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

Applications of the US-EPA CompTox Chemicals Dashboard to source pesticides data

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

May 2021

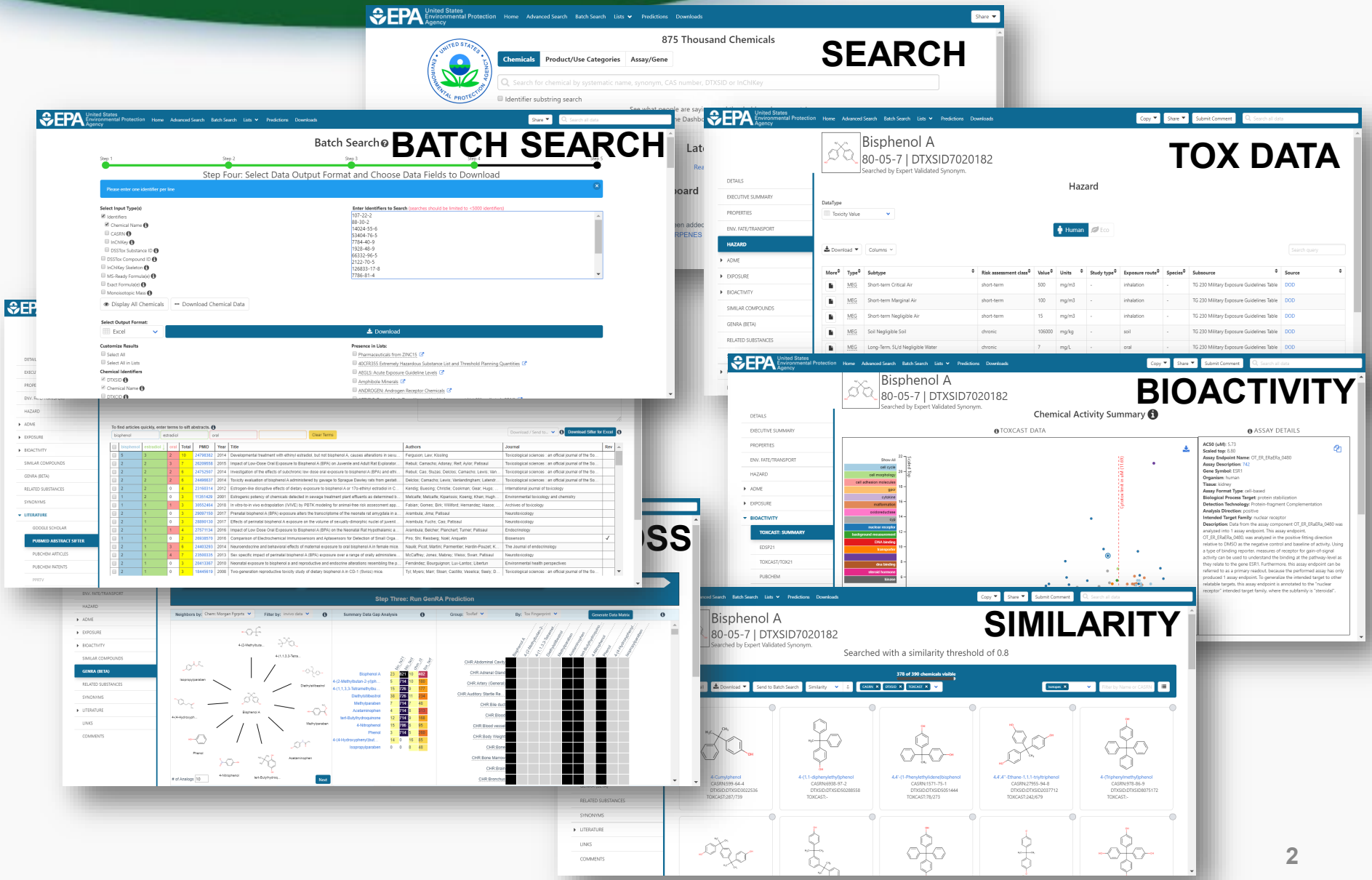
Latin American Pesticide Residue Workshop

Outline

- The Dashboard, the data and functionality
- Pesticide lists on the Dashboard
- Transformation product mappings
- Single searches vs Batch search
- Support for Mass Spectrometry
- Work in progress – prototypes

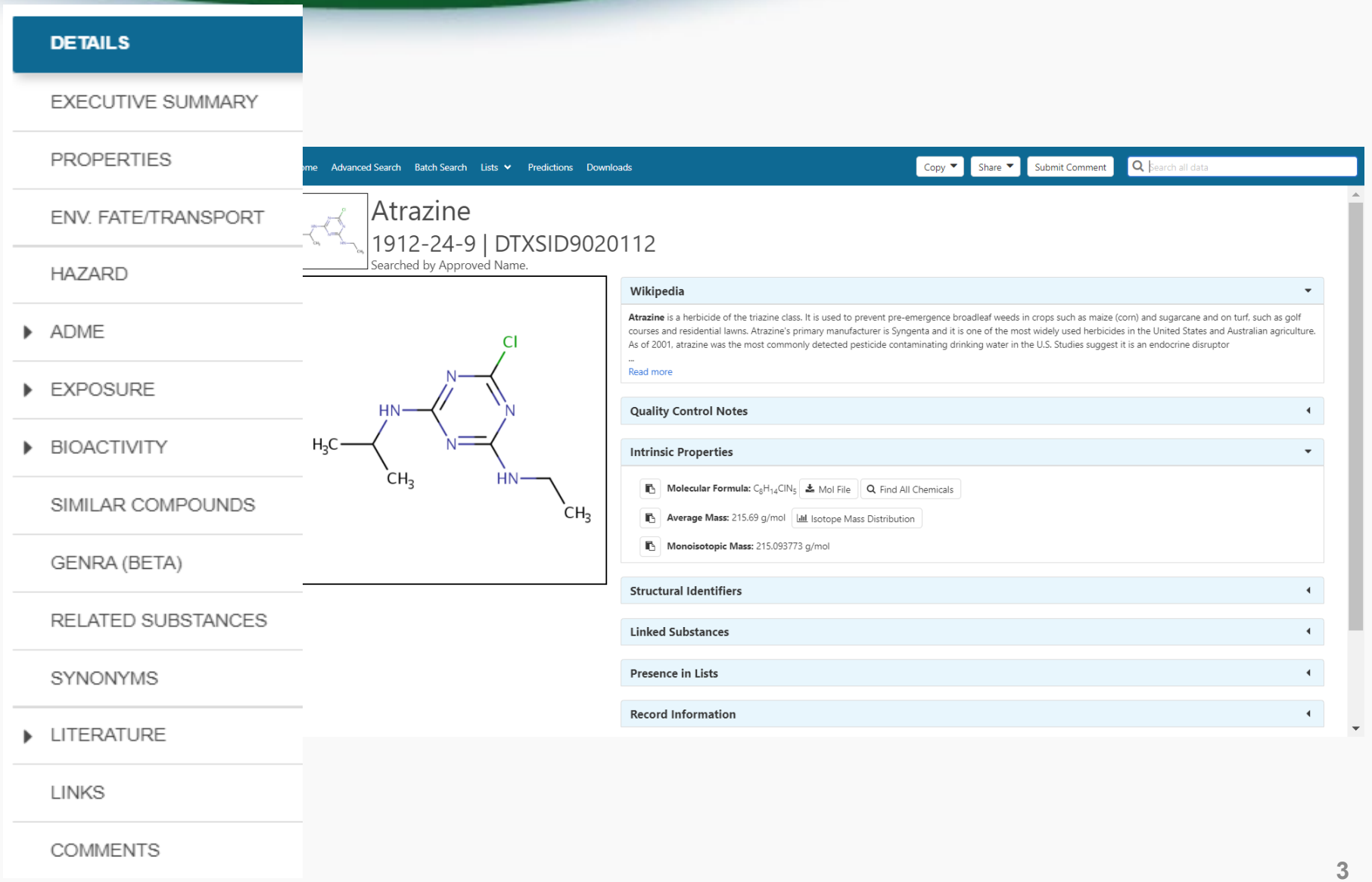
CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard

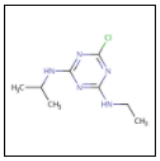


Detailed Chemical Pages for ~900,000 chemicals

EPA United States Environmental Protection Agency



Properties, Fate and Transport



Atrazine

1912-24-9 | DTXSID9020112

Searched by Approved Name.

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

Property

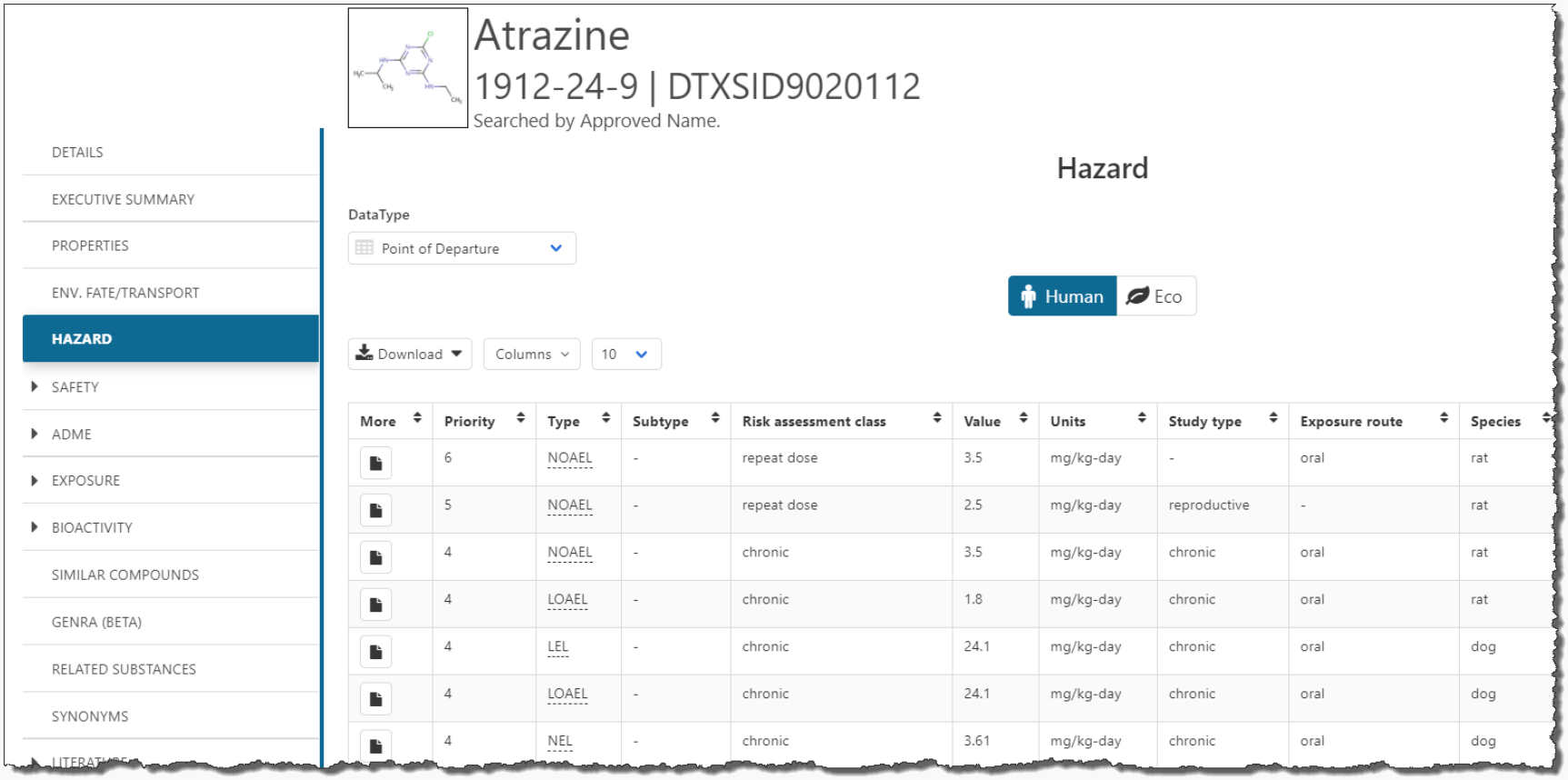
Summary

Summary

Download | Columns

Property	Experimental average	Predicted average	Experimental median	Predicted median
LogKow: Octanol-Water	2.61 (1)	2.68		2.65
Melting Point	174 (6)	147	174	154
Boiling Point	-	333		325
Water Solubility	2.12e-4 (5)	1.76	1.62e-4	7.17e-4
Vapor Pressure	2.89e-7 (1)	4.49e-6		4.97e-7
Flash Point	-	176		176
Thermal Conductivity	-	158		
Density	-	1.25		1.25
Surface Tension	-	53.8		
Index of Refraction	-	1.60		

Hazard



Atrazine

1912-24-9 | DTXSID9020112

Searched by Approved Name.

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

Hazard

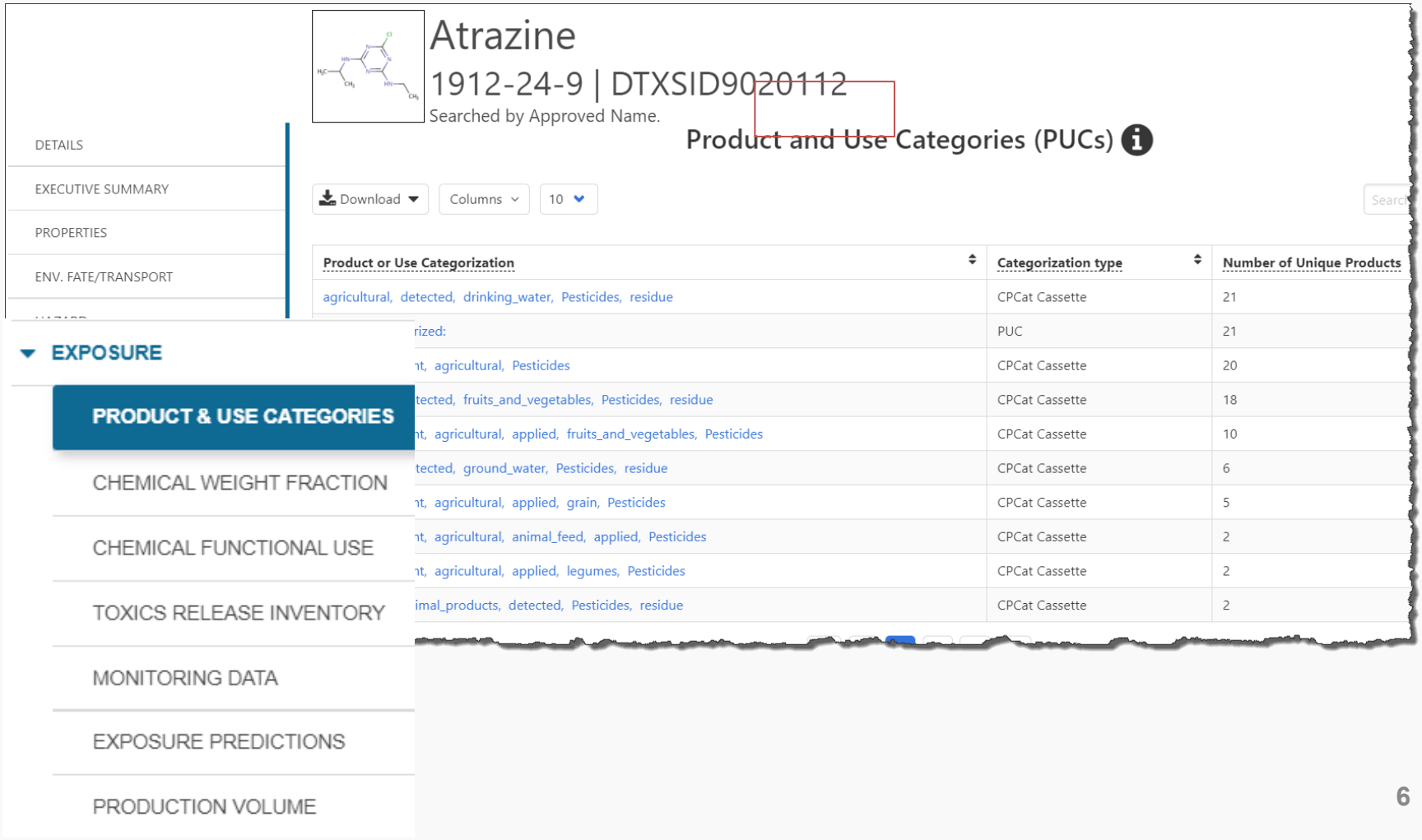
DataType: Point of Departure

Human | Eco

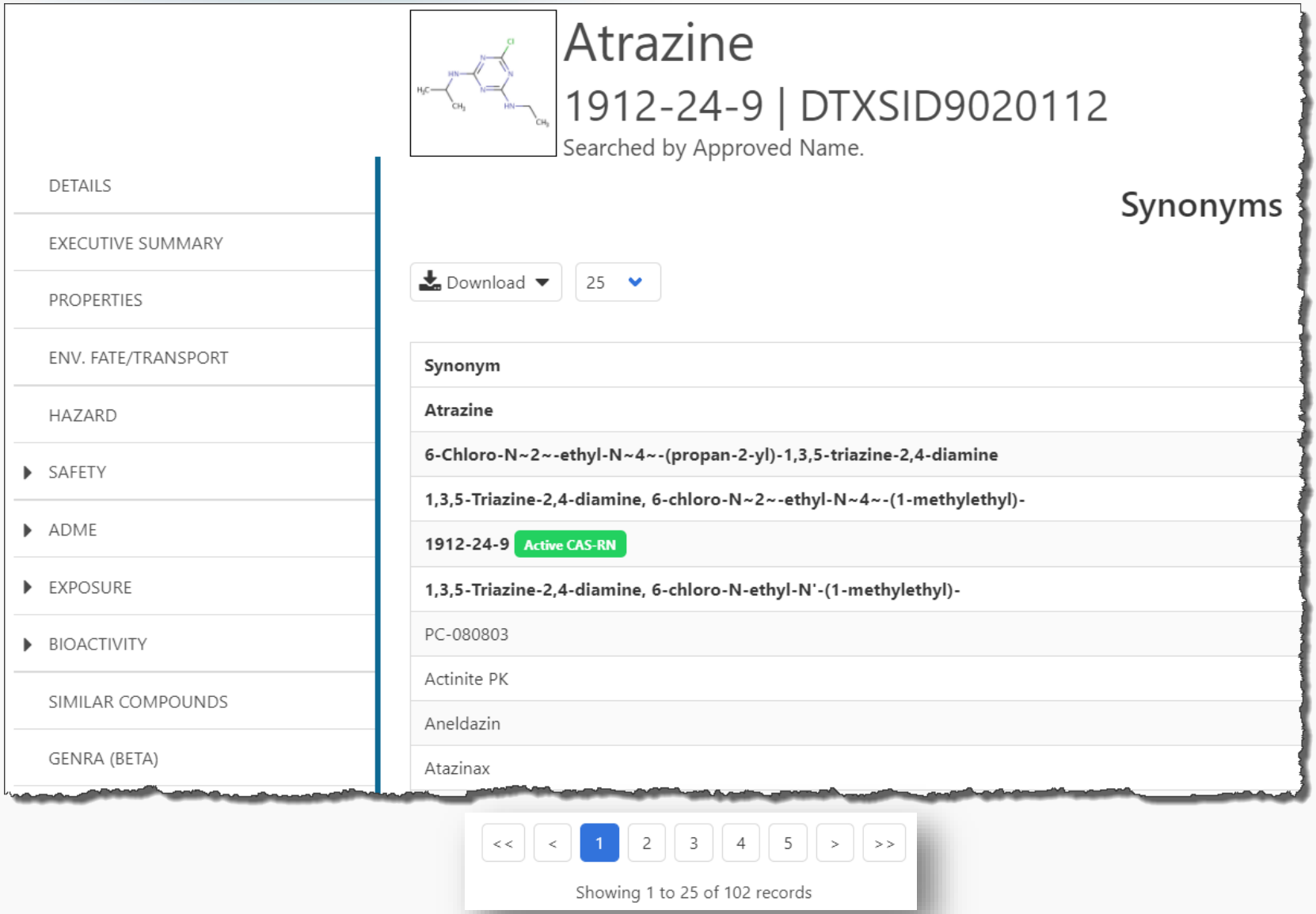
Download | Columns | 10

More	Priority	Type	Subtype	Risk assessment class	Value	Units	Study type	Exposure route	Species
	6	NOAEL	-	repeat dose	3.5	mg/kg-day	-	oral	rat
	5	NOAEL	-	repeat dose	2.5	mg/kg-day	reproductive	-	rat
	4	NOAEL	-	chronic	3.5	mg/kg-day	chronic	oral	rat
	4	LOAEL	-	chronic	1.8	mg/kg-day	chronic	oral	rat
	4	LEL	-	chronic	24.1	mg/kg-day	chronic	oral	dog
	4	LOAEL	-	chronic	24.1	mg/kg-day	chronic	oral	dog
	4	NEL	-	chronic	3.61	mg/kg-day	chronic	oral	dog

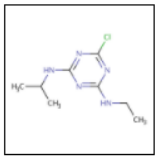
Sources of Exposure to Chemicals



Identifiers to Support Searches



Link Access



Atrazine

1912-24-9 | DTXSID9020112

Searched by Approved Name.

General

- EPA Substance Registry Service
- PubChem
- Chemspider
- CPCat
- DrugBank
- Wikipedia
- MSDS Lookup
- ChEMBL
- ToxPlanet
- ACS Reagent Chemicals
- Wolfram Alpha
- ECHA Infocard
- ChemAgora
- Consumer Product Information Database
- ChEBI
- NIST Chemistry Webbook
- WEBWISER
- PubChem Safety Sheet

Toxicology

- ACToR
- DrugPortal
- CCRIS
- ChemView
- CTD
- eChemPortal
- Gene-Tox
- HSDB
- ACToR PDF Report
- CREST
- National Air Toxics Assessment
- ChemView
- Chemical Checker
- BindingDB
- CalEPA OEHHA
- NIOSH IDLH Values
- LactMed
- ECOTOX

Publications

- Toxline
- PPRTVWEB
- PubMed
- IRIS Assessments
- EPA HERO
- NIOSH Skin Notation Profiles
- NIOSH Pocket Guide
- RSC Publications
- BioCaddie DataMed
- Springer Materials
- Bielefeld Academic Search Engine
- CORE Literature Search
- Google Books (Text Search)
- Google Patents (Text search)
- Google Scholar (Text search)
- Google Patents (Structure search)
- Google Books (Structure Search)
- Google Scholar (Structure search)

Analytical

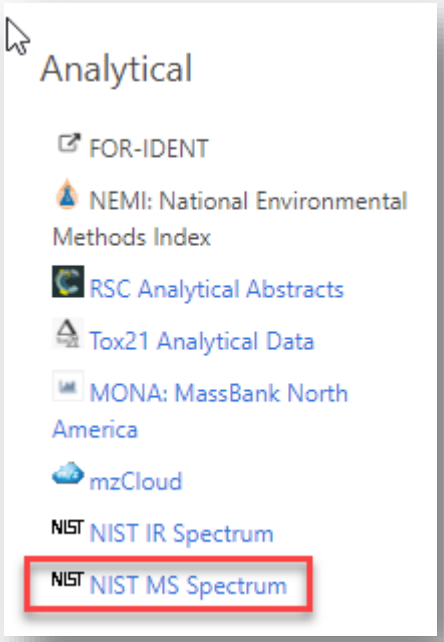
- RSC Analytical Abstracts
- Tox21 Analytical Data
- MONA: MassBank North America
- mzCloud
- NIST IR Spectrum
- NIST MS Spectrum
- MassBank
- NIST Antoine Constants
- IR Spectra on PubChem
- NIST Kovats Index values
- Protein DataBank
- National Environmental Methods Index

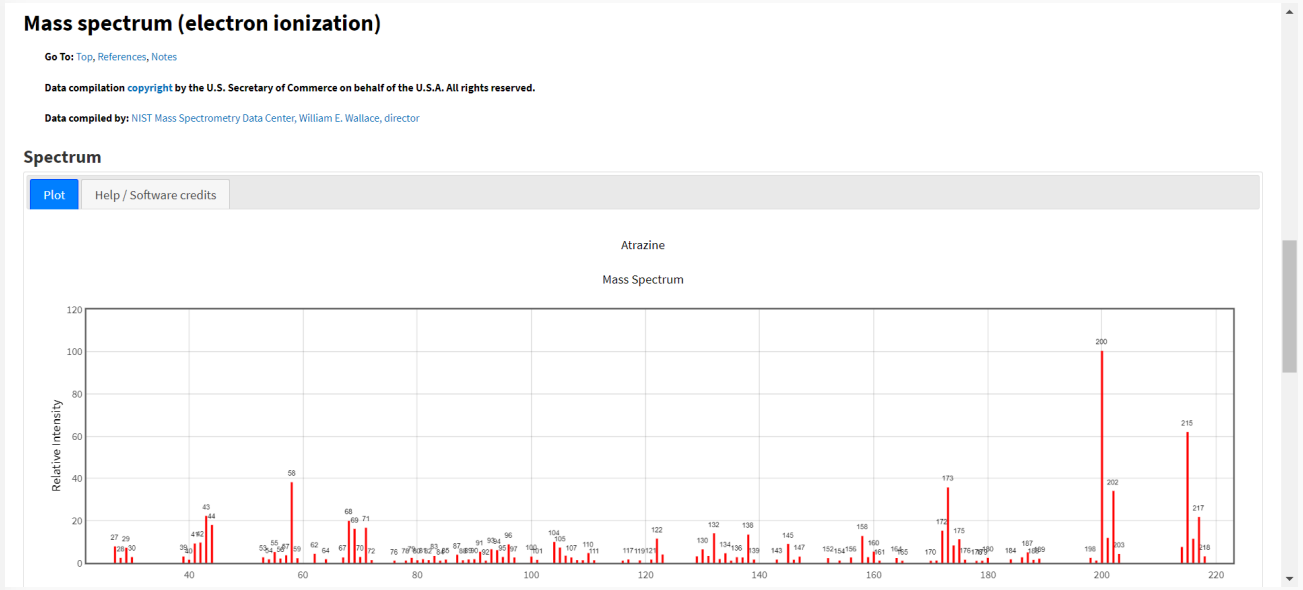
Prediction

- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRTP Predictor
- LSERD

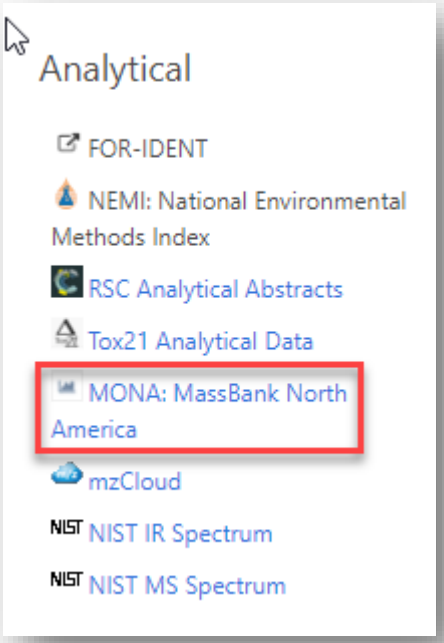
NIST WebBook
https://webbook.nist.gov/chemistry/

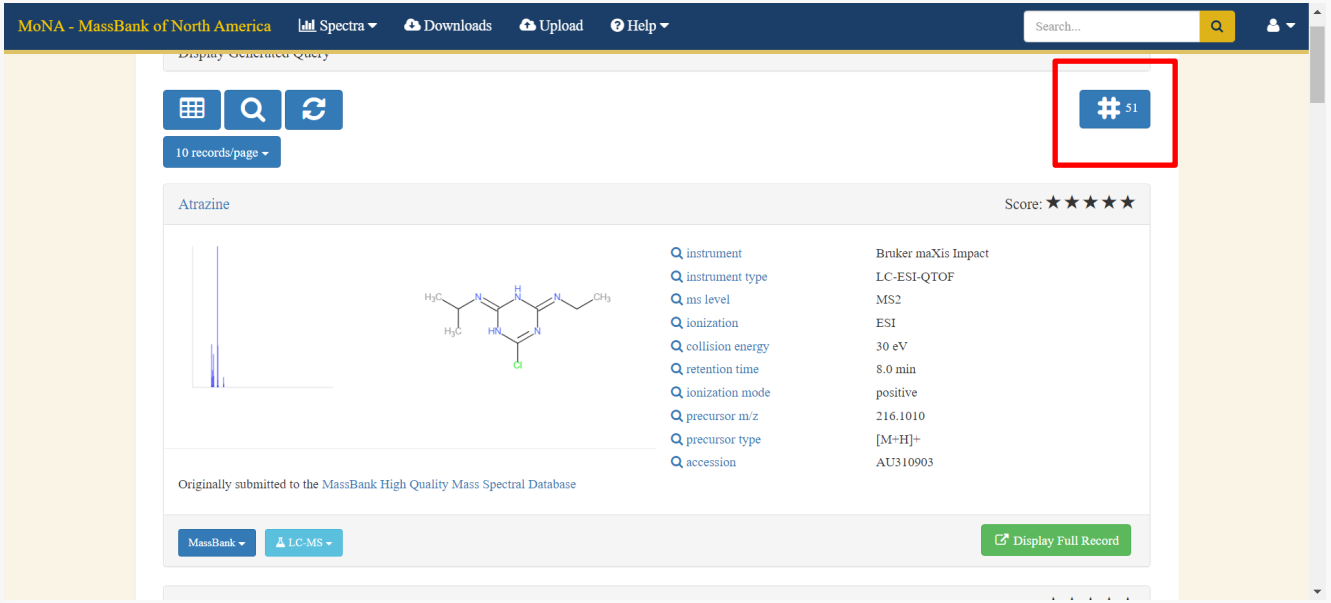
EPA United States Environmental Protection Agency



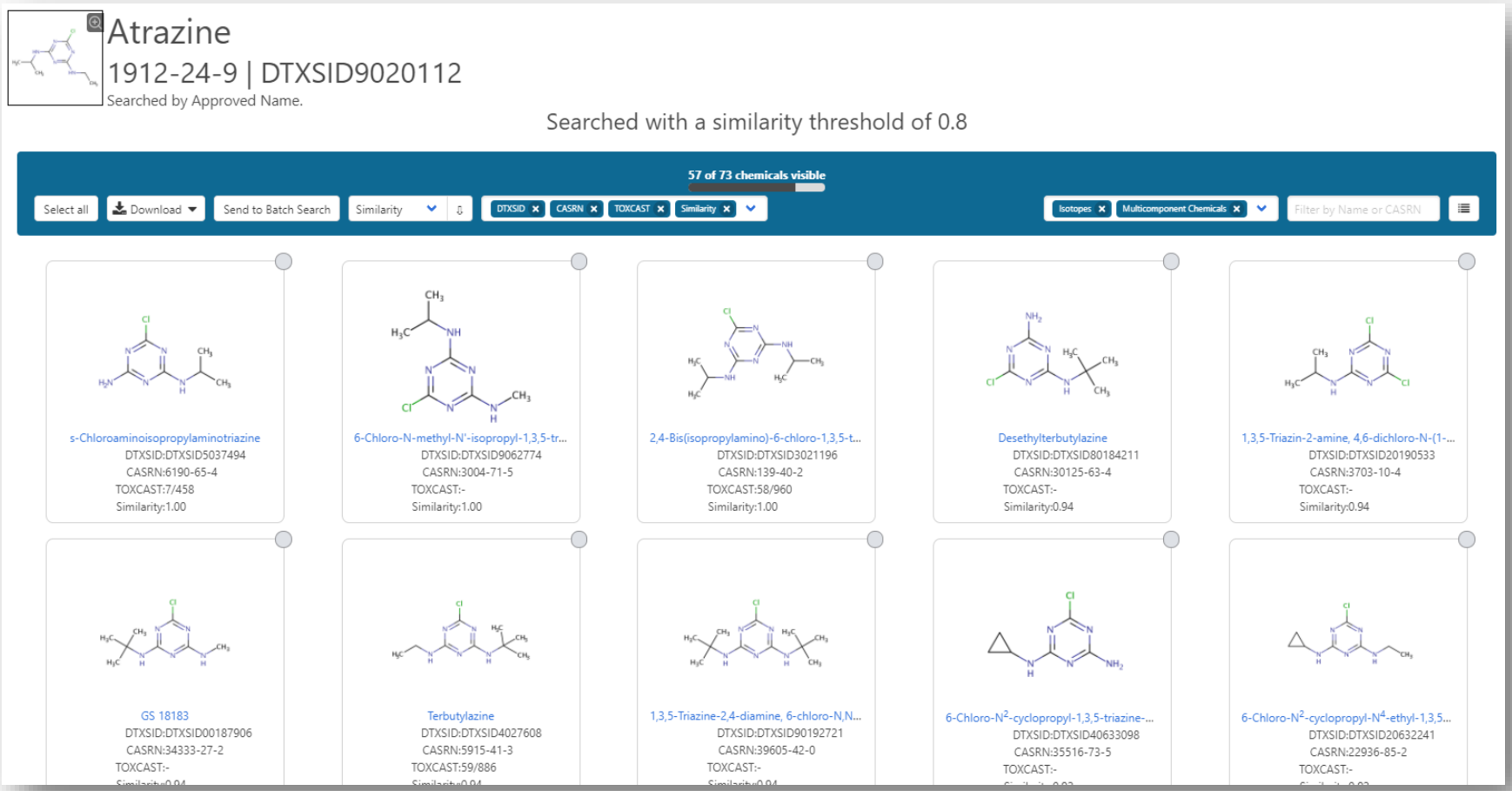


MassBank of North America
https://mona.fiehnlab.ucdavis.edu

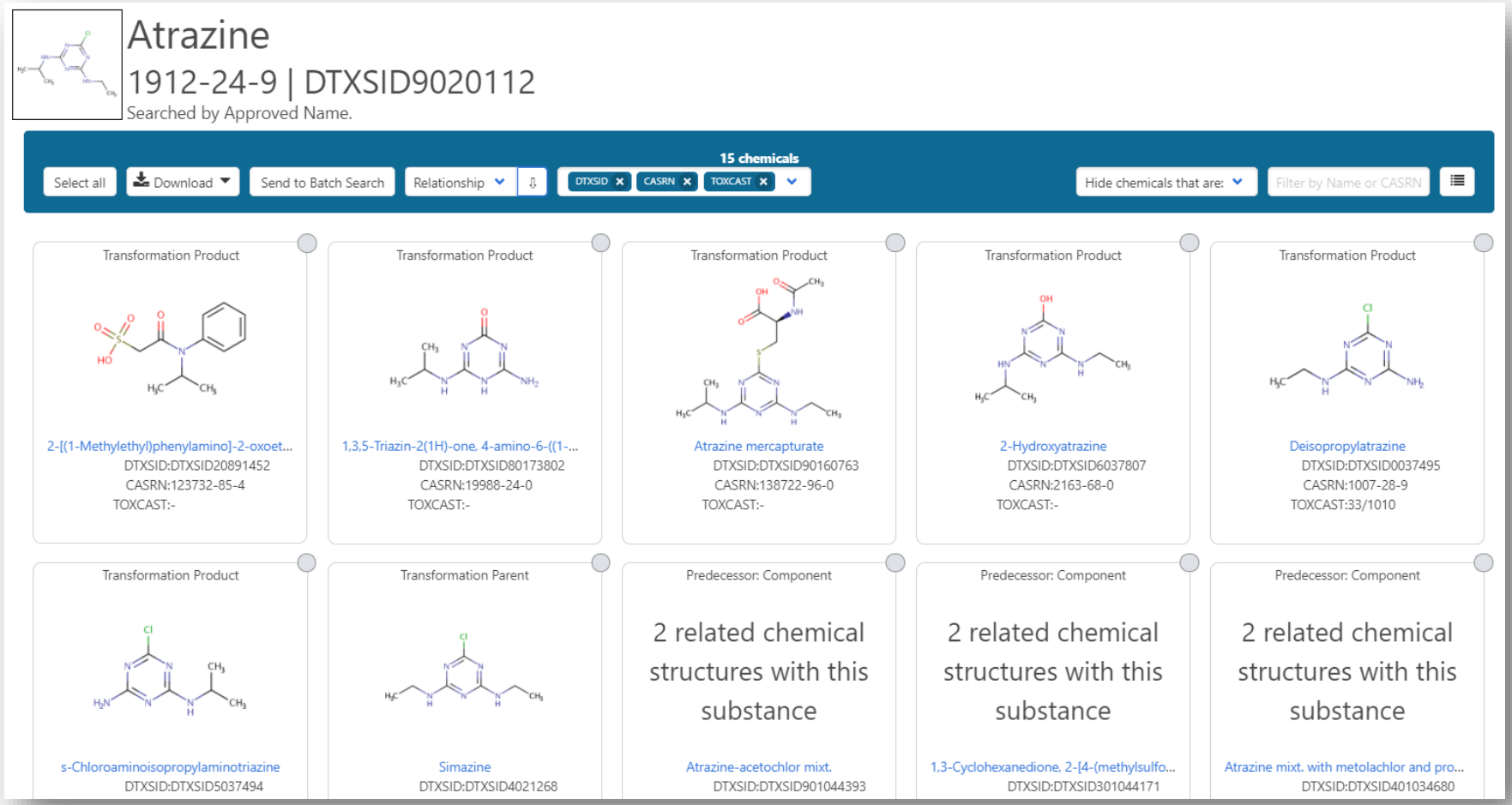




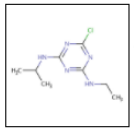
Structurally Similar Compounds



Related Substances: Transformation products, metabolites, mixtures



Literature searching of PubMed



Atrazine

1912-24-9 | DTXSID9020112

Searched by Approved Name.

Abstract Sifter

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

Metabolism/PK/PD | Retrieve Articles

2852 of 2852 articles loaded...

Optionally, edit the query before retrieving.

("1912-24-9" OR "Atrazine") AND (metabolism OR metabolite OR tissue distribution OR pharmacokinetics OR pharmacodynamics)

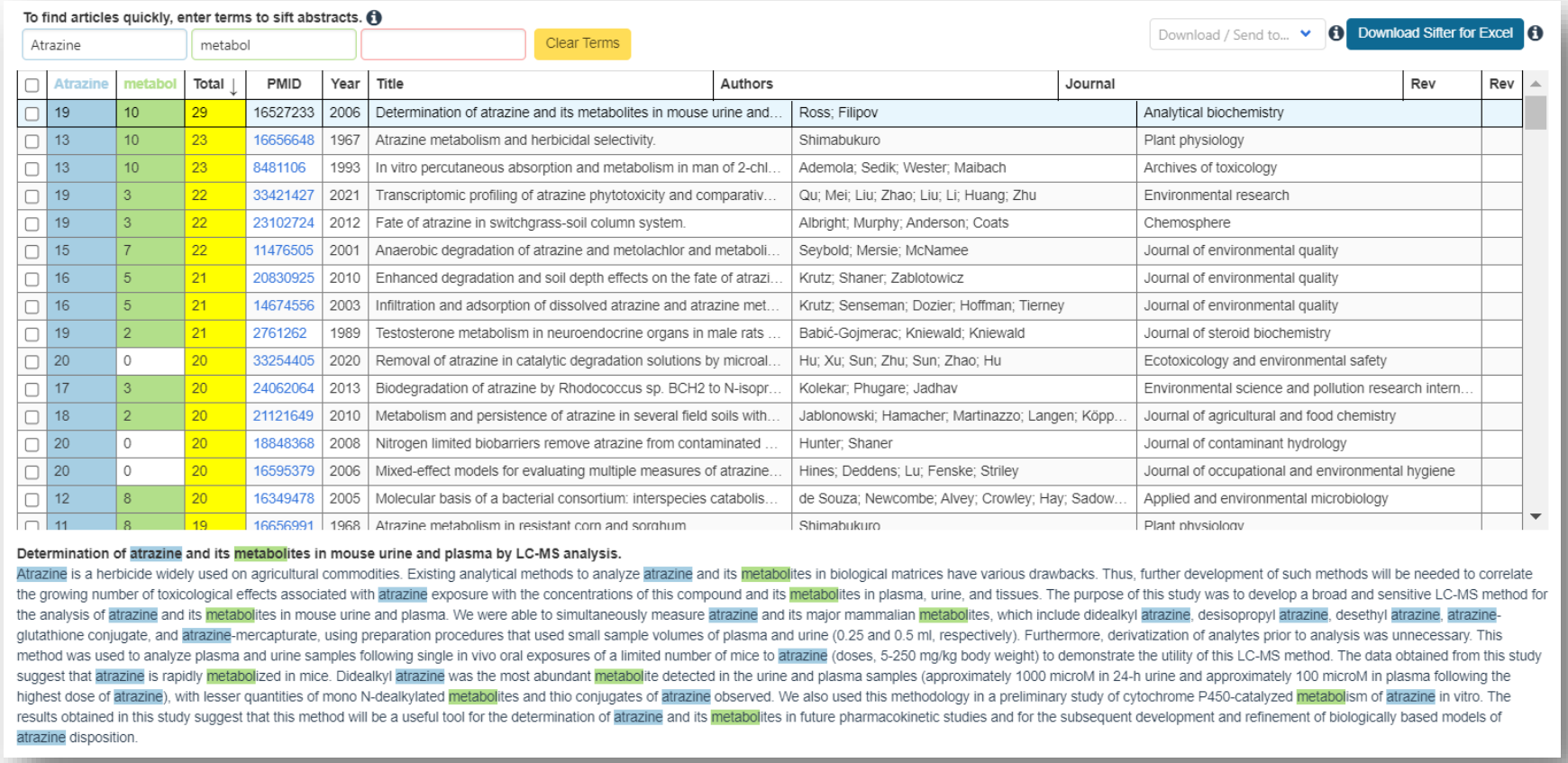
To find articles quickly, enter terms to sift abstracts.

Atrazine | metabol | | Clear Terms

Download / Send to... | Download Sifter for Excel

	Atrazine	metabol	Total ↓	PMID	Year	Title	Authors	Journal	Rev	Rev
☐	19	10	29	16527233	2006	Determination of atrazine and its metabolites in mouse urine and…	Ross; Filipov	Analytical biochemistry		
☐	13	10	23	16656648	1967	Atrazine metabolism and herbicidal selectivity.	Shimabukuro	Plant physiology		
☐	13	10	23	8481106	1993	In vitro percutaneous absorption and metabolism in man of 2-chl…	Ademola; Sedik; Wester; Maibach	Archives of toxicology		
☐	19	3	22	33421427	2021	Transcriptomic profiling of atrazine phytotoxicity and comparativ…	Qu; Mei; Liu; Zhao; Liu; Li; Huang; Zhu	Environmental research		
☐	19	3	22	23102724	2012	Fate of atrazine in switchgrass-soil column system.	Albright; Murphy; Anderson; Coats	Chemosphere		
☐	15	7	22	11476505	2001	Anaerobic degradation of atrazine and metolachlor and metaboli…	Seybold; Mersie; McNamee	Journal of environmental quality		
☐	16	5	21	20830925	2010	Enhanced degradation and soil depth effects on the fate of atrazi…	Krutz; Shaner; Zablotowicz	Journal of environmental quality		
☐	16	5	21	14674556	2003	Infiltration and adsorption of dissolved atrazine and atrazine met…	Krutz; Senseman; Dozier; Hoffman; Tierney	Journal of environmental quality		
☐	19	2	21	2761262	1989	Testosterone metabolism in neuroendocrine organs in male rats …	Babić-Gojmerac; Kniewald; Kniewald	Journal of steroid biochemistry		
☐	20	0	20	33254405	2020	Removal of atrazine in catalytic degradation solutions by microal…	Hu; Xu; Sun; Zhu; Sun; Zhao; Hu	Ecotoxicology and environmental safety		
☐	17	3	20	24062064	2013	Biodegradation of atrazine by Rhodococcus sp. BCH2 to N-isopr…	Kolekar; Phugare; Jadhav	Environmental science and pollution research intern…		
☐	18	2	20	21121649	2010	Metabolism and persistence of atrazine in several field soils with…	Jablonowski; Hamacher; Martinazzo; Langen; Köpp…	Journal of agricultural and food chemistry		
☐	20	0	20	18848368	2008	Nitrogen limited biobarriers remove atrazine from contaminated …	Hunter; Shaner	Journal of contaminant hydrology		
☐	20	0	20	16595379	2006	Mixed-effect models for evaluating multiple measures of atrazine…	Hines; Deddens; Lu; Fenske; Striley	Journal of occupational and environmental hygiene		
☐	12	8	20	16349478	2005	Molecular basis of a bacterial consortium: interspecies catabolis…	de Souza; Newcombe; Alvey; Crowley; Hay; Sadow…	Applied and environmental microbiology		
☐	11	8	19	16656991	1968	Atrazine metabolism in resistant corn and sorghum	Shimabukuro	Plant physiology		

Markup of abstracts



To find articles quickly, enter terms to sift abstracts.

Atrazine | metabol | Clear Terms | Download / Send to... | Download Sifter for Excel

	Atrazine	metabol	Total	PMID	Year	Title	Authors	Journal	Rev	Rev
	19	10	29	16527233	2006	Determination of atrazine and its metabolites in mouse urine and…	Ross; Filipov	Analytical biochemistry		
	13	10	23	16656648	1967	Atrazine metabolism and herbicidal selectivity.	Shimabukuro	Plant physiology		
	13	10	23	8481106	1993	In vitro percutaneous absorption and metabolism in man of 2-chl…	Ademola; Sedik; Wester; Maibach	Archives of toxicology		
	19	3	22	33421427	2021	Transcriptomic profiling of atrazine phytotoxicity and comparativ…	Qu; Mei; Liu; Zhao; Liu; Li; Huang; Zhu	Environmental research		
	19	3	22	23102724	2012	Fate of atrazine in switchgrass-soil column system.	Albright; Murphy; Anderson; Coats	Chemosphere		
	15	7	22	11476505	2001	Anaerobic degradation of atrazine and metolachlor and metaboli…	Seybold; Mersie; McNamee	Journal of environmental quality		
	16	5	21	20830925	2010	Enhanced degradation and soil depth effects on the fate of atrazi…	Krutz; Shaner; Zablotowicz	Journal of environmental quality		
	16	5	21	14674556	2003	Infiltration and adsorption of dissolved atrazine and atrazine met…	Krutz; Senseman; Dozier; Hoffman; Tierney	Journal of environmental quality		
	19	2	21	2761262	1989	Testosterone metabolism in neuroendocrine organs in male rats …	Babić-Gojmerac; Kniewald; Kniewald	Journal of steroid biochemistry		
	20	0	20	33254405	2020	Removal of atrazine in catalytic degradation solutions by microal…	Hu; Xu; Sun; Zhu; Sun; Zhao; Hu	Ecotoxicology and environmental safety		
	17	3	20	24062064	2013	Biodegradation of atrazine by Rhodococcus sp. BCH2 to N-isopr…	Kolekar; Phugare; Jadhav	Environmental science and pollution research intern…		
	18	2	20	21121649	2010	Metabolism and persistence of atrazine in several field soils with…	Jablonowski; Hamacher; Martinazzo; Langen; Köpp…	Journal of agricultural and food chemistry		
	20	0	20	18848368	2008	Nitrogen limited biobarriers remove atrazine from contaminated …	Hunter; Shaner	Journal of contaminant hydrology		
	20	0	20	16595379	2006	Mixed-effect models for evaluating multiple measures of atrazine…	Hines; Deddens; Lu; Fenske; Striley	Journal of occupational and environmental hygiene		
	12	8	20	16349478	2005	Molecular basis of a bacterial consortium: interspecies catabolis…	de Souza; Newcombe; Alvey; Crowley; Hay; Sadow…	Applied and environmental microbiology		
	11	8	19	16656991	1968	Atrazine metabolism in resistant corn and sorghum	Shimabukuro	Plant physiology		

Determination of atrazine and its metabolites in mouse urine and plasma by LC-MS analysis.

Atrazine is a herbicide widely used on agricultural commodities. Existing analytical methods to analyze atrazine and its metabolites in biological matrices have various drawbacks. Thus, further development of such methods will be needed to correlate the growing number of toxicological effects associated with atrazine exposure with the concentrations of this compound and its metabolites in plasma, urine, and tissues. The purpose of this study was to develop a broad and sensitive LC-MS method for the analysis of atrazine and its metabolites in mouse urine and plasma. We were able to simultaneously measure atrazine and its major mammalian metabolites, which include didealkyl atrazine, desisopropyl atrazine, desethyl atrazine, atrazine-glutathione conjugate, and atrazine-mercapturate, using preparation procedures that used small sample volumes of plasma and urine (0.25 and 0.5 ml, respectively). Furthermore, derivatization of analytes prior to analysis was unnecessary. This method was used to analyze plasma and urine samples following single in vivo oral exposures of a limited number of mice to atrazine (doses, 5-250 mg/kg body weight) to demonstrate the utility of this LC-MS method. The data obtained from this study suggest that atrazine is rapidly metabolized in mice. Didealkyl atrazine was the most abundant metabolite detected in the urine and plasma samples (approximately 1000 microM in 24-h urine and approximately 100 microM in plasma following the highest dose of atrazine), with lesser quantities of mono N-dealkylated metabolites and thio conjugates of atrazine observed. We also used this methodology in a preliminary study of cytochrome P450-catalyzed metabolism of atrazine in vitro. The results obtained in this study suggest that this method will be a useful tool for the determination of atrazine and its metabolites in future pharmacokinetic studies and for the subsequent development and refinement of biologically based models of atrazine disposition.

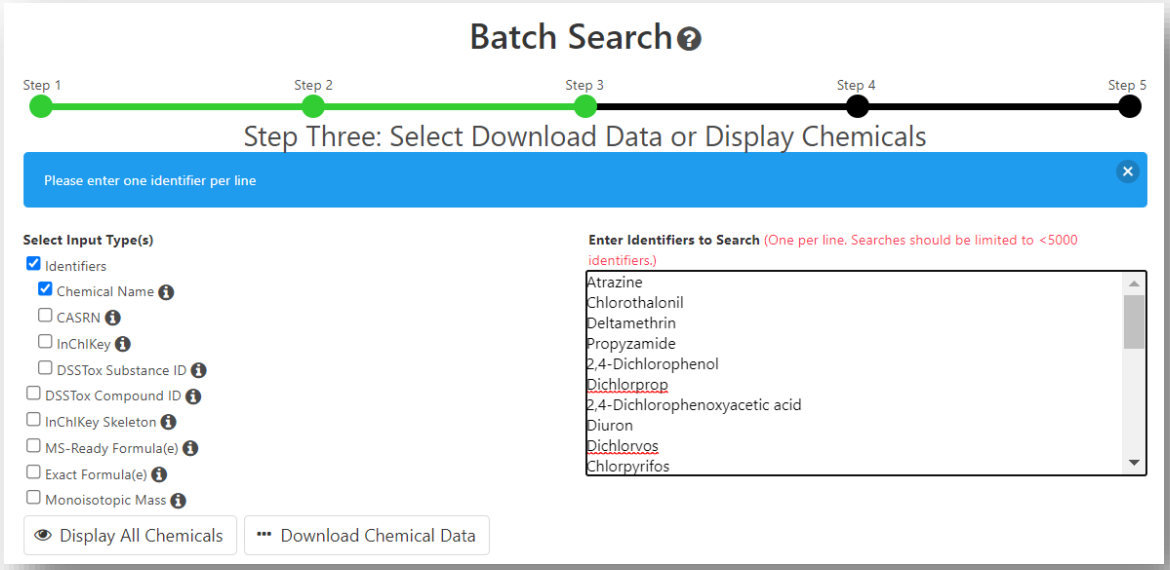
Batch Searching

Batch Search Names

*Untitled - Notepad

File Edit Format View Help

```
Atrazine
Chlorothalonil
Deltamethrin
Propyzamide
2,4-Dichlorophenol
Dichlorprop
2,4-Dichlorophenoxyacetic acid
Diuron
Dichlorvos
Chlorpyrifos
Dimethoate
Endosulfan
Ethylene thiourea
Ethylene urea
(R,R)-Fenvalerate
Picloram
Saccharin
Simazine
Thiram
Thiabendazole
Folpet
Trifluralin
Phthalic acid
3,4-Dichloroaniline
4-Chlorophenol
2,6-Dimethyl morpholine
4-Fluoroaniline
2,6-Dichlorobenzamide
```



Excel Download

	A	B	C	D
1	INPUT	DTXSID	PREFERRED_NAME	CASRN
2	Atrazine	DTXSID9020112	Atrazine	1912-24-9
3	Chlorothalonil	DTXSID0020319	Chlorothalonil	1897-45-6
4	Deltamethrin	DTXSID8020381	Deltamethrin	52918-63-5
5	Propyzamide	DTXSID2020420	Propyzamide	23950-58-5
6	2,4-Dichlorophenol	DTXSID1020439	2,4-Dichlorophenol	120-83-2
7	Dichlorprop	DTXSID0020440	Dichlorprop	120-36-5
8	2,4-Dichlorophenoxy	DTXSID0020442	2,4-Dichlorophenoxyac	94-75-7
9	Diuron	DTXSID0020446	Diuron	330-54-1
10	Dichlorvos	DTXSID5020449	Dichlorvos	62-73-7
11	Chlorpyrifos	DTXSID4020458	Chlorpyrifos	2921-88-2
12	Dimethoate	DTXSID7020479	Dimethoate	60-51-5
13	Endosulfan	DTXSID1020560	Endosulfan	115-29-7
14	Ethylene thiourea	DTXSID5020601	Ethylene thiourea	96-45-7
15	Ethylene urea	DTXSID0020602	Ethylene urea	120-93-4
16	(R,R)-Fenvalerate	DTXSID3020621	(R,R)-Fenvalerate	67614-33-9

Add Other Data of Interest

EPA United States Environmental Protection Agency

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

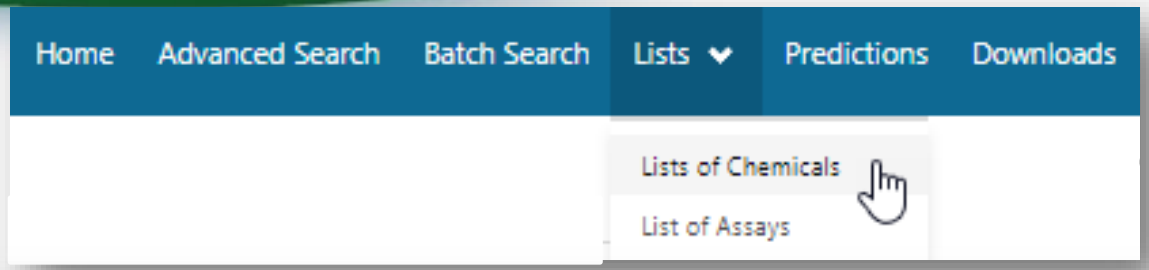
Intrinsic And Predicted Properties

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

B	C	D	E	F	G	H	I	J	K
DTXSID	PREFERRED_NAI	CASRN	MOLECULAR_FO	MONOISOTOF	TOXVAL_	TOXCAST_	TOXCAST_	IRIS_LINK	WIKIPEDIA
DTXSID9020112	Atrazine	1912-24-9	C8H14ClN5	215.0937732	Y	6.05	62/1024	Y	Y
DTXSID0020319	Chlorothalonil	1897-45-6	C8Cl4N2	263.8815588	Y	36.36	389/1070	Y	Y
DTXSID8020381	Deltamethrin	52918-63-5	C22H19Br2NO3	502.973169	Y	12.92	80/619	-	Y
DTXSID2020420	Propyzamide	23950-58-5	C12H11Cl2NO	255.0217694	Y	10.47	102/974	Y	-
DTXSID1020439	2,4-Dichlorophenol	120-83-2	C6H4Cl2O	161.9639201	Y	3.95	35/886	Y	Y
DTXSID0020440	Dichlorprop	120-36-5	C9H8Cl2O3	233.9850495	Y	2.39	22/921	-	Y
DTXSID0020442	2,4-Dichlorophenoxy	94-75-7	C8H6Cl2O3	219.9693995	Y	4.1	40/975	Y	Y
DTXSID0020446	Diuron	330-54-1	C9H10Cl2N2O	232.0170184	Y	10.45	104/995	Y	Y
DTXSID5020449	Dichlorvos	62-73-7	C4H7Cl2O4P	219.9459011	Y	6.03	57/946	Y	Y
DTXSID4020458	Chlorpyrifos	2921-88-2	C9H11Cl3NO3PS	348.9262845	Y	17.56	150/854	Y	Y
DTXSID7020479	Dimethoate	60-51-5	C5H12NO3PS2	228.9996226	Y	3.13	31/989	-	Y
DTXSID1020560	Endosulfan	115-29-7	C9H6Cl6O3S	403.8168814	Y	32.63	311/953	Y	Y
DTXSID5020601	Ethylene thiourea	96-45-7	C3H6N2S	102.0251694	Y	3.33	36/1082	Y	Y
DTXSID0020602	Ethylene urea	120-93-4	C3H6N2O	86.04801282	Y	0.43	1/235	-	-
DTXSID3020621	(R,R)-Fenvalerate	67614-33-9	C25H22ClNO3	419.1288213	Y	11.79	54/458	Y	Y
DTXSID1021160	Picloram	1918-02-1	C6H3Cl3N2O2	239.9260104	Y	1.46	14/958	Y	Y
DTXSID5021251	Saccharin	81-07-2	C7H5NO3S	182.9990142	Y	1.13	5/442	-	Y
DTXSID4021268	Simazine	122-34-9	C7H12ClN5	201.0781231	Y	3.99	40/1002	Y	Y
DTXSID5021332	Thiram	137-26-8	C6H12N2S4	239.9883331	Y	28.18	290/1029	-	Y
DTXSID0021337	Thiabendazole	148-79-8	C10H7N3S	201.0360684	Y	8.17	81/991	-	Y
DTXSID0021385	Folpet	133-07-3	C9H4Cl3NO2S	294.9028326	Y	20.64	206/998	-	-
DTXSID4021395	Trifluralin	1582-09-8	C13H16F3N3O4	335.1092905	Y	9.87	96/973	Y	Y
DTXSID8021484	Phthalic acid	88-99-3	C8H6O4	166.0266087	Y	0.0	0/235	-	Y

Chemical Lists of Interest…

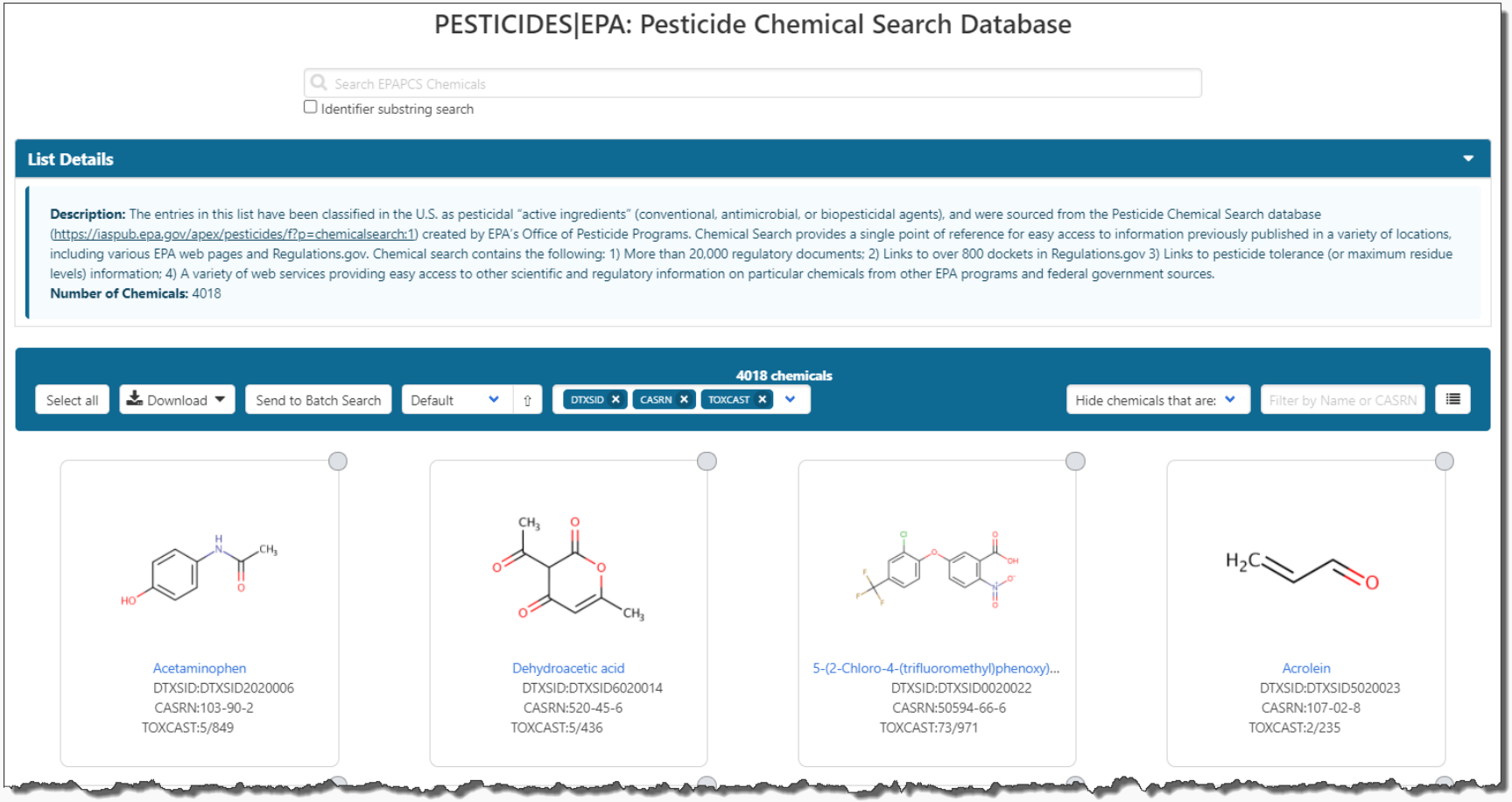
~300 Chemical Lists (and growing)



Download | Columns | pesticides | Copy Filtered Lists URL

List Acronym	List Name	Last Updated	Number of Chemicals	List Description
INERTNONFOOD	CATEGORY|PESTICIDES; InertFinder Non-Food Chemicals	2020-11-17	4814	List of chemicals listed in InertFinder as Non-Food Use Chemicals
EPAPCS	PESTICIDES|EPA: Pesticide Chemical Search Database	2017-11-07	4018	The entries in this list have been classified in the U.S. as pesticidal "active ingredients" (conventional, antimicrobial, or biopesticidal agents), and were sourced from the Pesticide Chemical Search database.
OPPIN	CATEGORY|EPA|PESTICIDES: Office of Pesticide Programs Information Network	2019-10-30	3998	Office of Pesticide Programs integrated system, the Office of Pesticide Programs Information Network (OPPIN).
PESTINERTS	PESTICIDES|EPA: List of Inert Ingredients Food and Nonfood Use UPDATED 10/25/2019	2019-11-17	1654	List of Inert Ingredients Food and Nonfood Use UPDATED 10/25/2019
SWISSPEST19	PESTICIDES|NORMAN: Swiss Pesticides and Metabolites from Keifer et al 2019	2019-11-16	876	Swiss pesticides contributed to the NORMAN Suspect List Exchange
PESTACTIVES	PESTICIDES|EPA: List of Active Ingredients UPDATED 10/25/2019	2019-11-17	510	List of active ingredients in pesticides UPDATED 10/25/2019

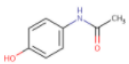
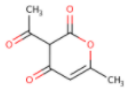
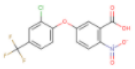
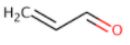
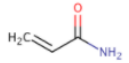
EPA Pesticide Chemical Search
Tile Mode



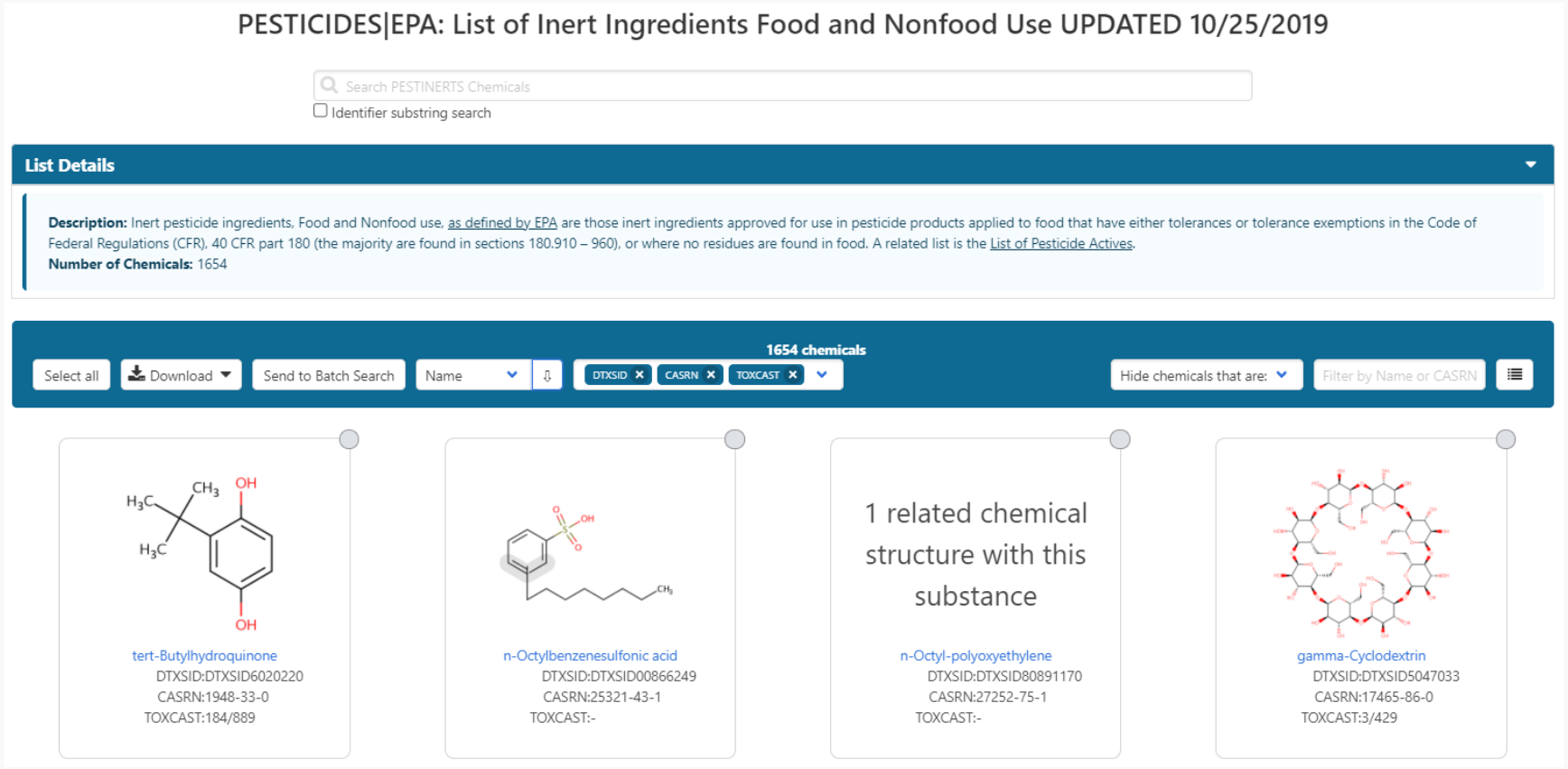
EPA Pesticide Chemical Search
Table Mode

EPA United States Environmental Protection Agency

Select all | Download | Send to Batch Search | Default | 4018 chemicals | Hide chemicals that are: | Filter by Name or CASRN

Structure	DTXSID	Preferred Name	CASRN	QC Level	CPDat Count	Number of Sources	PubChem Data Sources	PubMed Ref. Counts	Monoisotopic Mass
	DTXSID2020006 ToxCast™	Acetaminophen	103-90-2	Level 2	24	199	396	15759	151.063329
	DTXSID6020014 ToxCast™	Dehydroacetic acid	520-45-6	Level 1	15	112	141	71	168.042259
	DTXSID0020022 ToxCast™	5-(2-Chloro-4-(trifluoromethyl)phenoxy)-2-nitrobenzoic acid	50594-66-6	Level 2	23	106	98	50	360.996485
	DTXSID5020023 ToxCast™	Acrolein	107-02-8	Level 2	14	189	139	2468	56.026215
	DTXSID5020027 ToxCast™	Acrylamide	79-06-1	Level 2	180	195	257	2402	71.037114

List of Inert Ingredients Food and Nonfood Use



Natural Product Insecticides

EPA United States Environmental Protection Agency

PESTICIDES|NORMAN: Natural Product Insecticides

Search NPINSECT Chemicals

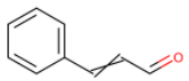
Identifier substring search

List Details

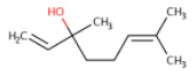
Description: A list of naturally occurring insecticides curated and provided to the NORMAN Suspect List Exchange (https://www.norman-network.com/nds/SLE/) by Reza Aalizadeh (University of Athens). DOI: https://doi.org/10.5281/zenodo.3544742
Number of Chemicals: 84

84 chemicals

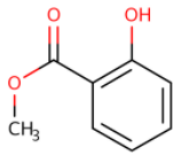
Select all | Download | Send to Batch Search | Default | DTXSID | CASRN | TOXCAST | Hide chemicals that are: | Filter by Name or CASRN



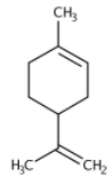
3-Phenylprop-2-enal
DTXSID:DTXSID1024835
CASRN:104-55-2
TOXCAST:1/238



Linalool
DTXSID:DTXSID7025502
CASRN:78-70-6
TOXCAST:1/433

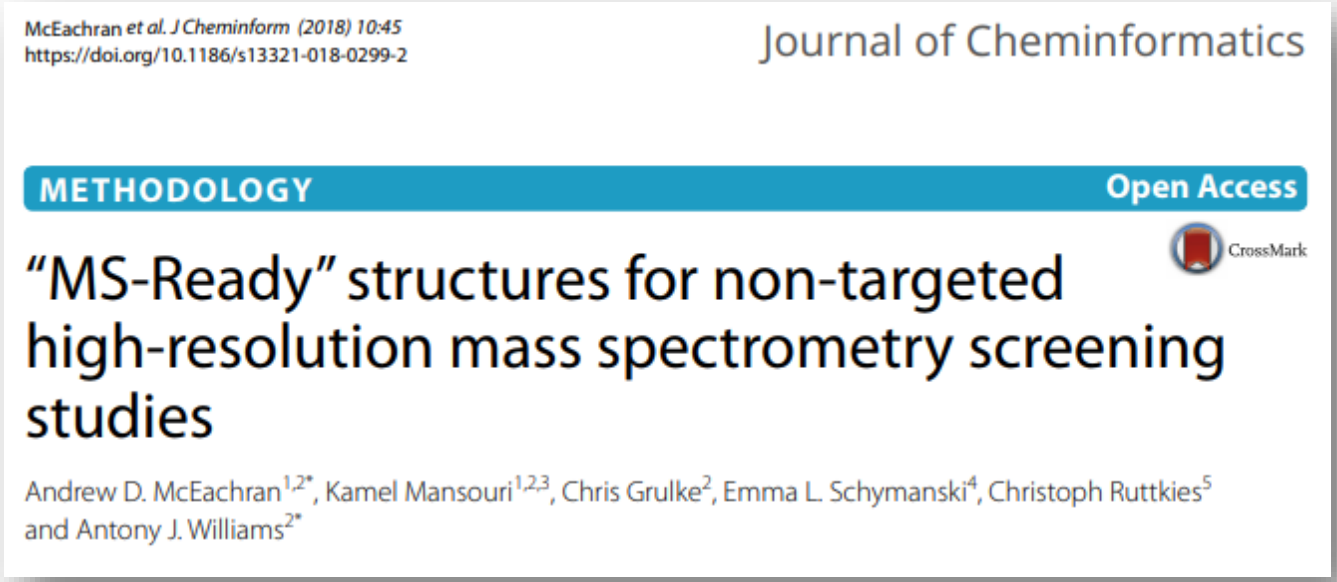


Methyl salicylate
DTXSID:DTXSID5025659
CASRN:119-36-8
TOXCAST:46/855



Limonene
DTXSID:DTXSID2029612
CASRN:138-86-3
TOXCAST:4/433

"MS-ready" structures



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

Journal of Cheminformatics

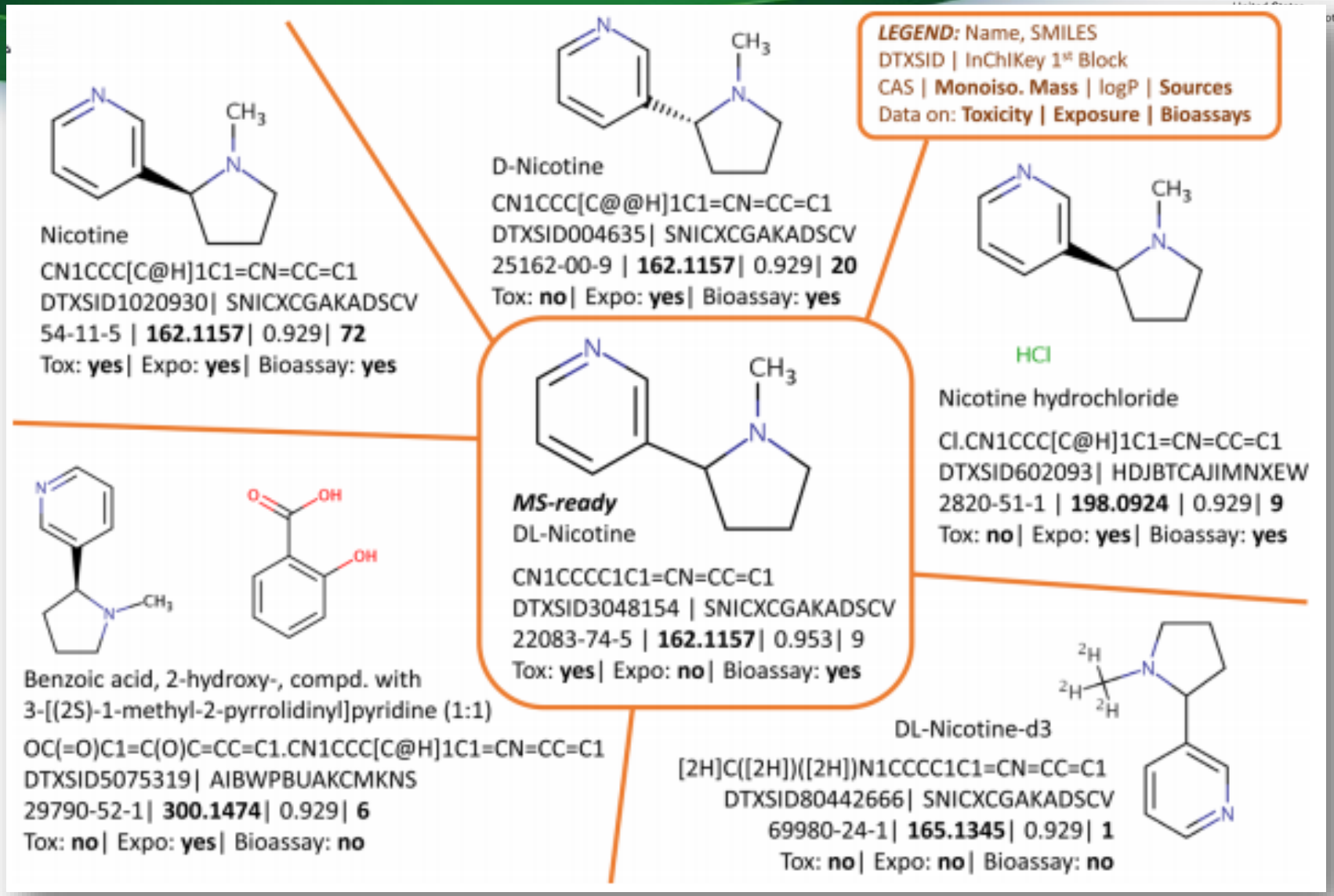
METHODOLOGY Open Access

"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

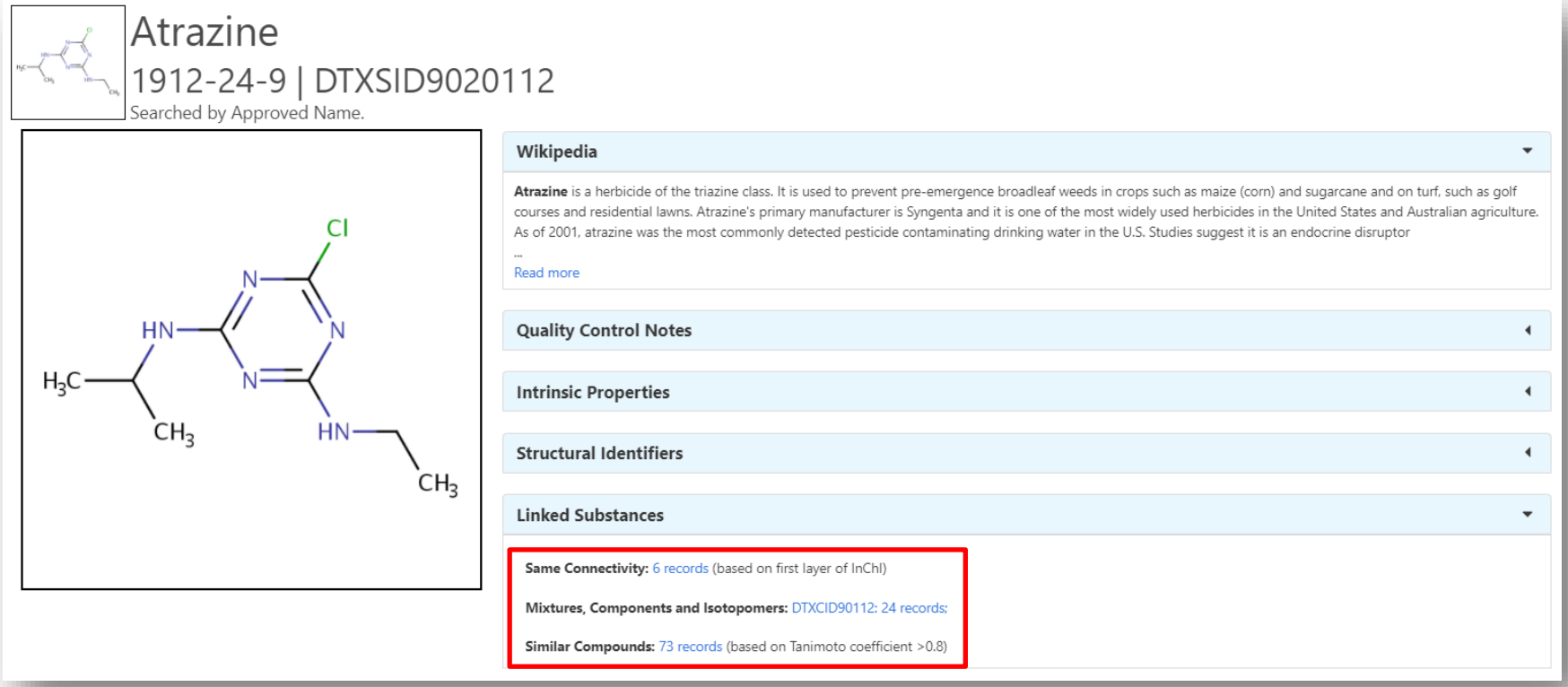


Environmental Science & Technology
Viewpoint
pubs.acs.org/est

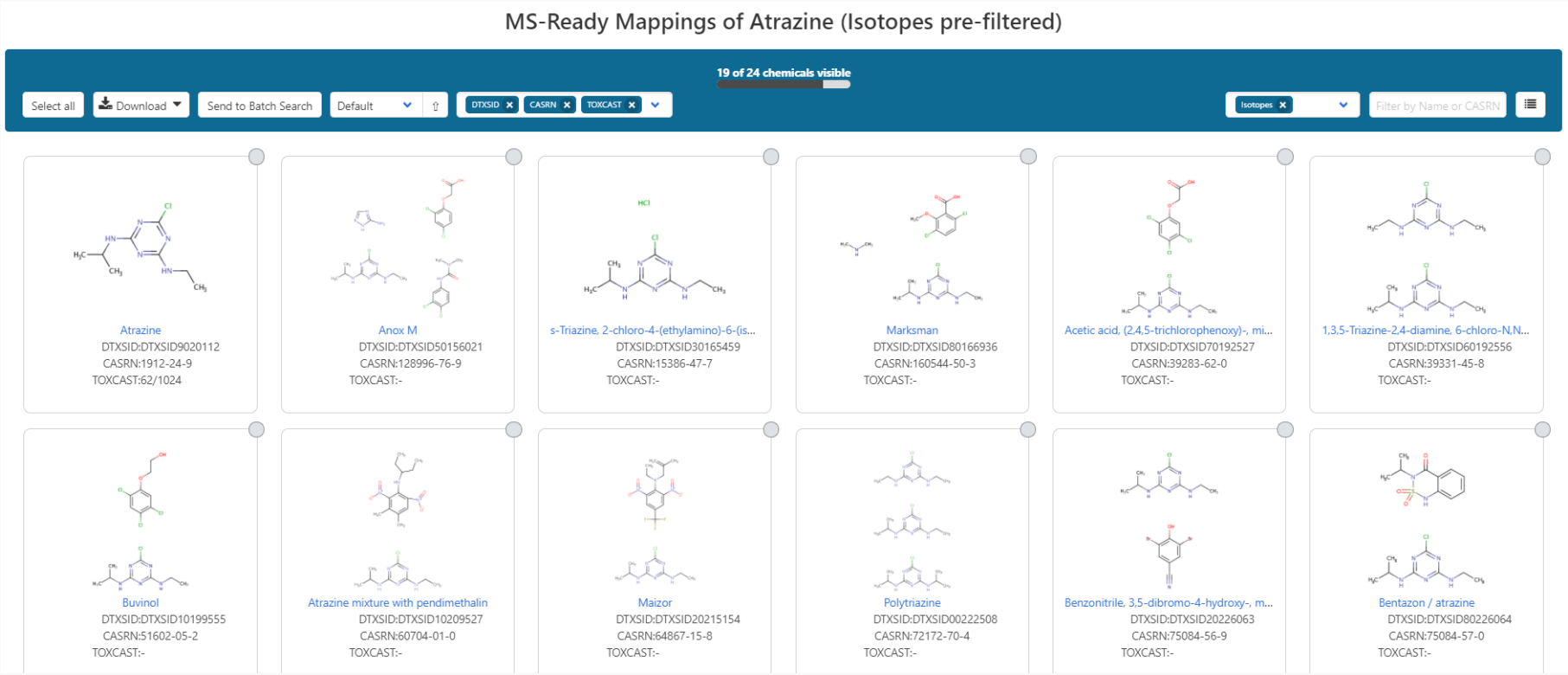
Open Science for Identifying "Known Unknown" Chemicals

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

MS-Ready Mappings from Details Page

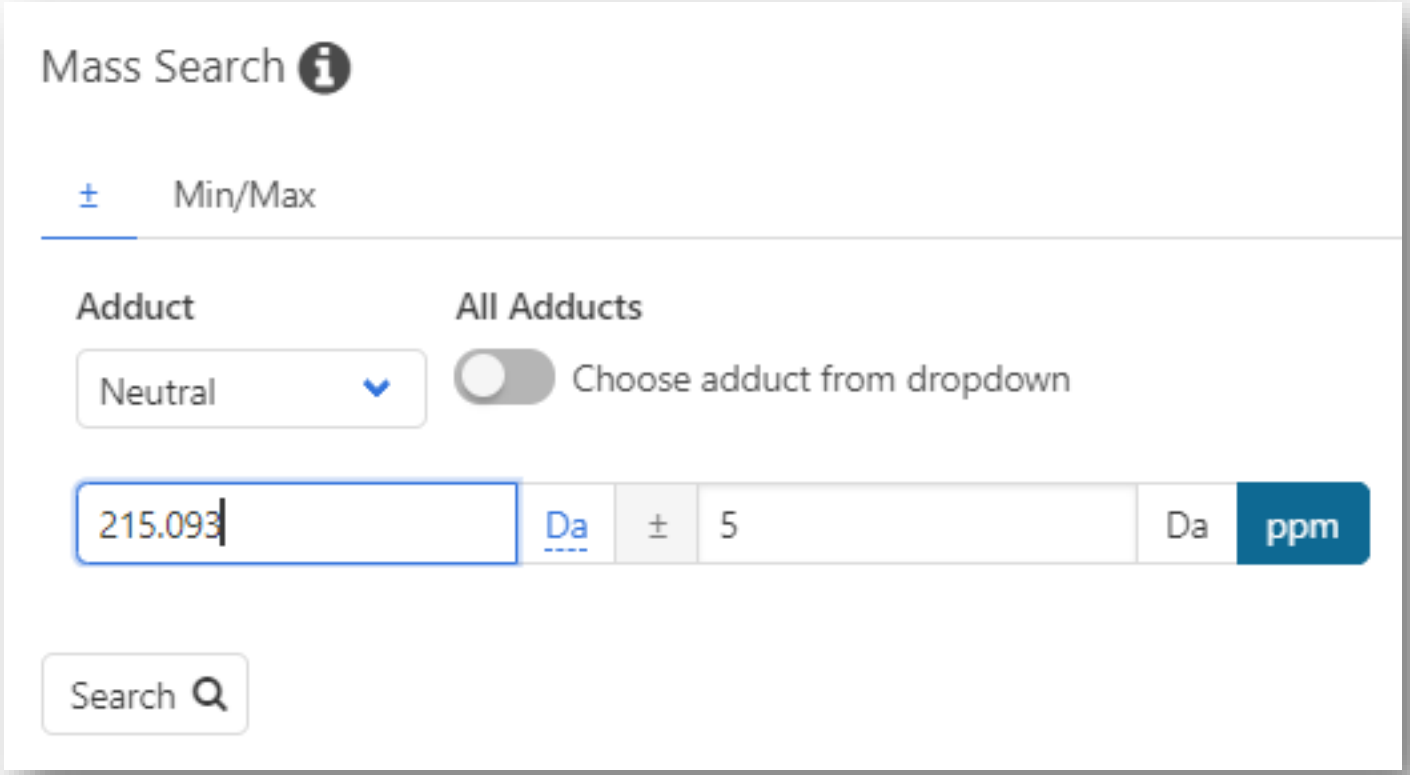


MS-Ready Mappings Set of 24 substances for “Atrazine”

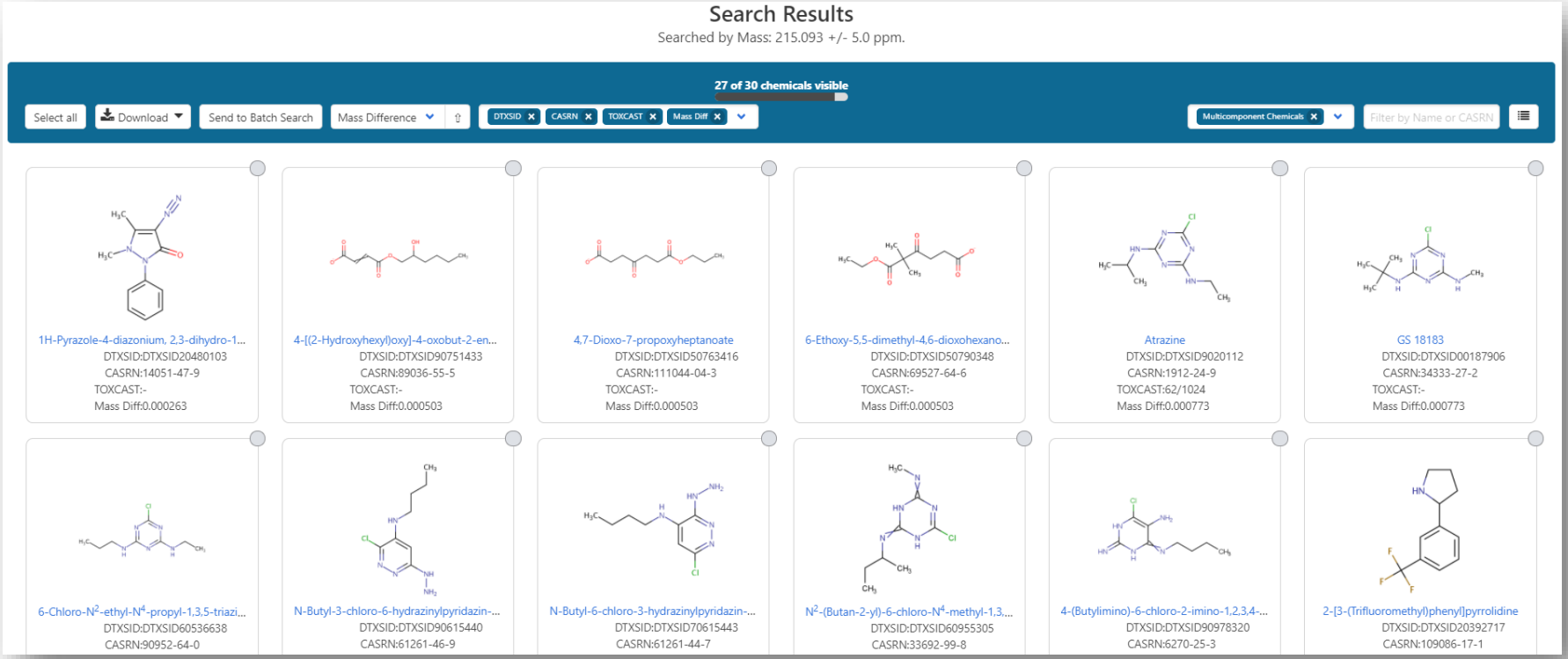


Mass and Formula Searching

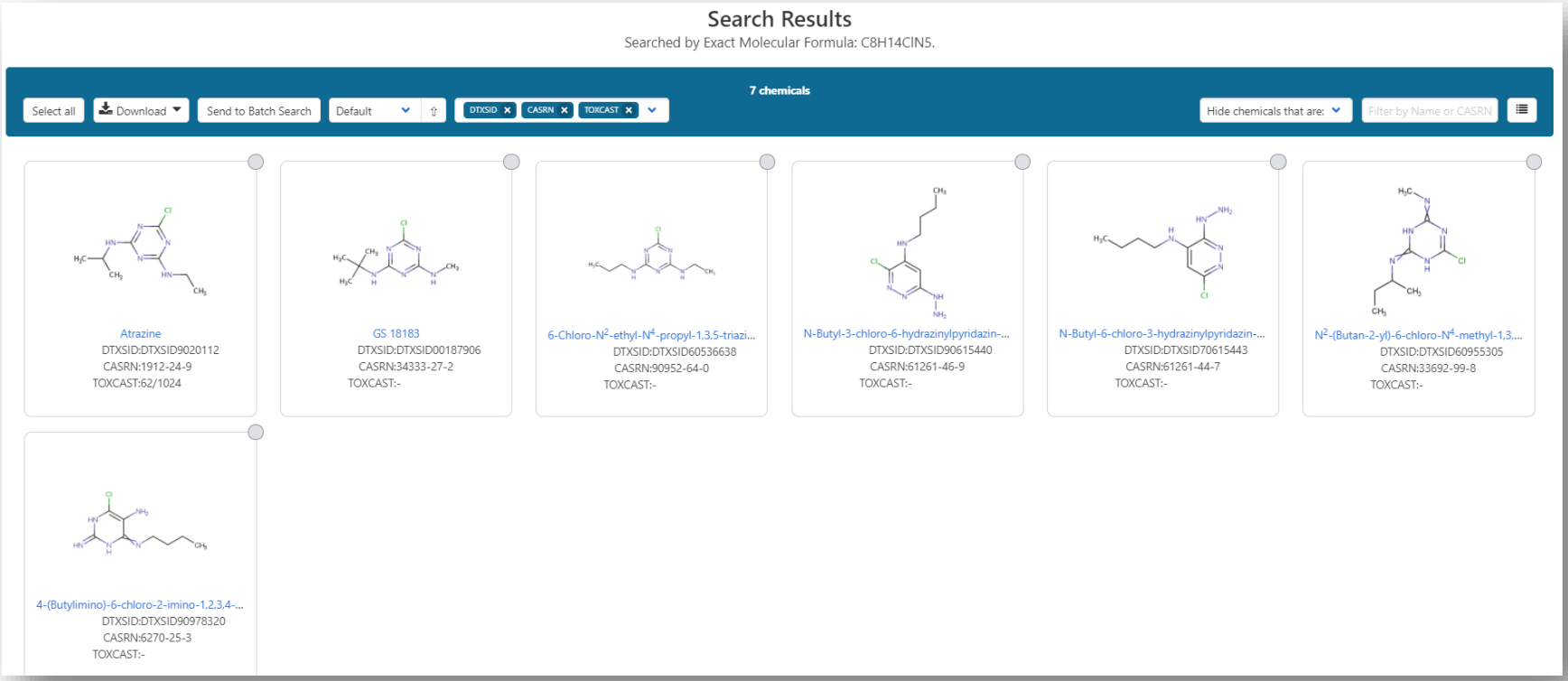
Advanced Searches: **Mass** Search
Same mass as Atrazine +/-5ppm



Advanced Searches: **Mass** Search
Same mass as Atrazine +/-5ppm



Advanced Searches: **Formula** Search
Same Formula as Atrazine



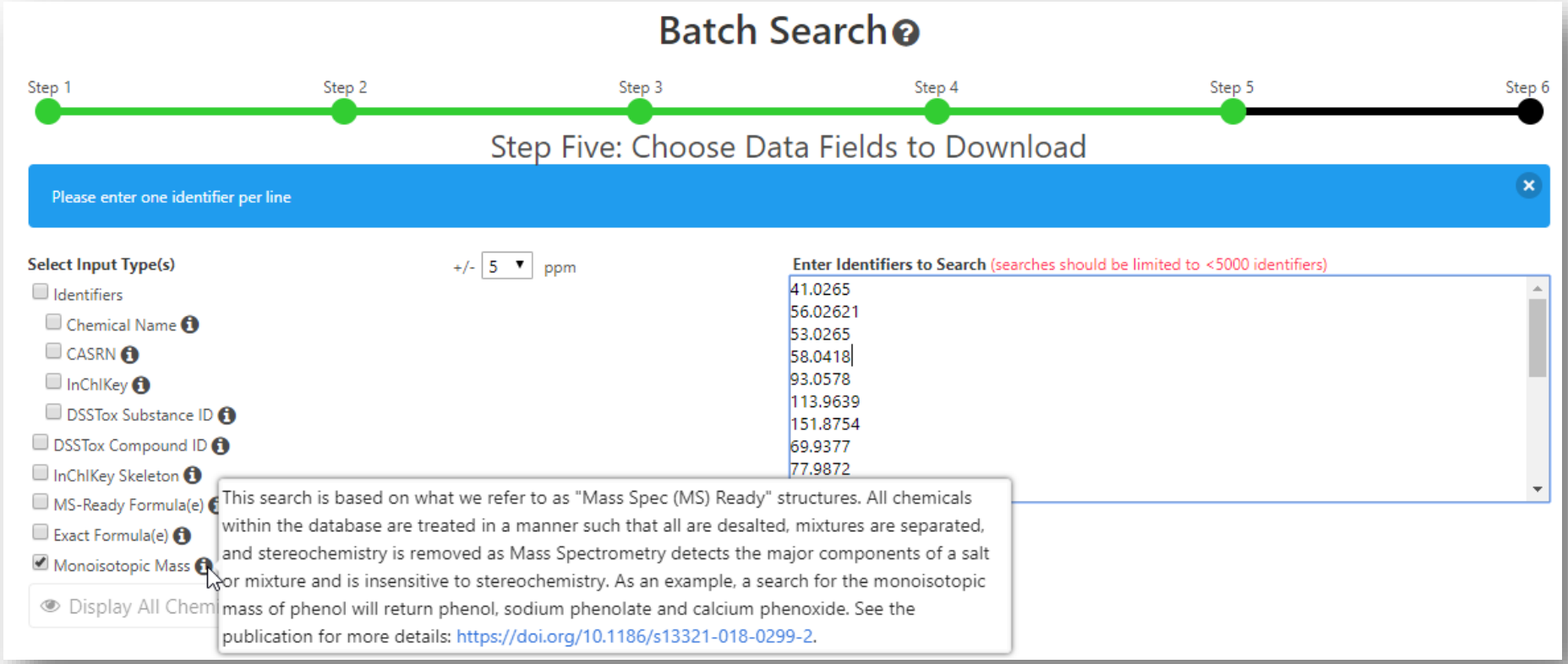
Batch Searching mass and formula

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

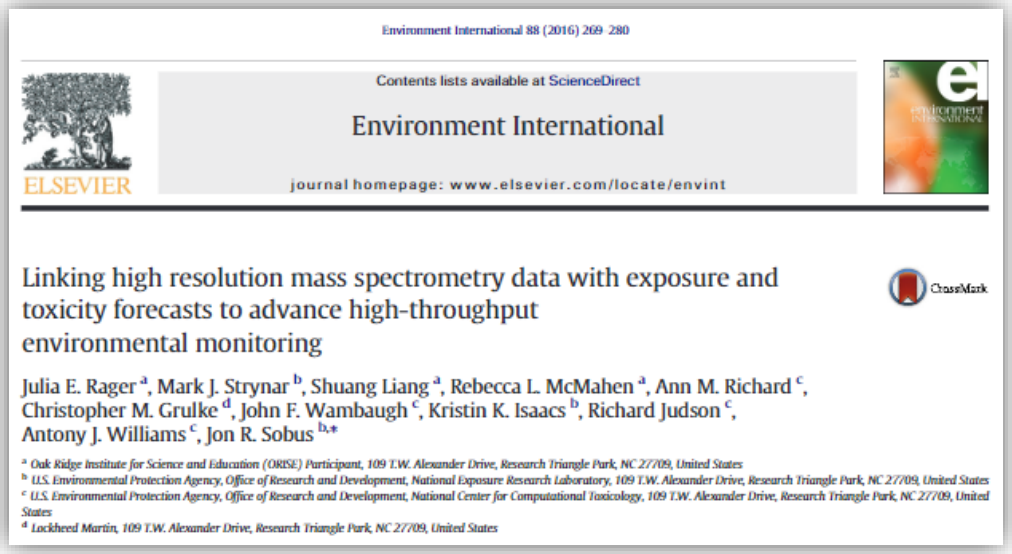
	A	B	C	D	E	F	G
1	INPUT	DTXSID	CASRN	PREFERRED NAME	MOL FORMULA	MONOISOTOPIC MASS	DATA SOURCES
2	C14H22N2O3	DTXSID2022628	29122-68-7	Atenolol	C14H22N2O3	266.163042576	46
3	C14H22N2O3	DTXSID0021179	6673-35-4	Practolol	C14H22N2O3	266.163042576	32
4	C14H22N2O3	DTXSID4048854	841-73-6	Bucolome	C14H22N2O3	266.163042576	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-quina	C14H18N4O3	290.137890456	4
30	C14H18N4O3	DTXSID30213742	63931-79-3	1H-1,2,4-Benzotriazepine-3-carboxylic acid, 4,5-dihydro-4-	C14H18N4O3	290.137890456	3
31	C14H18N4O3	DTXSID30219608	69449-07-6	2,4-Pyrimidinediamine, 5-((3,4,5-trimethoxyphenyl)methyl)-	C14H20N4O4	308.14845514	3
32	C14H18N4O3	DTXSID20241155	94232-27-6	L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl	C18H25N5O7	423.175398165	3
33	C14H18N4O3	DTXSID80241156	94232-28-7	L-Glutamic acid, compound with 5-((3,4,5-trimethoxypheny	C19H27N5O7	437.191048229	3
34	C14H18N4O3	DTXSID20143781	101204-93-7	1H-Pyrido(2,3-e)-1,4-diazepine-2,3,5-trione, 4-(2-(diethylam	C14H18N4O3	290.137890456	3
35	C12H11N7	DTXSID6021373	396-01-0	Triamterene	C12H11N7	253.107593382	52
36	C12H11N7	DTXSID00204465	5587-93-9	Ampyrimine	C12H11N7	253.107593382	7
37	C12H11N7	DTXSID5064621	7300-26-7	Benzenamine, 4-azido-N-(4-azidophenyl)-	C12H9N7	251.091943318	4
38	C12H11N7	DTXSID00848025	90293-82-6	Sulfuric acid--6-phenylpteridine-2,4,7-triamine (1/1)	C12H13N7O4S	351.074973101	1
39	C12H11N7	DTXSID50575293	92310-83-3	(1E)-N-Phenyl-1,2-bis(1H-1,2,4-triazol-1-yl)ethan-1-imine	C12H11N7	253.107593382	1
40	C8H9NO2	DTXSID2020006	103-90-2	Acetaminophen	C8H9NO2	151.063328534	75
41	C8H9NO2	DTXSID6026567	134-20-3	Methyl 2-aminobenzoate	C8H9NO2	151.063328534	60

Batch Search in specific lists

INPUT	DTXSID	MASSBANKREF	NEMILIST	WRTMSD	NORMANPRI	SUSDAT
Buprenorph	DTXSID202	-	-	Y	-	Y
Codeine	DTXSID202	Y	Y	Y	Y	Y
Dextrometh	DTXSID302	Y	Y	Y	-	Y
Dihydrocod	DTXSID502	Y	-	Y	Y	Y
Dihydromor	DTXSID704	-	-	-	-	Y
Ethylmorph	DTXSID104	-	-	Y	-	Y
Fentanyl	DTXSID902	Y	-	Y	-	Y
Heroin	DTXSID604	Y	-	Y	Y	Y
Hydrocodor	DTXSID802	Y	Y	Y	Y	Y
Hydromorpl	DTXSID802	-	-	Y	-	Y
Ketamine	DTXSID802	Y	-	Y	-	Y
Meperidine	DTXSID902	Y	-	Y	-	Y
Methadone	DTXSID702	Y	Y	Y	-	Y
Morphine	DTXSID902	Y	Y	Y	Y	Y
Morphinone	DTXSID501	-	-	-	-	Y
Naloxone	DTXSID802	-	-	Y	-	Y
Naltriben	-	-	-	-	-	-
Oxycodone	DTXSID502	Y	Y	Y	Y	Y
Oxymorpho	DTXSID502	-	-	Y	-	Y
Propoxyphe	DTXSID102	Y	Y	Y	-	Y
Sufentanil	DTXSID602	-	-	Y	-	Y
Tramadol	DTXSID908	Y	Y	Y	Y	Y

Benefits of bringing it all together

- The true dashboard benefit is integration
- Rank potential candidates for toxicity using available data – hazard, exposure, *in vitro*



Environment International 88 (2016) 269–280

Contents lists available at ScienceDirect

Environment International

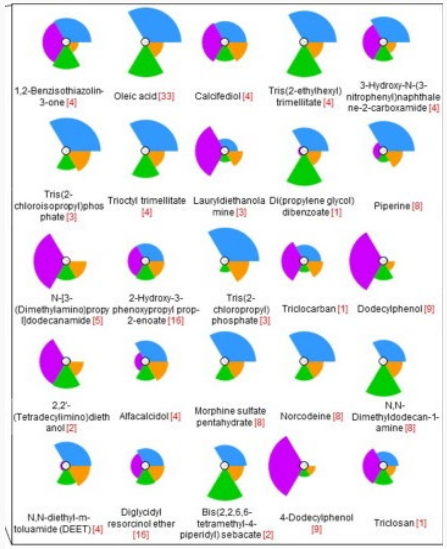
journal homepage: www.elsevier.com/locate/envint

ELSEVIER

Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring

Julia E. Rager [a], Mark J. Strynar [b], Shuang Liang [a], Rebecca L. McMahen [a], Ann M. Richard [c], Christopher M. Grulke [d], John F. Wambaugh [c], Kristin K. Isaacs [b], Richard Judson [c], Antony J. Williams [c], Jon R. Sobus [b,*]

[a] Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[b] U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[c] U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States
[d] Lockheed Martin, 109 T.W. Alexander Drive, Research Triangle Park, NC 27709, United States



EPA United States Environmental Protection Agency

Candidate ranking using metadata

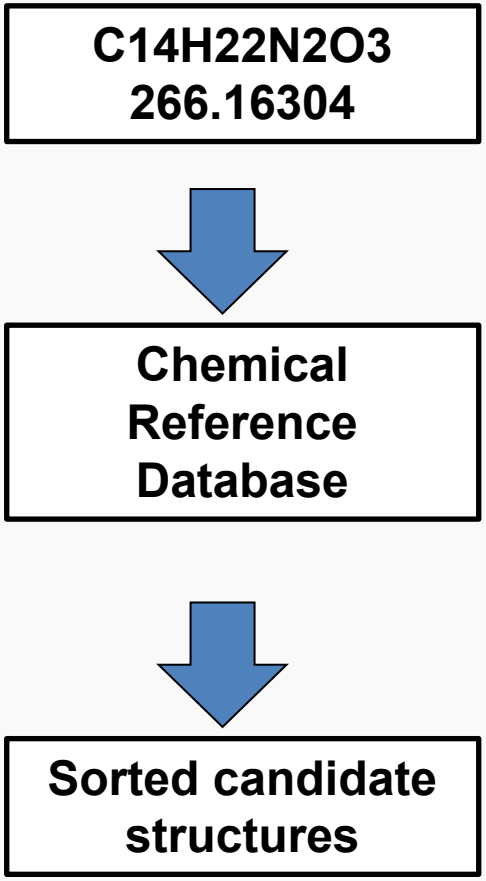
J. Am. Soc. Mass Spectrom. (2012) 23:179–185
DOI: 10.1007/s13361-011-0265-y

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



Data Streams for Ranking

- CompTox Dashboard Data Sources
- PubChem Data Source Count
- PubMed.gov Reference Count
- Toxcast *in* vitro bioactivity
- Presence in CPDat database
- OPERA PhysChem Properties
- Other possibilities – predicted media occurrence, frequency of InChIs online

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?

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?

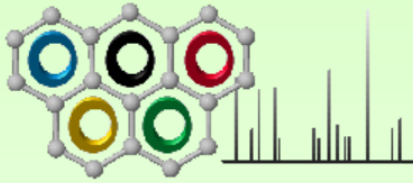
	Mass-based searching		Formula-based searching	
	Dashboard	ChemSpider	Dashboard	ChemSpider
Average rank position	1.3	2.2[a]	1.2	1.4
Percent in #1 position	85%	70%	88%	80%

[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

CASMI

EPA United States Environmental Protection Agency



CASMI 2017

CASMI 2017
- Important Dates
- Contest Rules
- Challenge Data
- Solutions
- Preliminary results
- Results
- About the Team

CASMI 2016

CASMI 2014

CASMI 2013

CASMI 2012

Critical Assessment of Small Molecule Identification

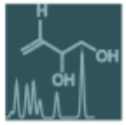
The experimental and computational mass spectrometry communities are invited to participate in the fifth round of an open contest on the identification of small molecules from mass spectrometry data.

This year the contest will test the applicability of MS and MS/MS on natural products chemistry identifications. With 45 (Category 1) and up to 243 (Categories 2&3) natural products challenges - including a few tricky ones - there's something for everyone!

CASMI 2017 is organised by Dr. Dejan Nikolic (University of Illinois at Chicago, USA), Dr. Nir Shahaf (Weizmann Institute of Science, Israel), Dr. Emma Schymanski (Eawag, Switzerland) and Dr. Steffen Neumann (IPB Halle, Germany).

Mailing lists

CASMI



metabolites

MDPI

Article

Revisiting Five Years of CASMI Contests with EPA Identification Tools

Table 2. Percentage of the total number of compounds from each CASMI contest year that were ranked in the top 5 by Competitive Fragmentation Modeling for Metabolite Identification (CFM-ID) only and by the summation of CFM-ID and DSSTox Data Source Counts (DS), alongside the percentage in the top 5 reported by the contest years' winning entry. Complete ranking results are provided in Supplemental File S1.

CASMI Year	CFM-ID Only	CFM-ID + DS	Winners' Results [1]	Total in DB/Total in Dataset [2]
2012	36%	64%	36%	14/14
2013	81%	88%	88%	16/16
2014	57%	76%	71%	42/42
2016-training	63%	96%		312/312
2016-challenge	66%	94%	81%	208/208
2017	59%	53%	74% [3]	227/243

Work in Progress

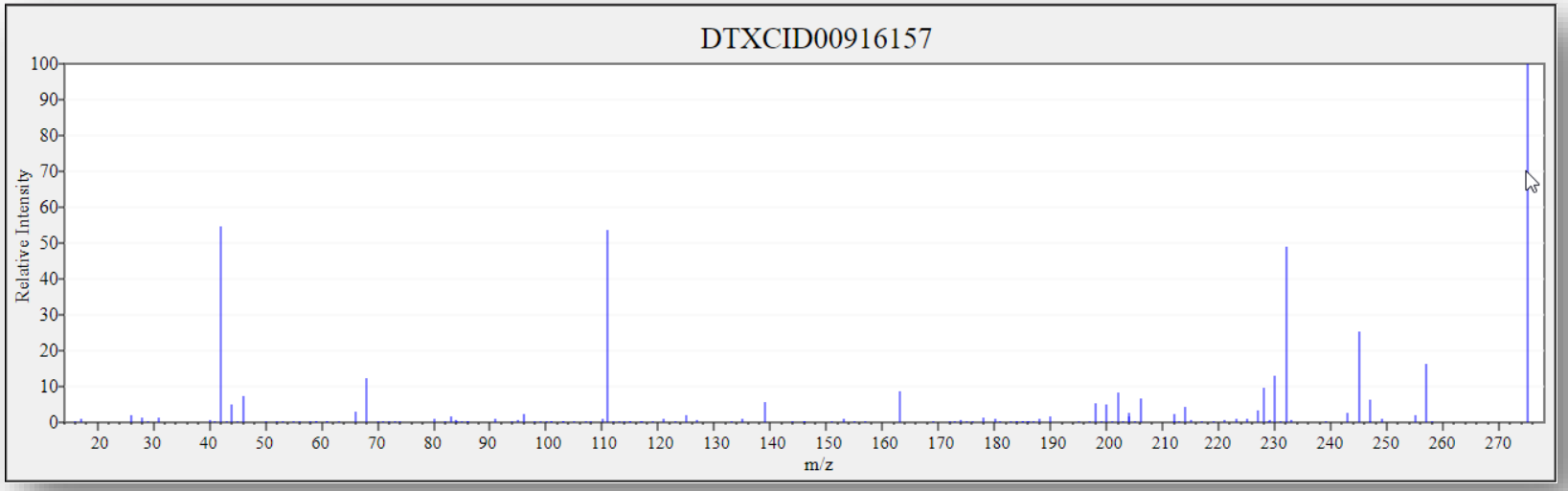
Prototype Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Structure/substructure/similarity search
- Access to API and web services
- Integration to EPA “Chemical Transformation Simulator”

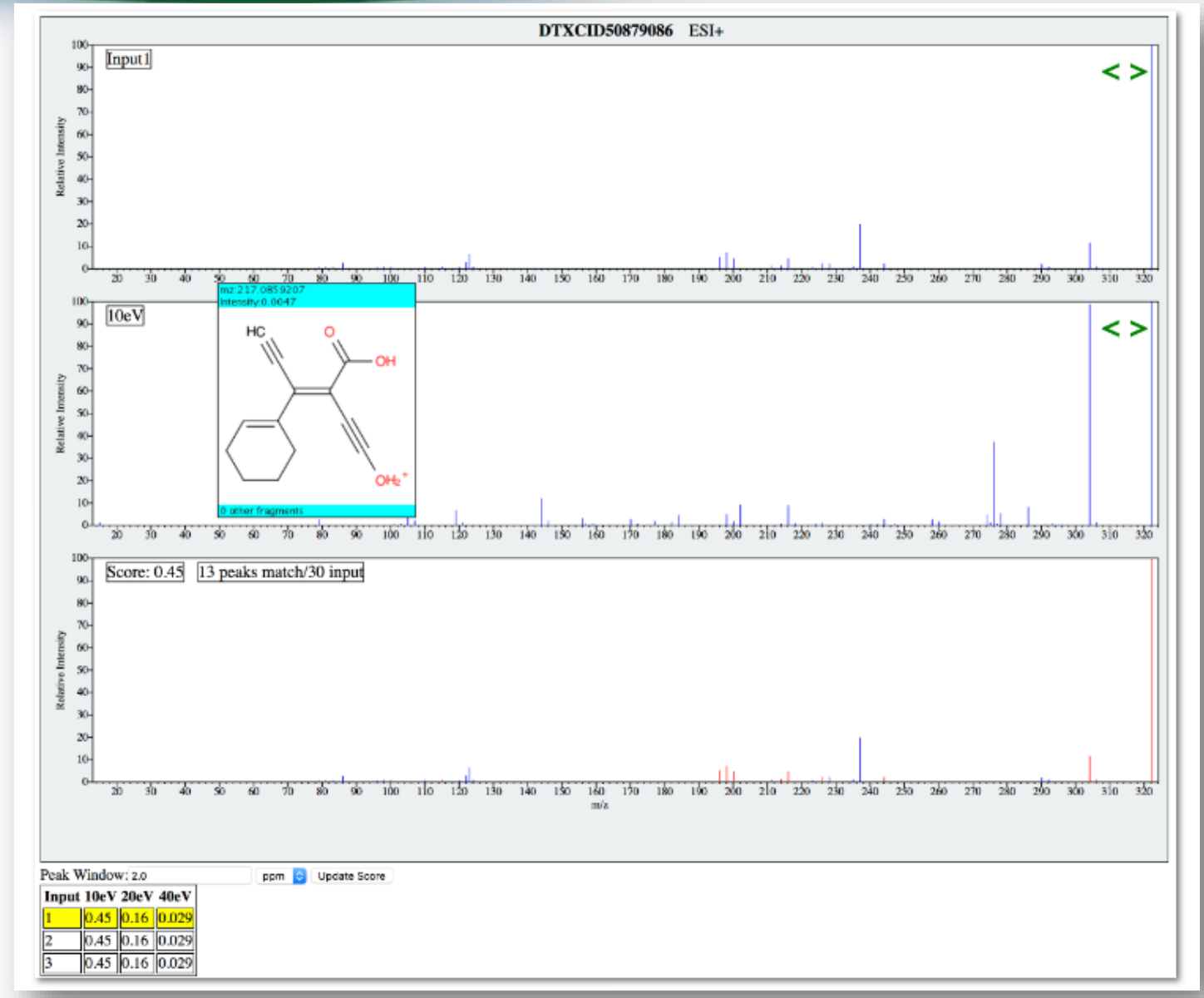
Predicted Mass Spectra

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



Spectral Viewer Comparison



Published: Alex Chao et al

Analytical and Bioanalytical Chemistry

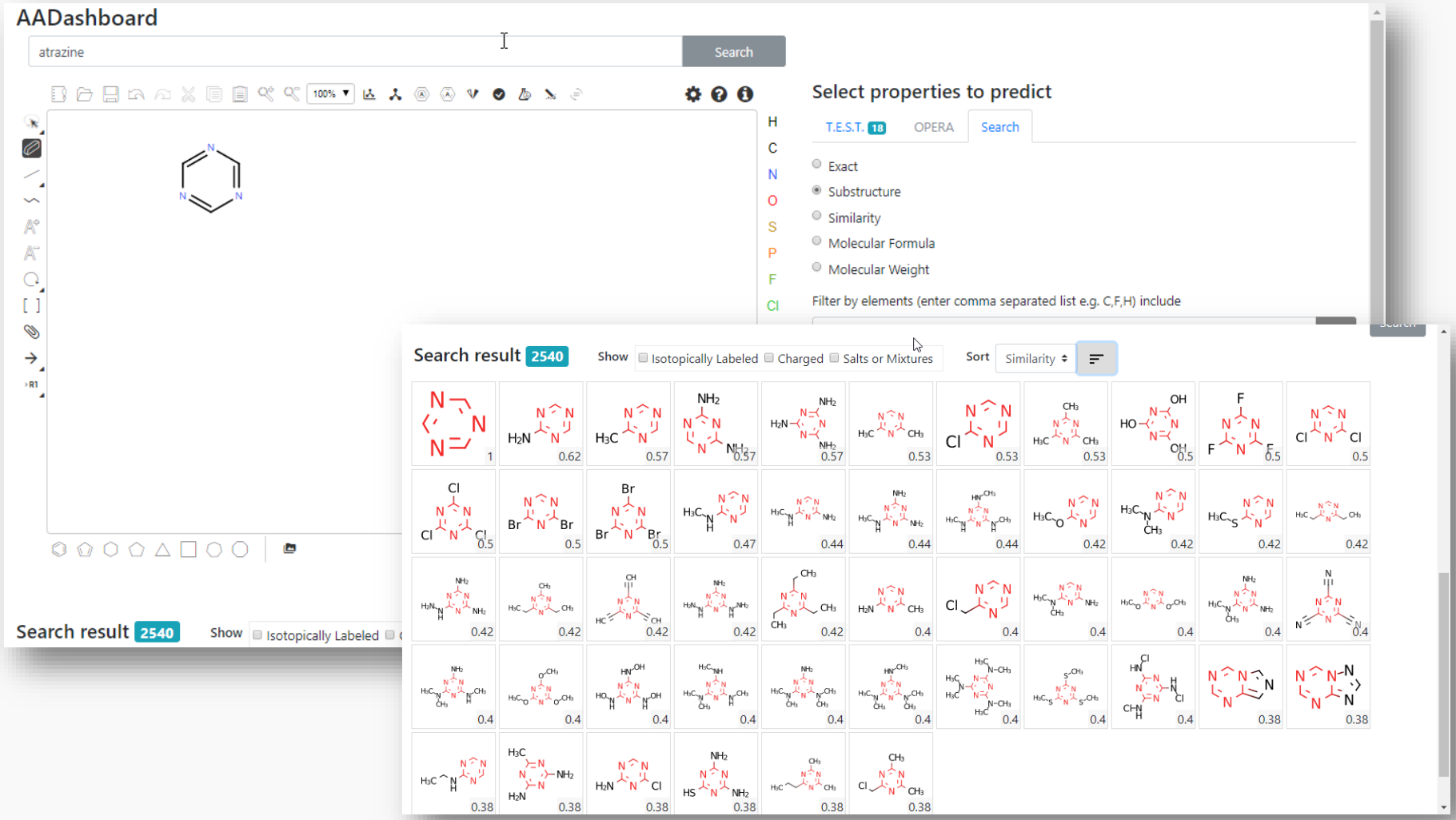
RESEARCH PAPER

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

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

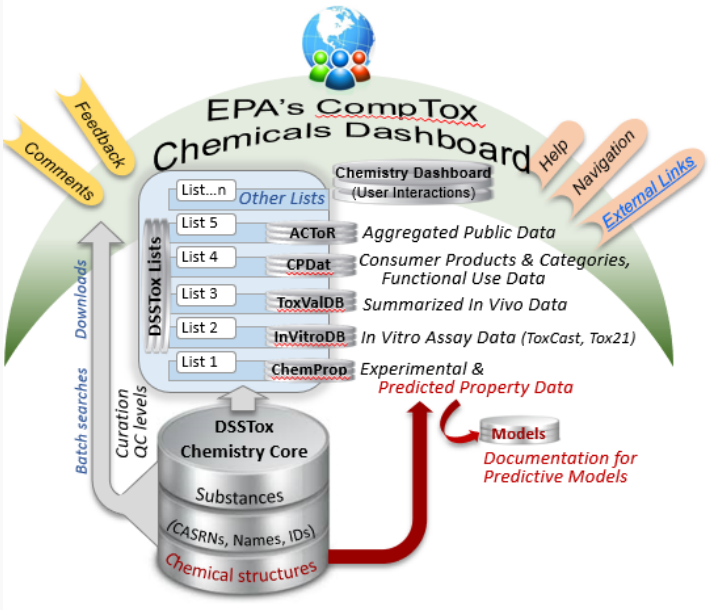
Received: 4 October 2019 / Revised: 27 November 2019 / Accepted: 11 December 2019

Prototype Development



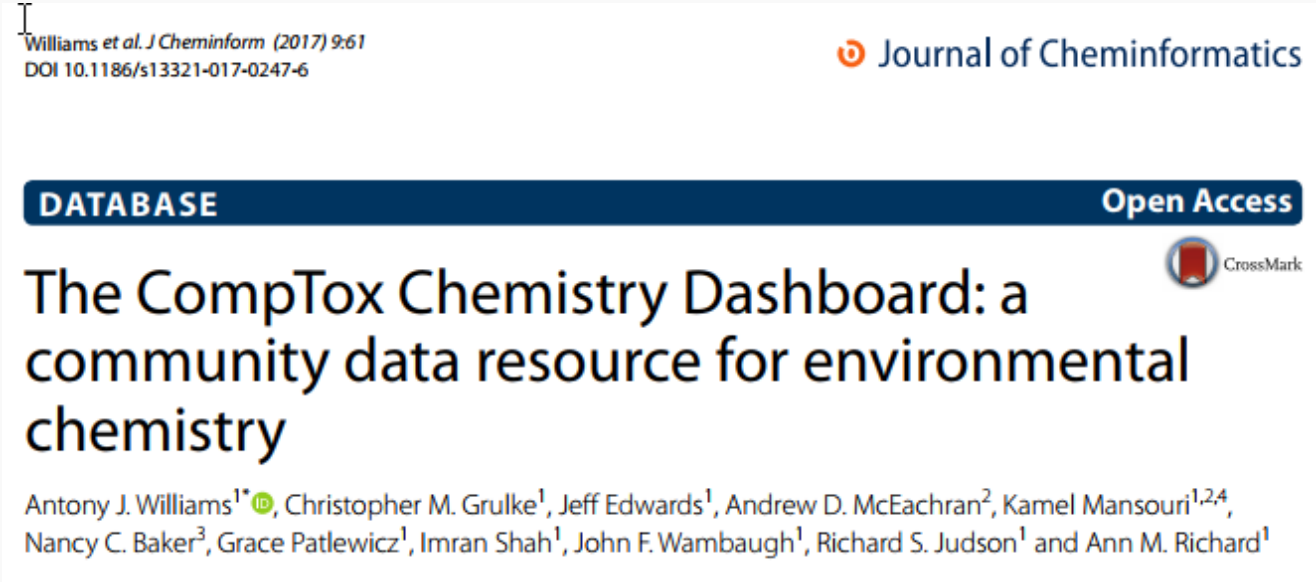
Conclusion

- Dashboard access to data for ~883,000 chemicals
- Pesticides lists sourced from EPA and other sources
- Relationship mappings between parents and products
- MS-Ready data facilitates structure identification
- Metadata for candidate ranking
- Data continues to grow with ongoing curation activities



Contact

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



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