Structures

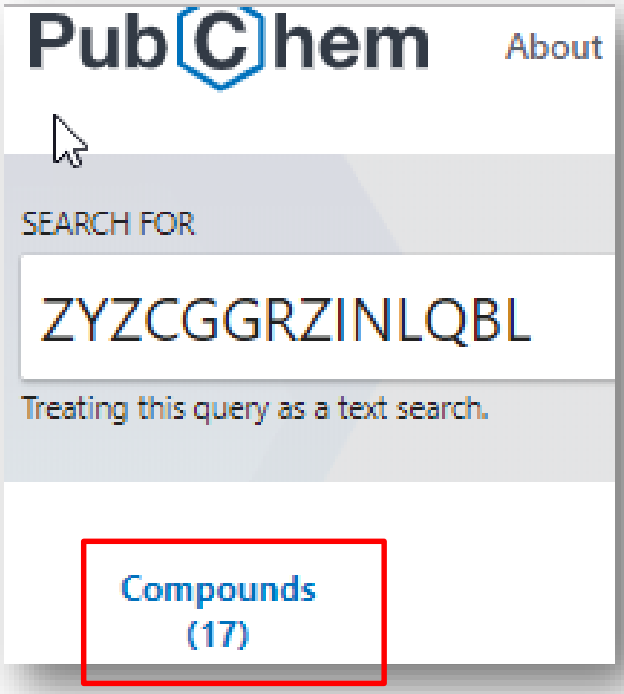


ChemSpiderID	Standard InChIKey Stereolayer
WIKIPEDIA	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
CompTox	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
4941647	t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+
393078	t28-,29-,30-,31+,34-,35-,36+,**37-**,38-,40+
57618348	t28-,29-,30-,31+,34-,35-,36+,**37-**,38-,40+
29342071	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

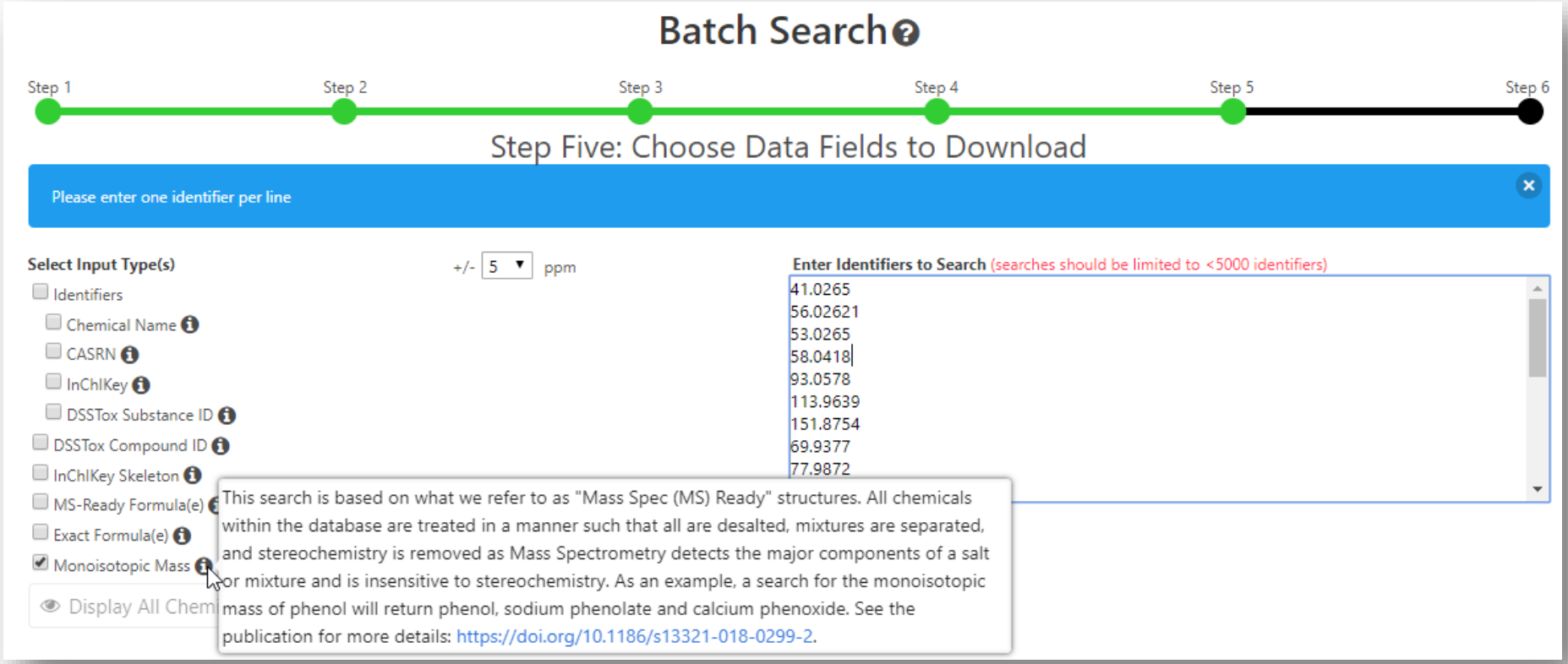
CMR. Query InChI...	src_id	Source	src_compound_id
...matches...	1	ChEMBL	CHEMBL444092
...matches...	4	Guide to Pharmacology	4735
...matches...	6	KEGG Ligand	C05371
...matches...	7	ChEBI	6925
...matches...	9	ZINC	ZINC000169715525
...matches...	9	ZINC	ZINC000255288110
...matches...	9	ZINC	ZINC000255288111
...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

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

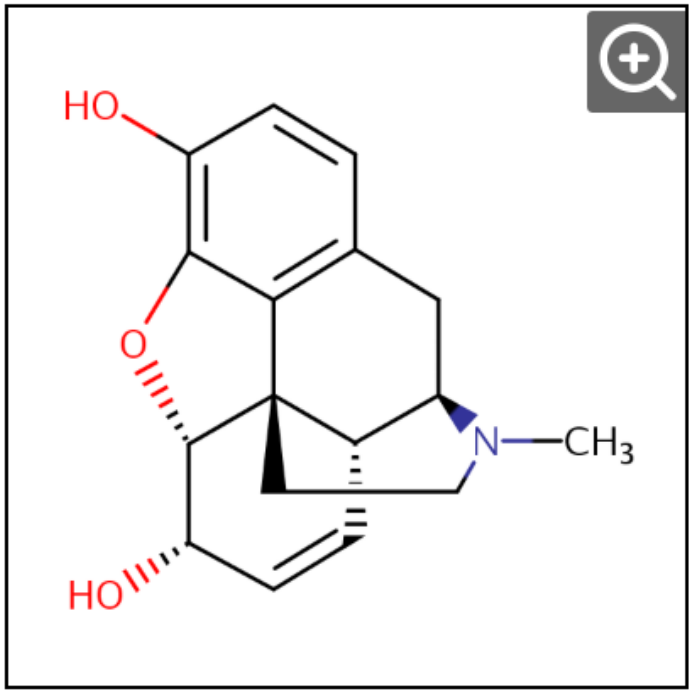
EPA United States Environmental Protection Agency

	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

Related Searches to Support Mass Spectrometry

Find me “related structures” 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

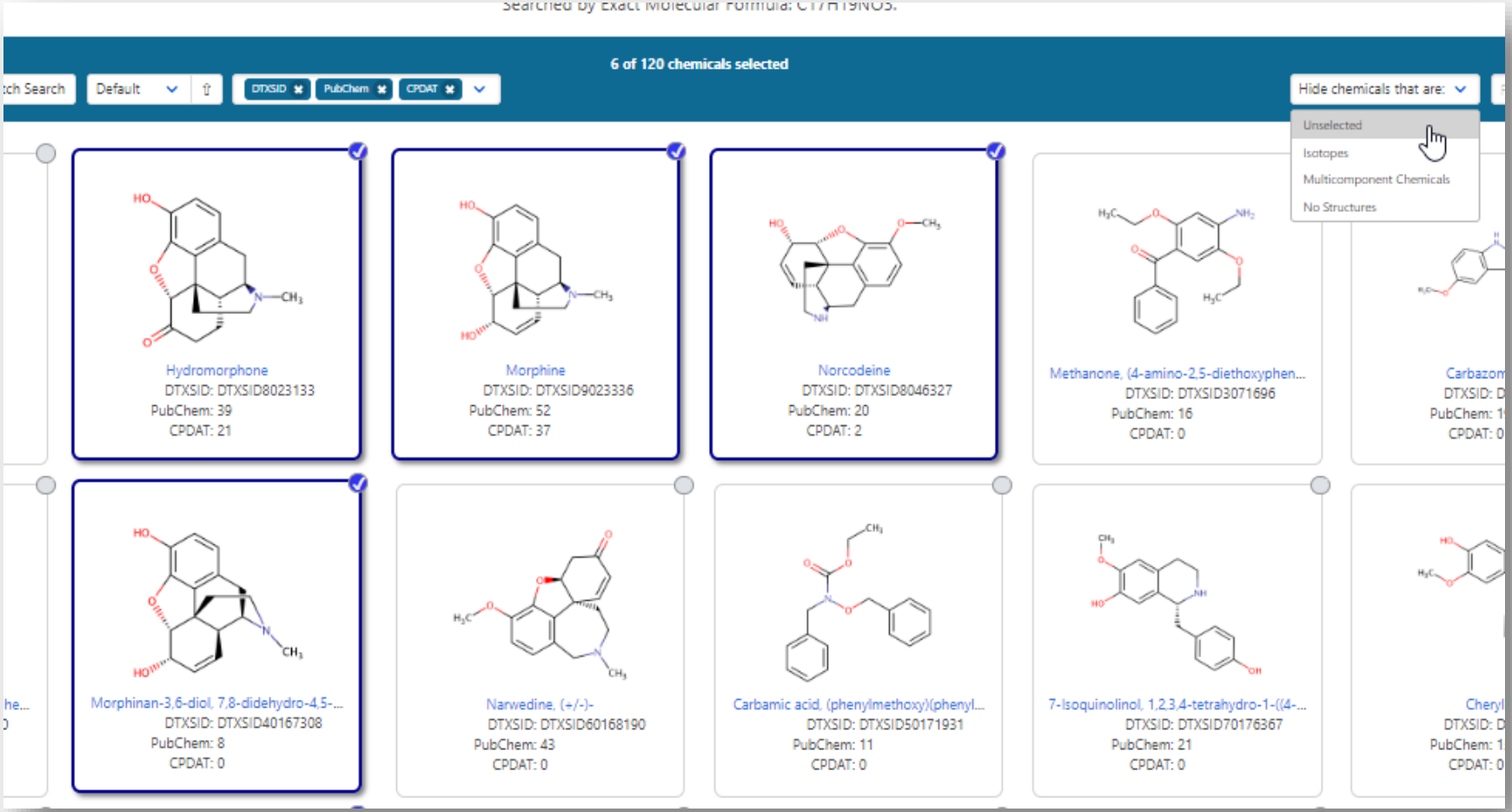
Molecular Formula: $C_{17}H_{19}NO_3$ Mol File Find All Chemicals

Average Mass: 285.343 g/mol Isotope Mass Distribution

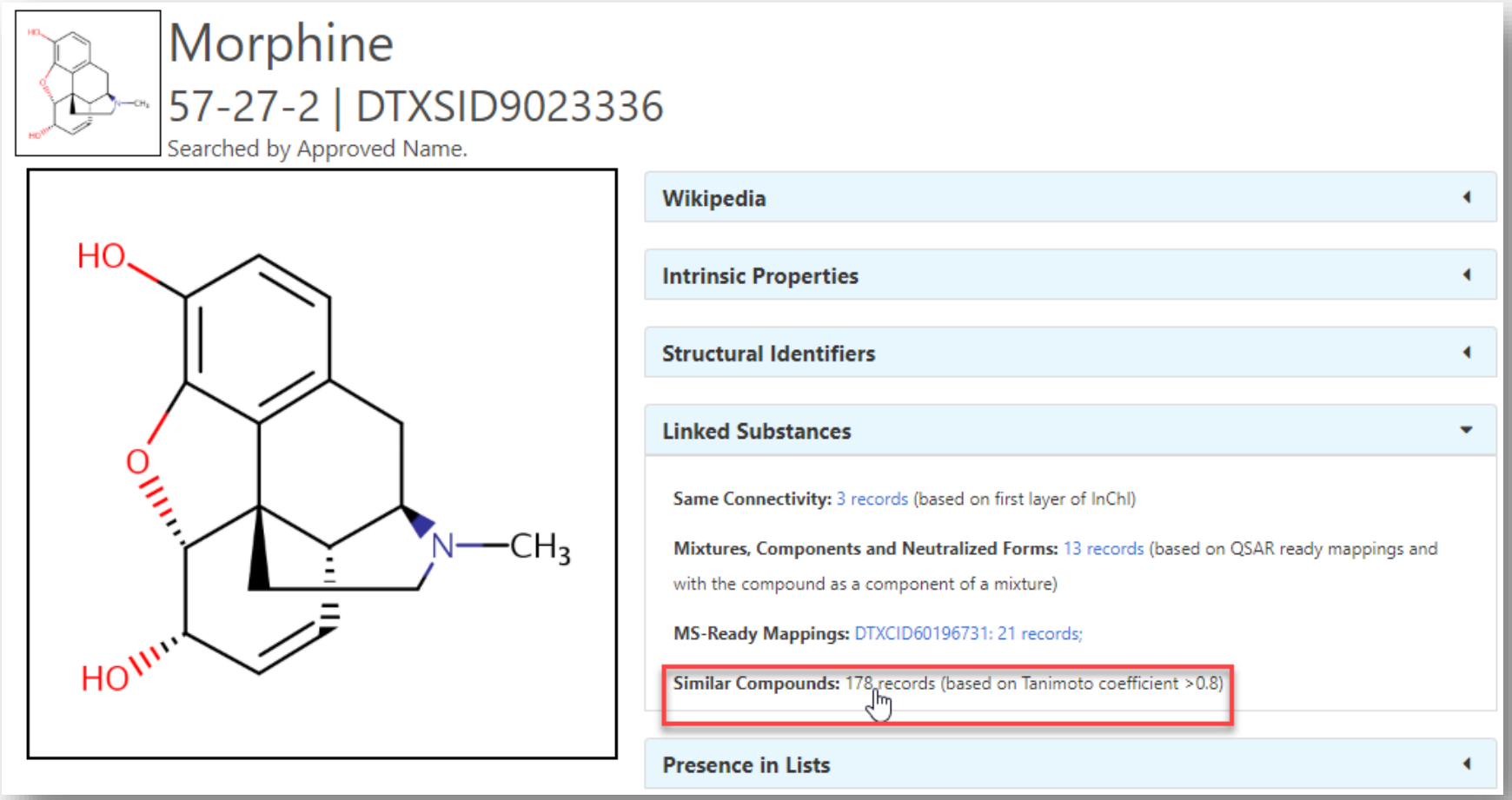
Monoisotopic Mass: 285.136493 g/mol

Structural Identifiers

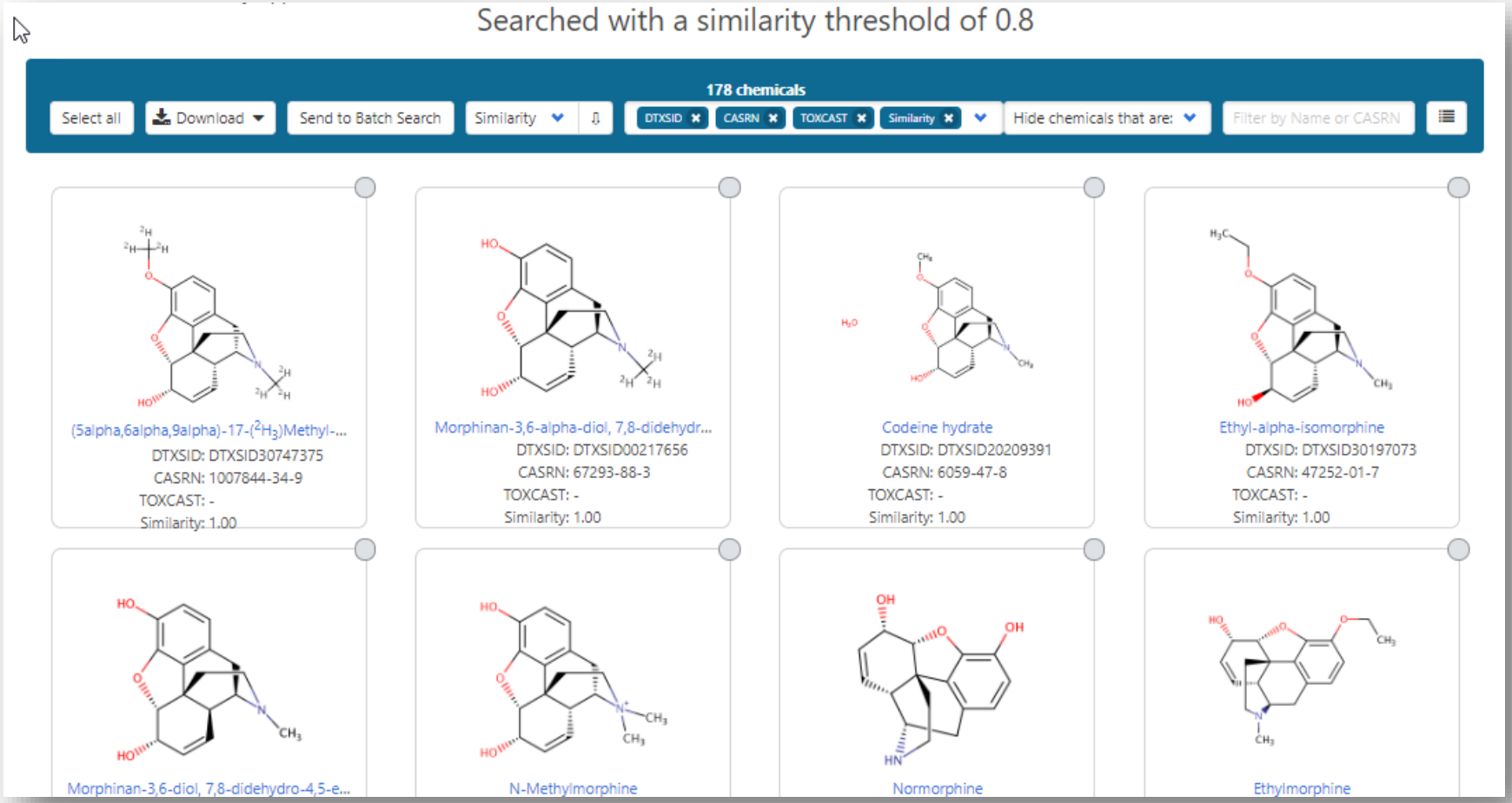
Select Chemicals of Interest



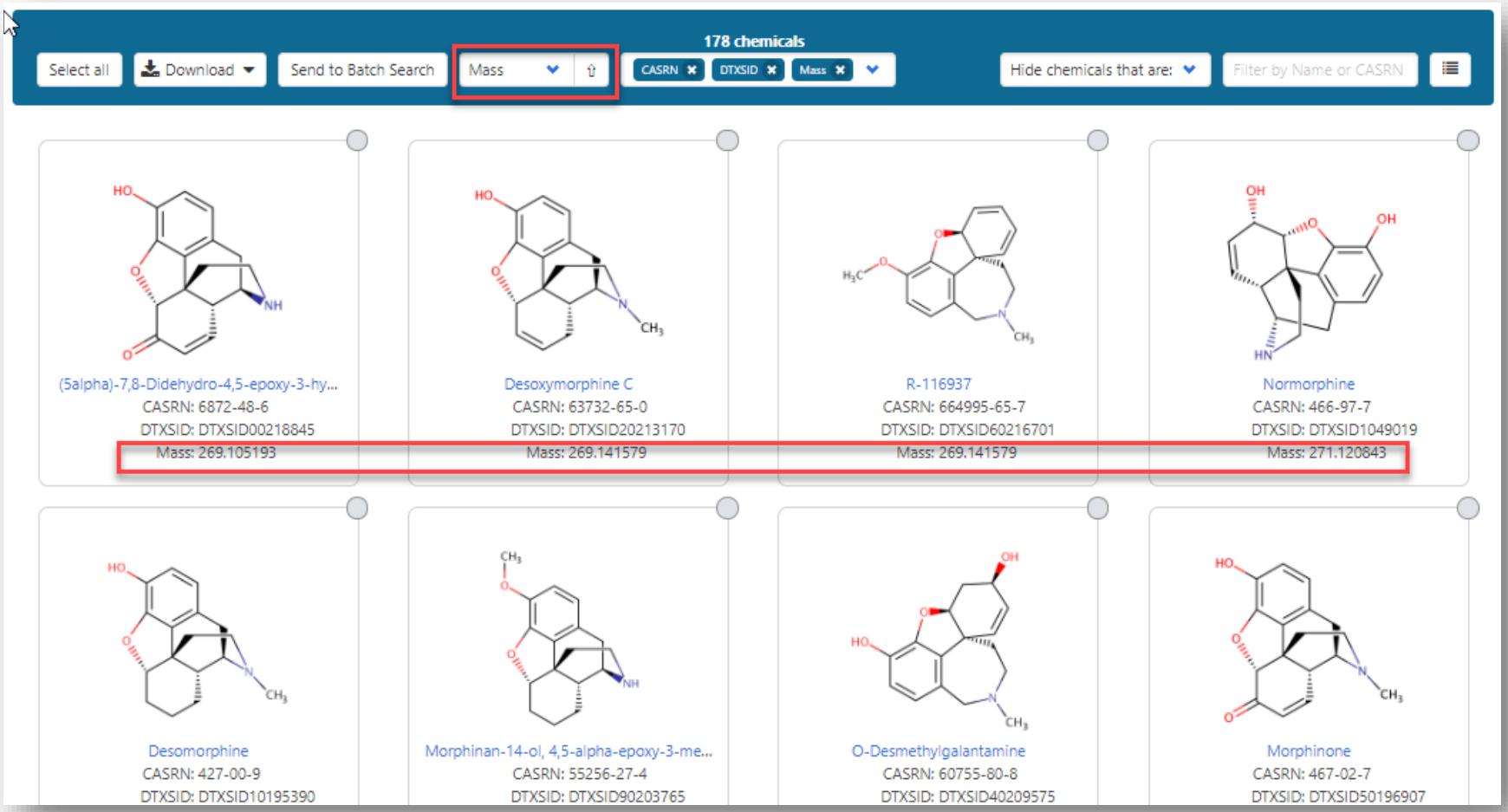
Find me “related structures” Based on Structure Similarity



Find me “related structures” Based on Structure Similarity

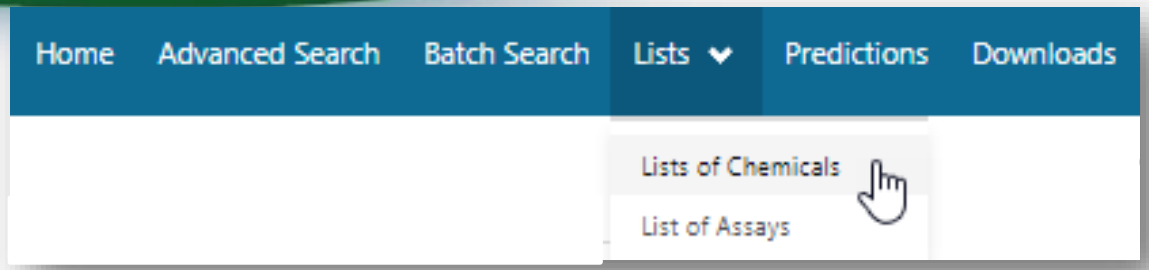


Find me “related structures”
Structure Similarity – sort on mass



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

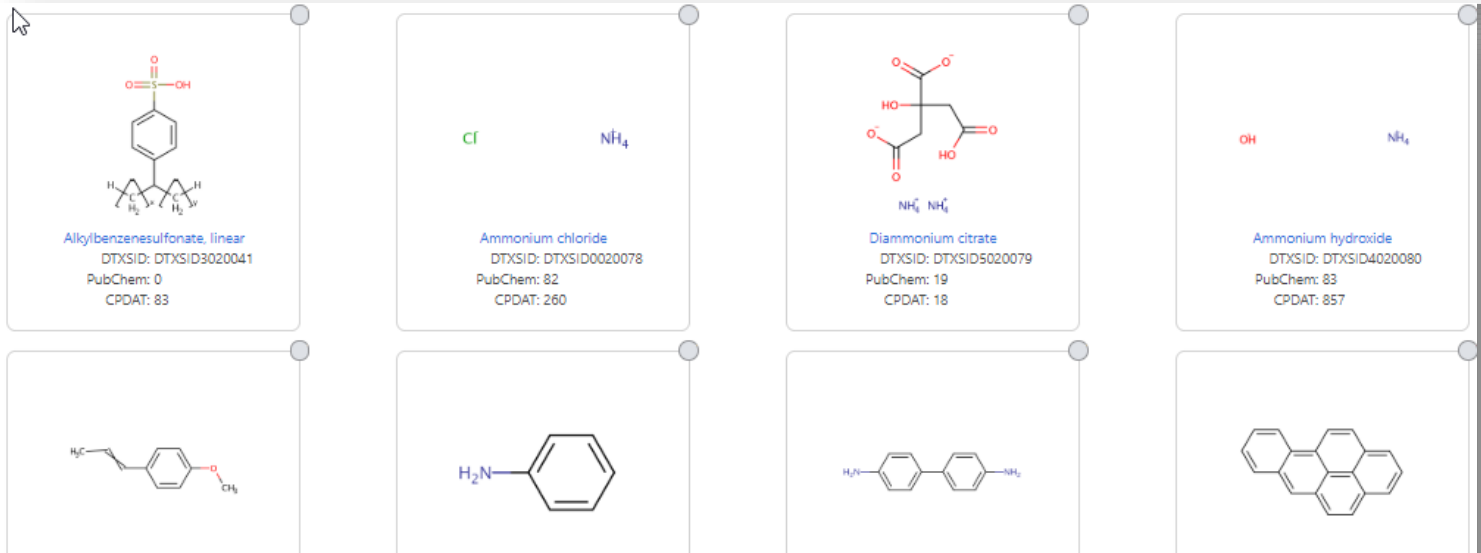
Identifier substring search

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



List of Opioids – Presence in Lists?

Trends in Environmental Analytical Chemistry

Volume 20, October 2018, e00059

Opioid occurrence in environmental water samples—A review

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

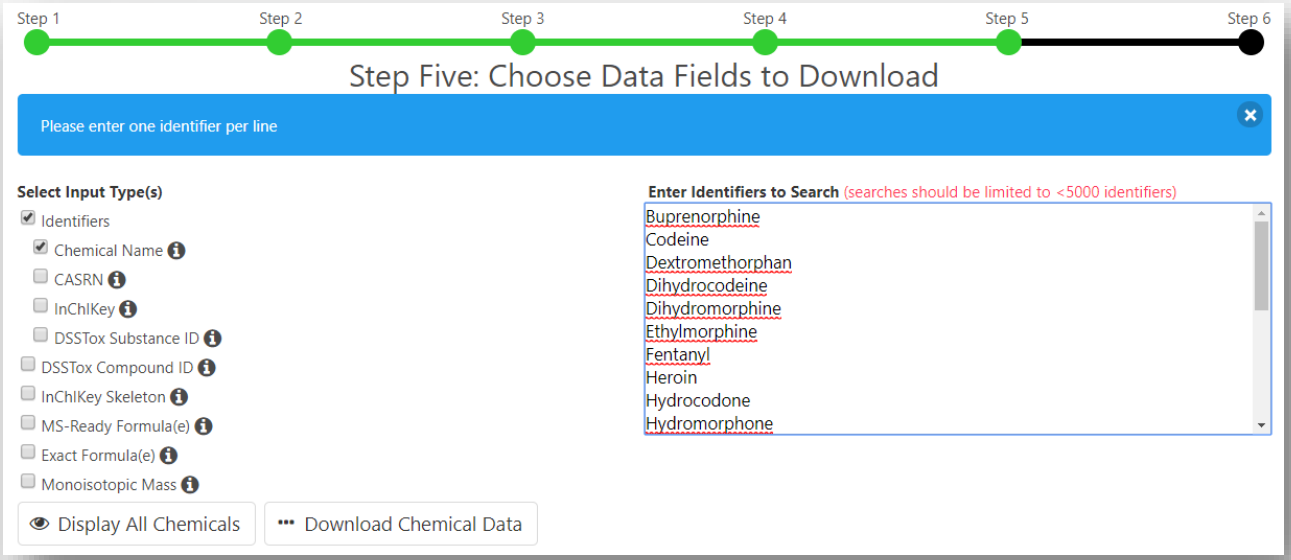
Show more

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

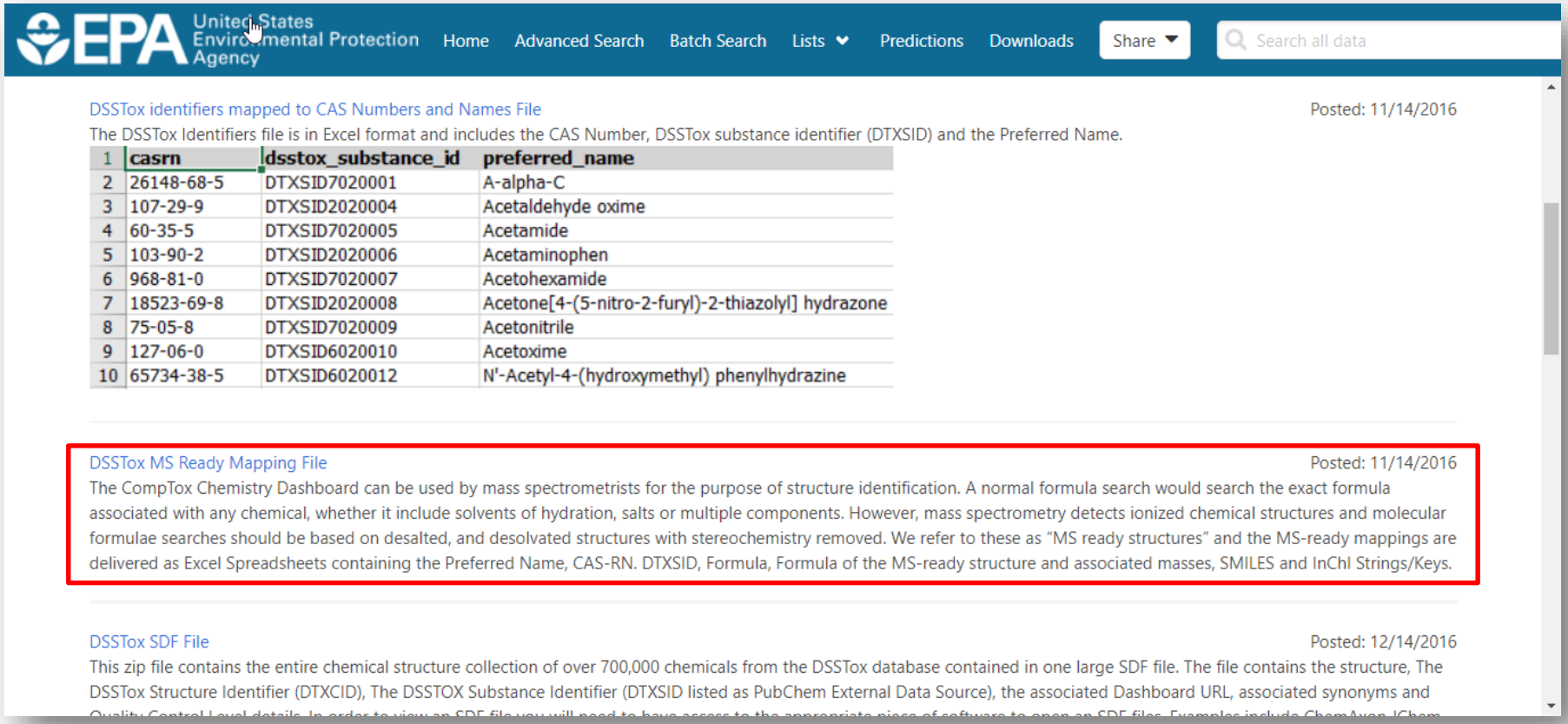
INPUT	FOUND_BY	DTXSID
Buprenorphine	Approved Name	DTXSID2022705
Codeine	Approved Name	DTXSID2020341
Dextromethorphan	Approved Name	DTXSID3022908
Dihydrocodeine	Approved Name	DTXSID5022936
Dihydromorphine	Approved Name	DTXSID7048908
Ethylmorphine	Approved Name	DTXSID1046760
Fentanyl	Approved Name	DTXSID9023049
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

API services and Open Data

- Available API and web services
- Open Data 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/

- Dozens of web services to provide access to data
- Data in UI, JSON and XML format

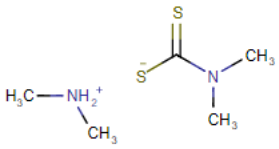
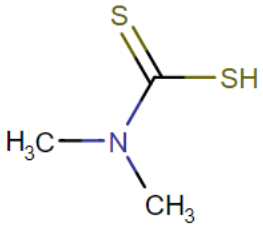
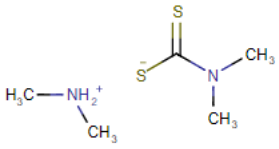
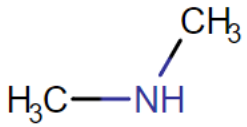
https://actorws.epa.gov/actorws/dsstox/v02/msready?identifier=80-05-7
https://actorws.epa.gov/actorws/dsstox/v02/msready.json?identifier=80-05-7
https://actorws.epa.gov/actorws/dsstox/v02/msready.xml?identifier=80-05-7

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

Example: 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

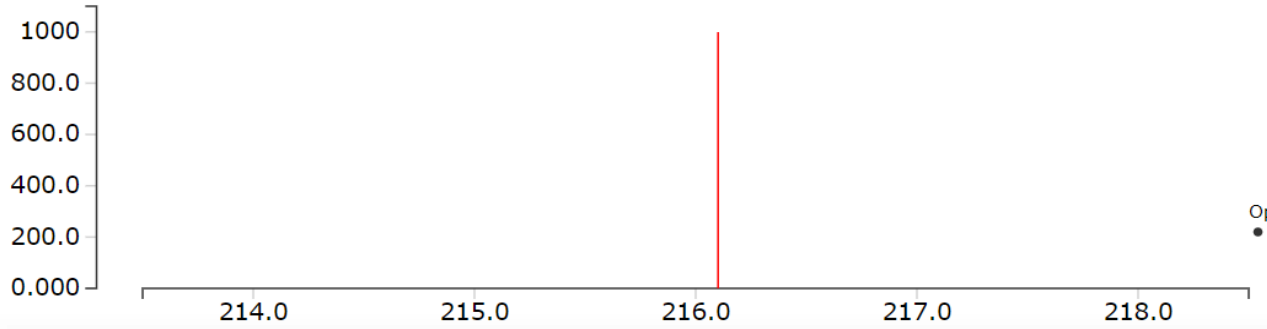
MassBank mapping to Dashboard

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



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

EPA United States Environmental Protection Agency

Benefits of Open Data

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

EPA United States Environmental Protection Agency

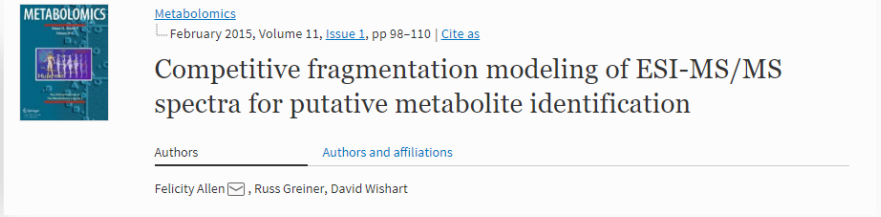
In Progress

Work in Progress

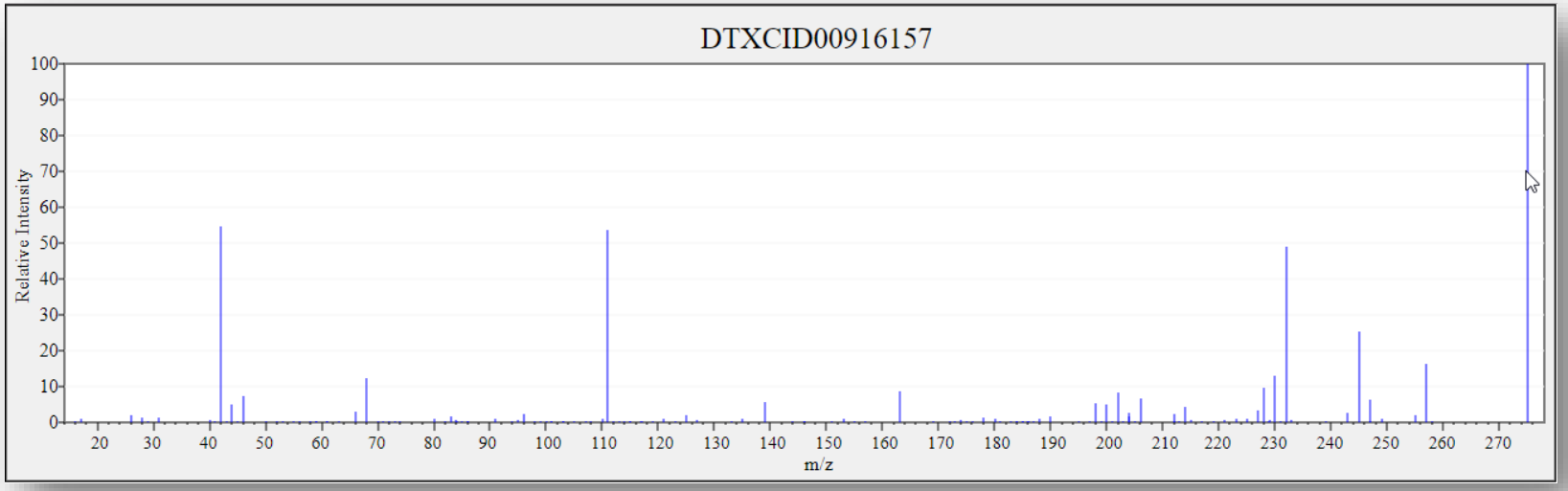
- Predicted Spectra for candidate ranking
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database

Predicted Mass Spectra

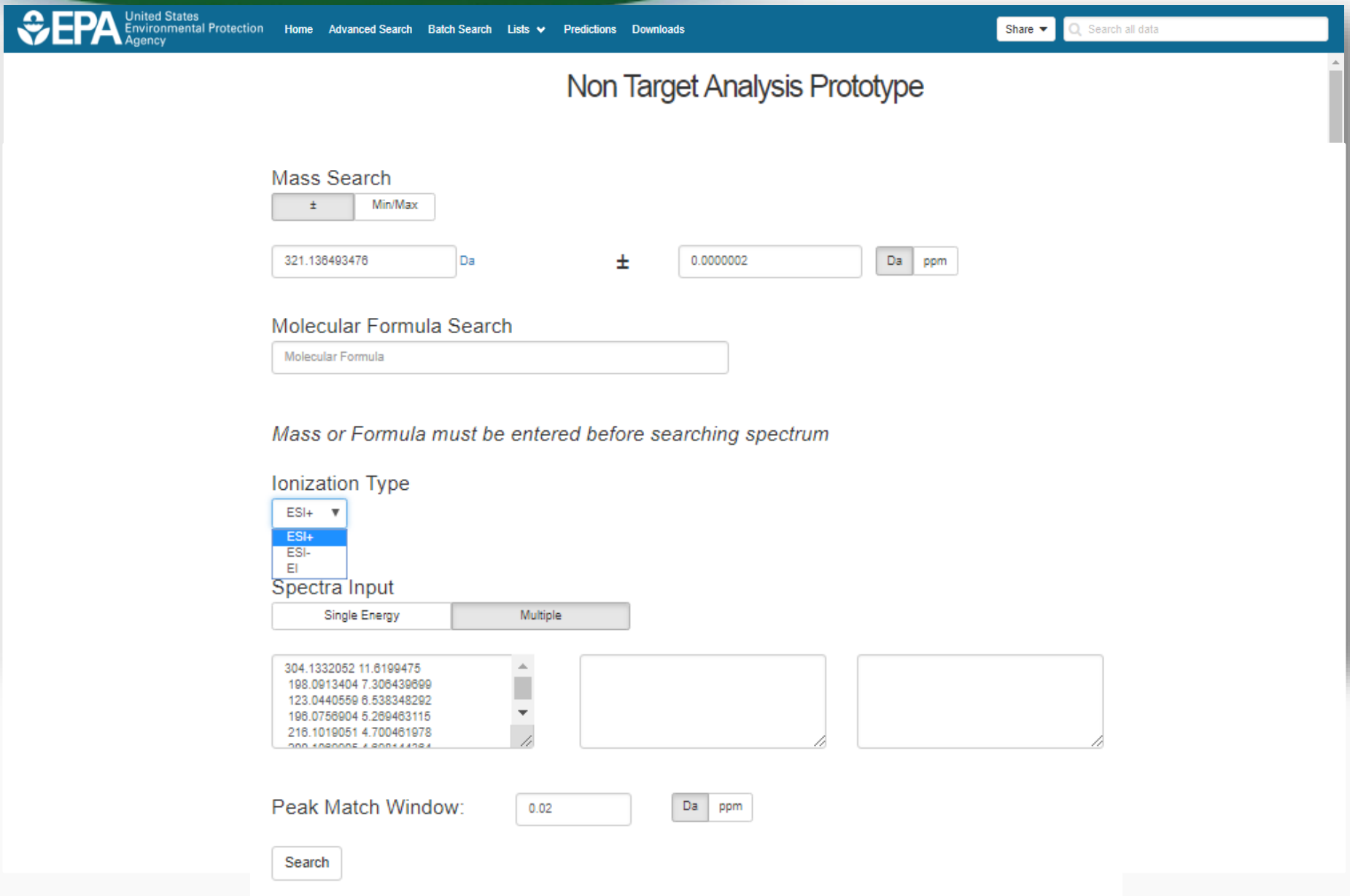
http://cfmid.wishartlab.com/



- 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



CFM-ID Predicted Library Available

- Predictions generated and stored for >700,000 structures
- Python code to score experimental vs predicted spectra
- Cosine dot product match score calculation

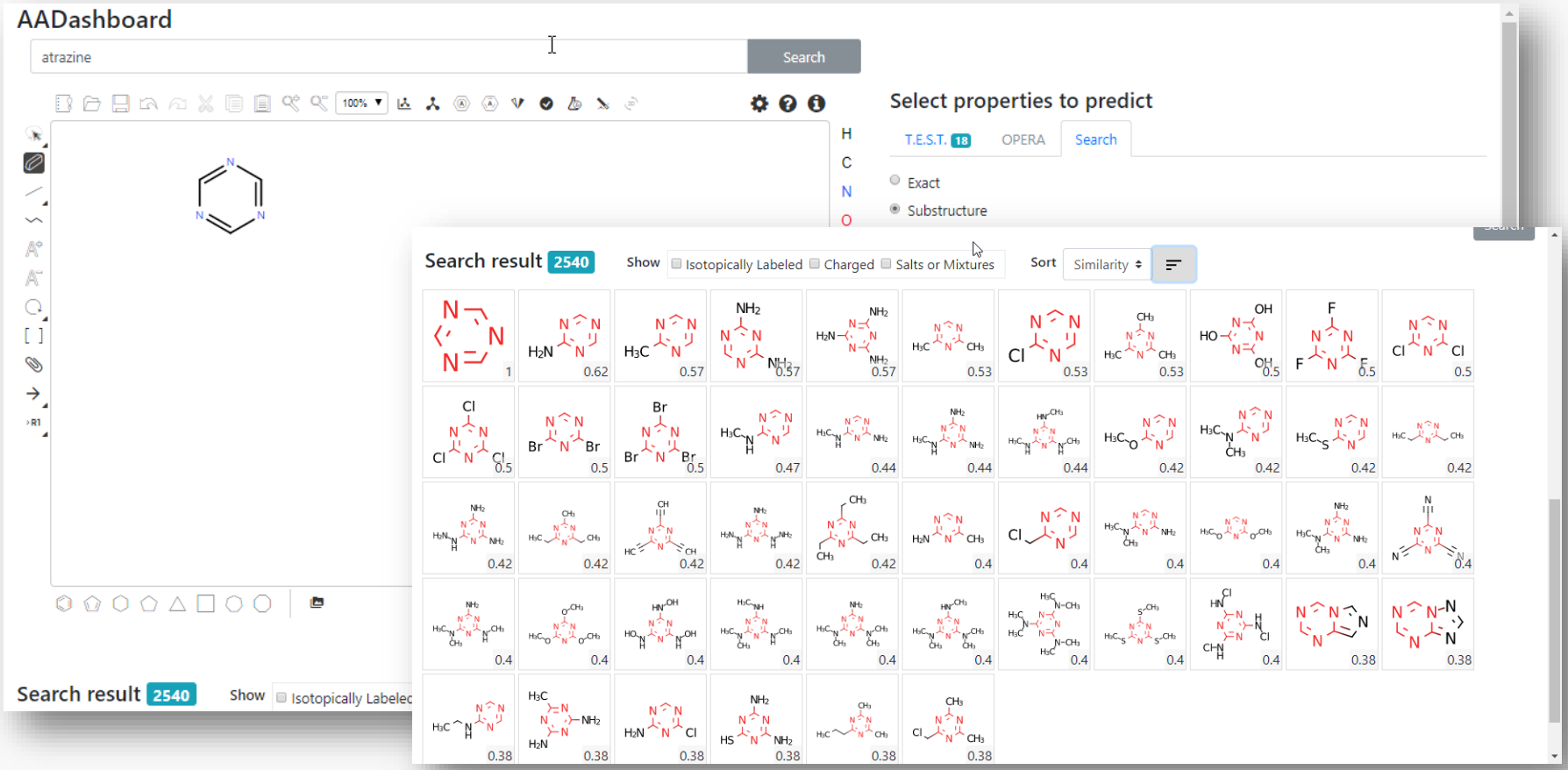
Data Descriptor | OPEN | Published: 02 August 2019

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

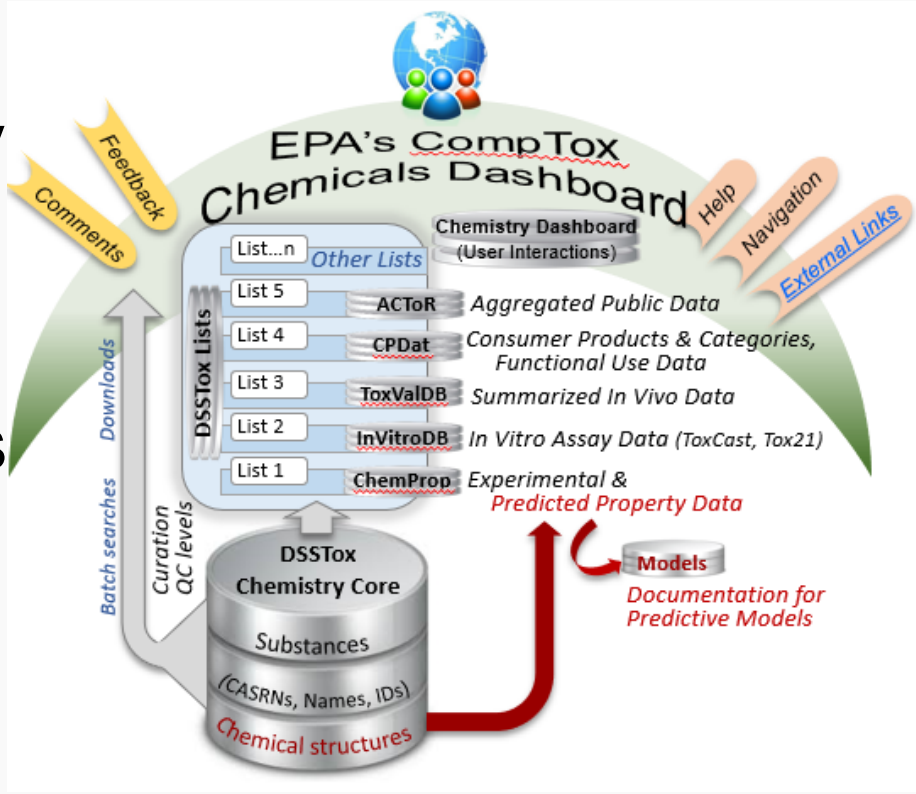
Scientific Data **6**, Article number: 141 (2019) | Download Citation

Prototype Development
Structure/substructure search



Conclusion

- Dashboard access to data for ~875,000 chemicals
- MS-Ready data facilitates structure identification
- Related metadata facilitates candidate ranking
- Relationship mappings and chemical lists of great utility
- Dashboard and contents are one part of the solution
- New API and Web Services are in development



Acknowledgements

- *NCCT IT development team*
- *Tommy Cathey, ACTOR Web Services*
- *Nancy Baker, Abstract Sifter*
- *Todd Martin & Valery Tkachenko, WebTEST*
- *Kathie Dionisio & Kristin Isaacs, CPDat*
- *Thanks to Emma Schymanski, University of Luxembourg, for coordinating all efforts with the NORMAN Network for curation of lists on the Suspect Exchange*

Contact

Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology
EMAIL: Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821