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

# Applications of the US EPA's CompTox chemicals dashboard to support structure identification and chemical forensics using mass spectrometry

***Antony Williams***[1], *Andrew D. McEachran*[2], *Jon R. Sobus*[3] *and Emma Schymanski*[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) *National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC*
4) *Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Campus Belval, Luxembourg*



*Spring 2019*
*ACS Spring Meeting, Orlando*

# Suspect Screening and Non-Targeted Analysis Workflows

EPA United States Environmental Protection Agency

**Suspect Screening**

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

**Non-Targeted Analysis**

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

***Color Key***

Red = Analytical Chemistry

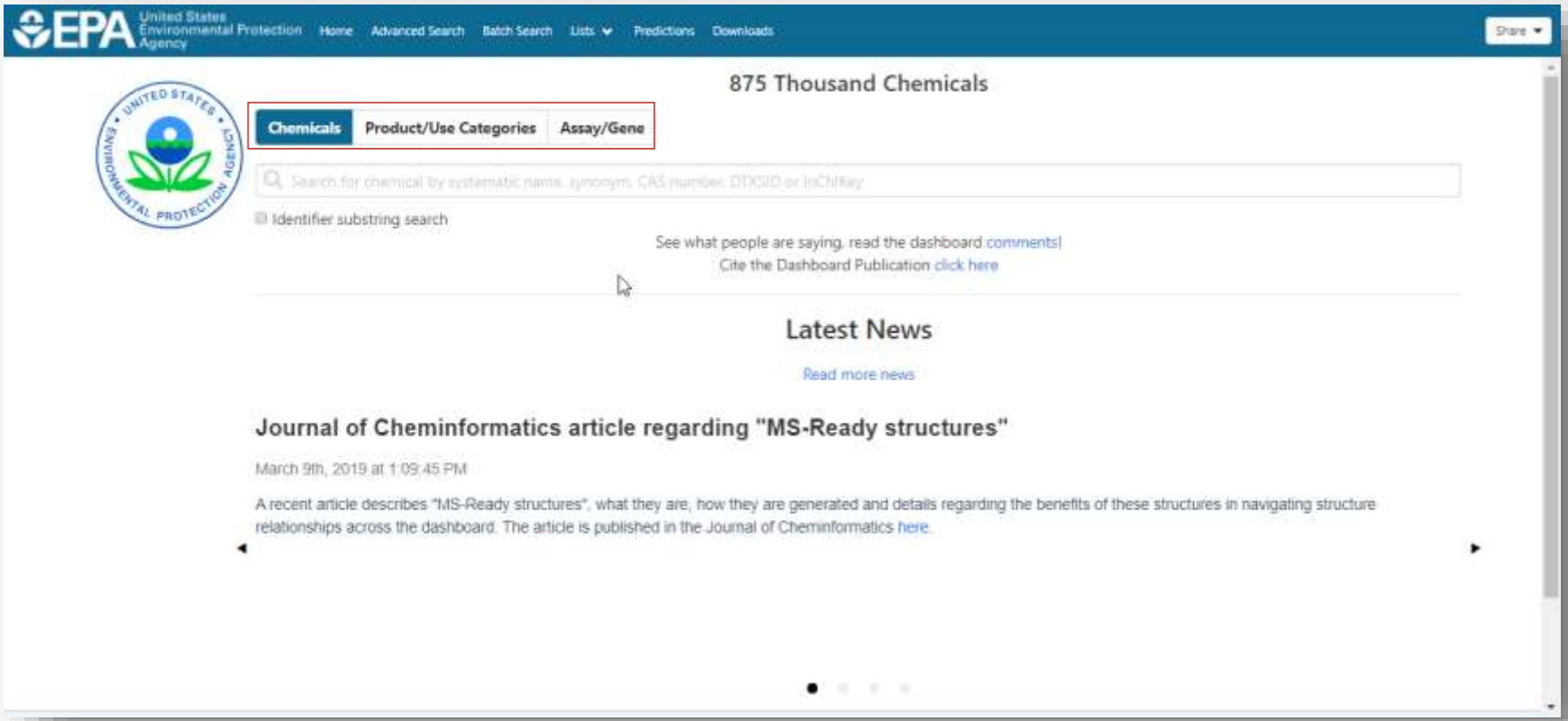
Blue = Data Processing & Analysis

Purple = Mathematical & QSPR Modeling

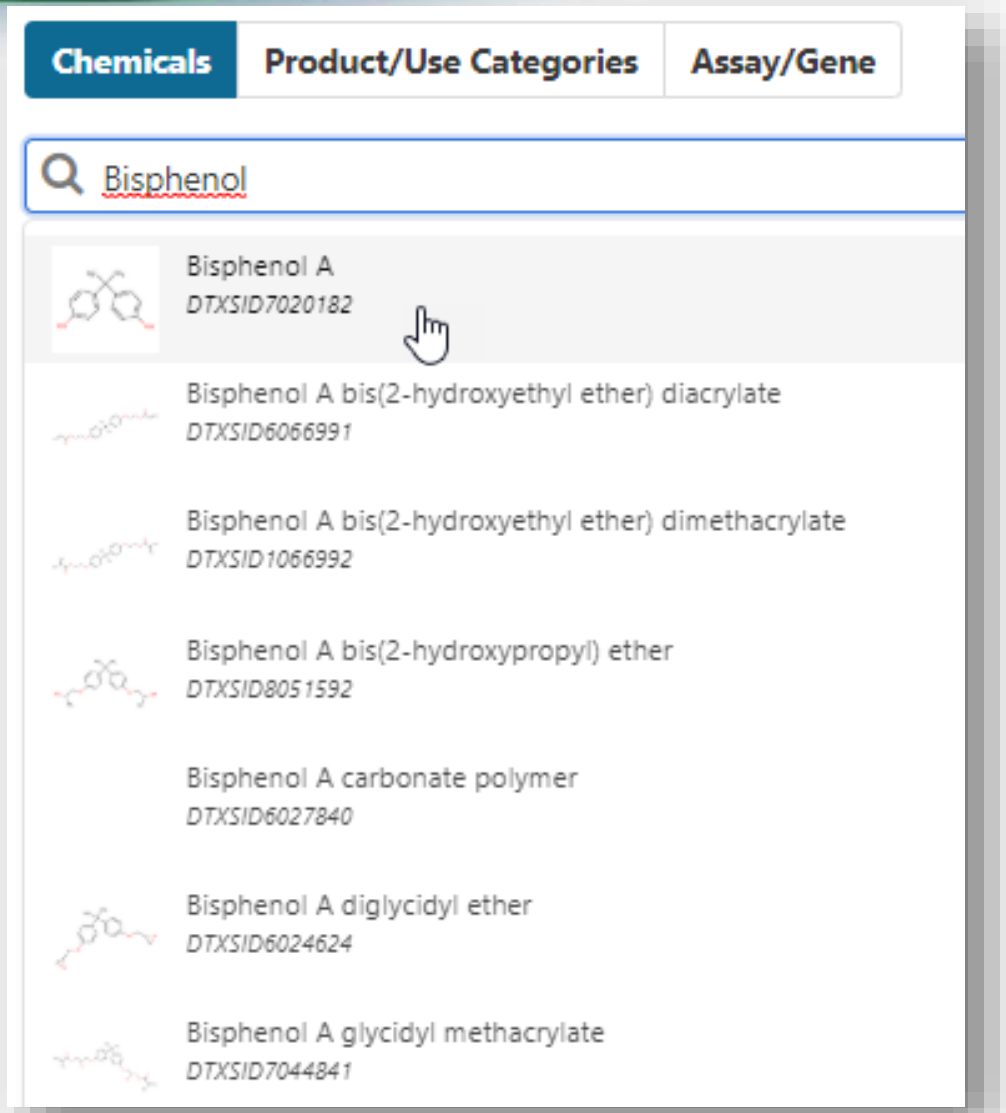
Green = Informatics & Web Services

# CompTox Chemicals Dashboard

https://comptox.epa.gov/dashboard



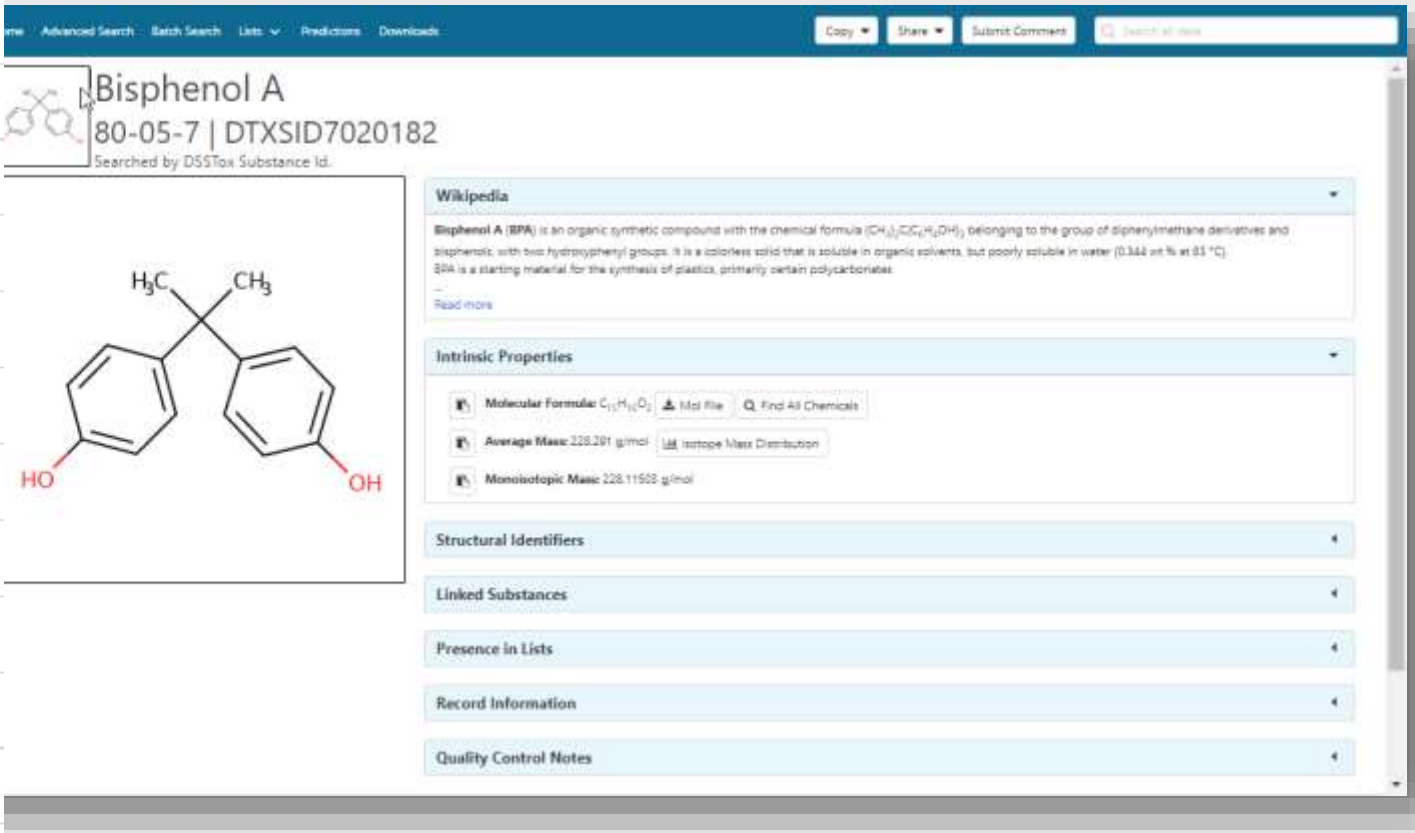
# BASIC Search



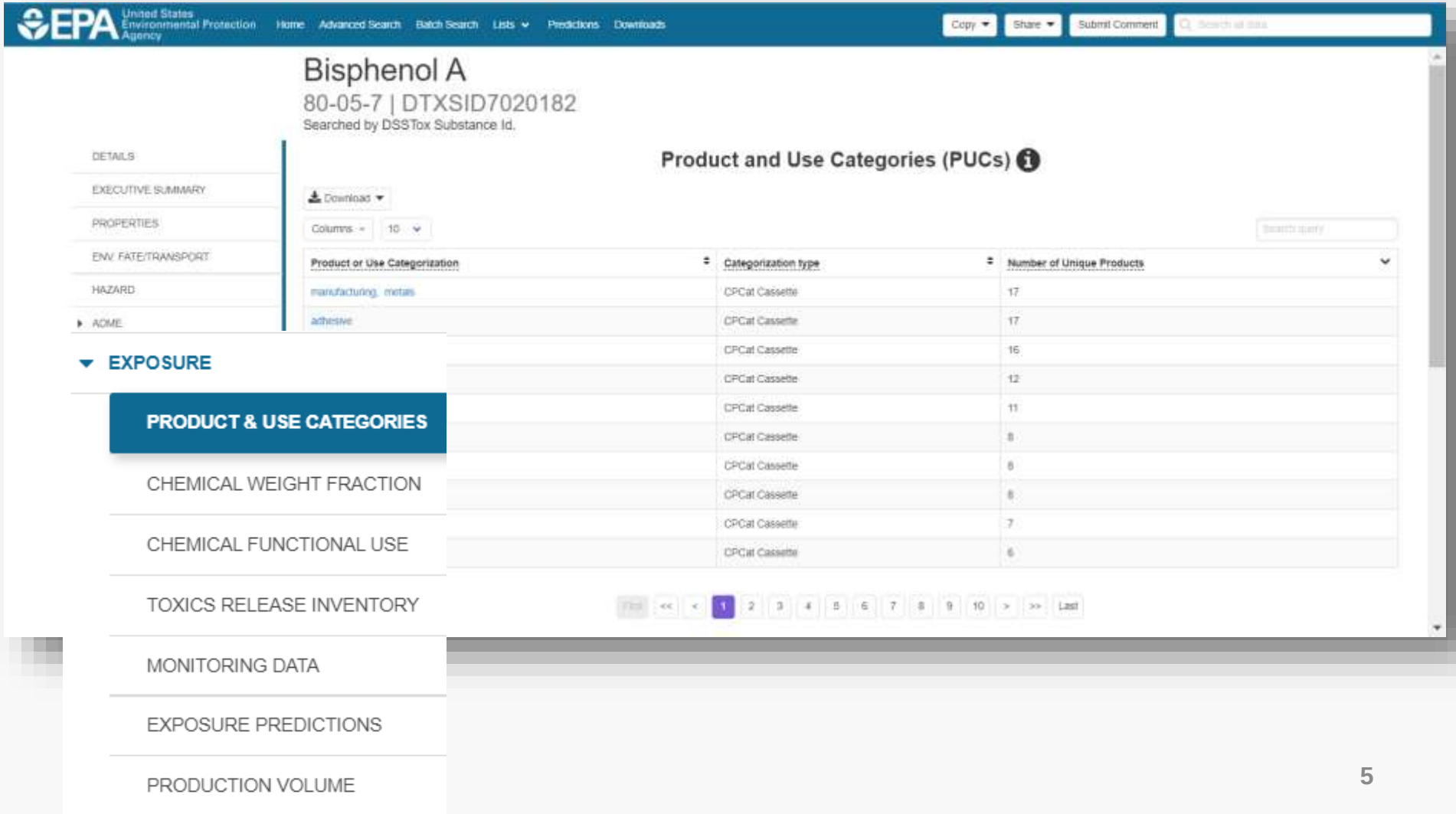
# Detailed Chemical Pages

EPA United States Environmental Protection Agency

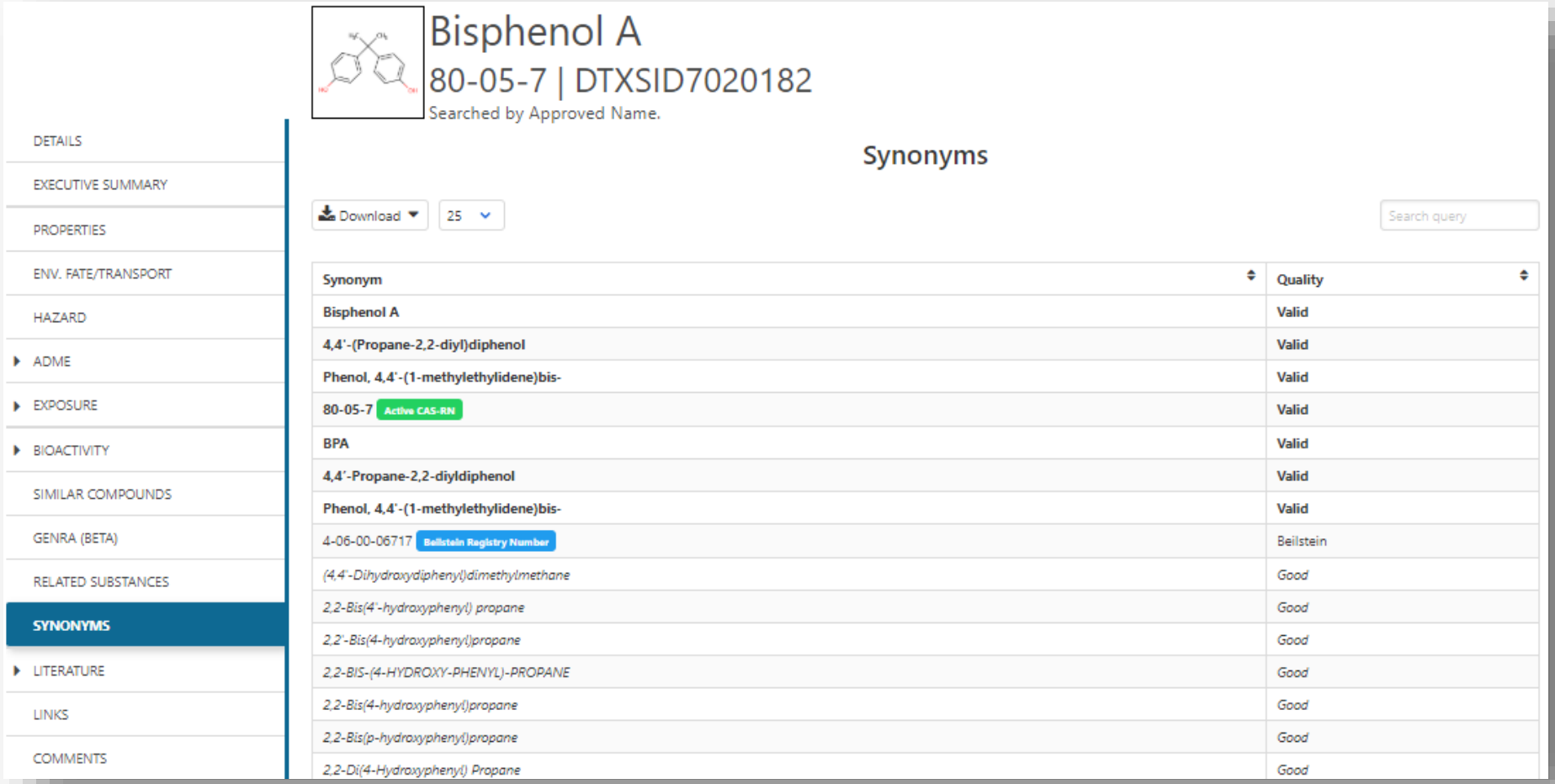
- **DETAILS**
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE
- LINKS
- COMMENTS



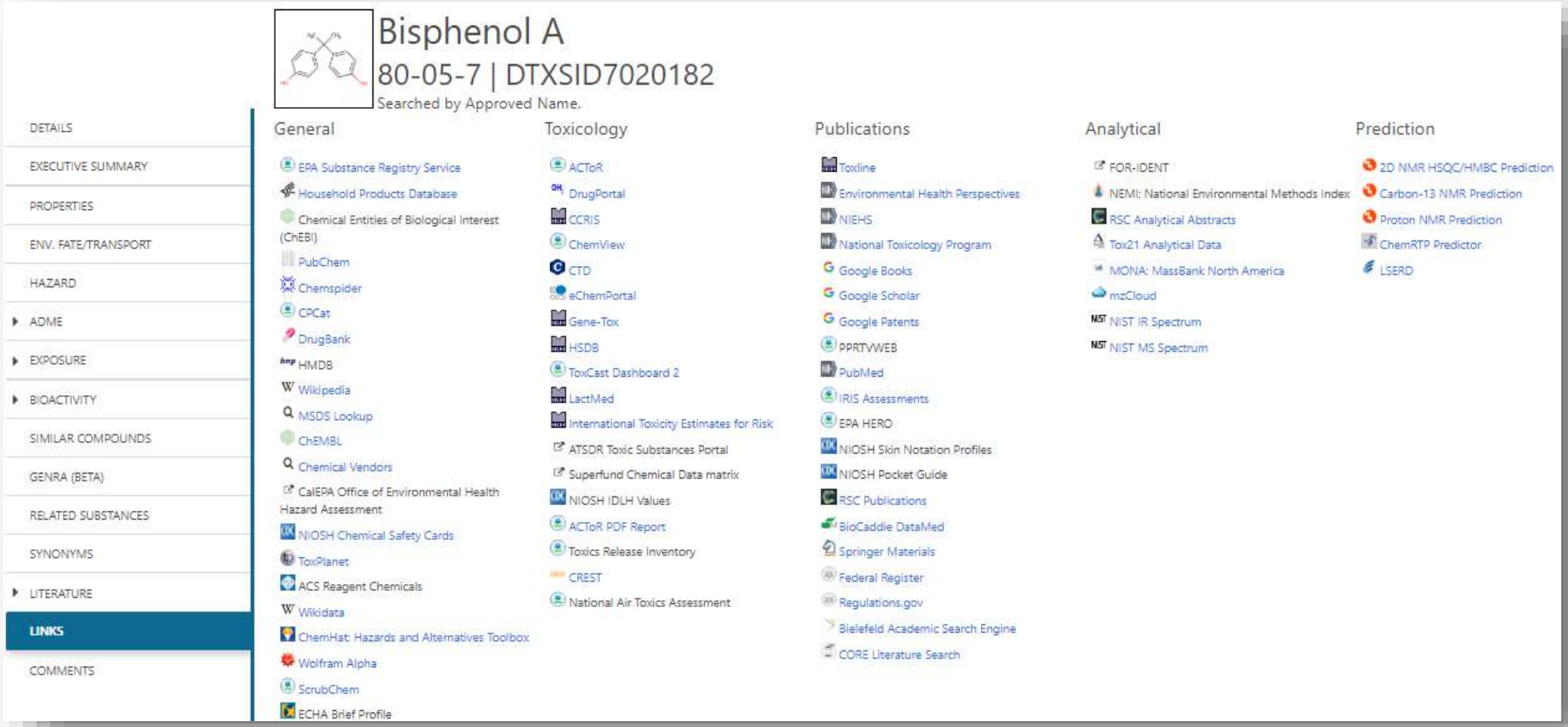
# Sources of Exposure to Chemicals



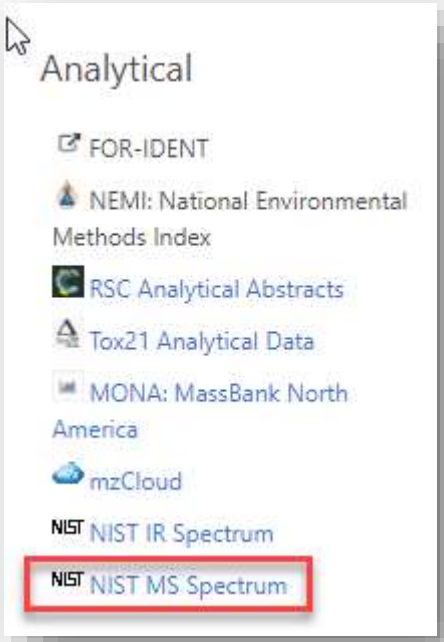
# Identifiers to Support Searches

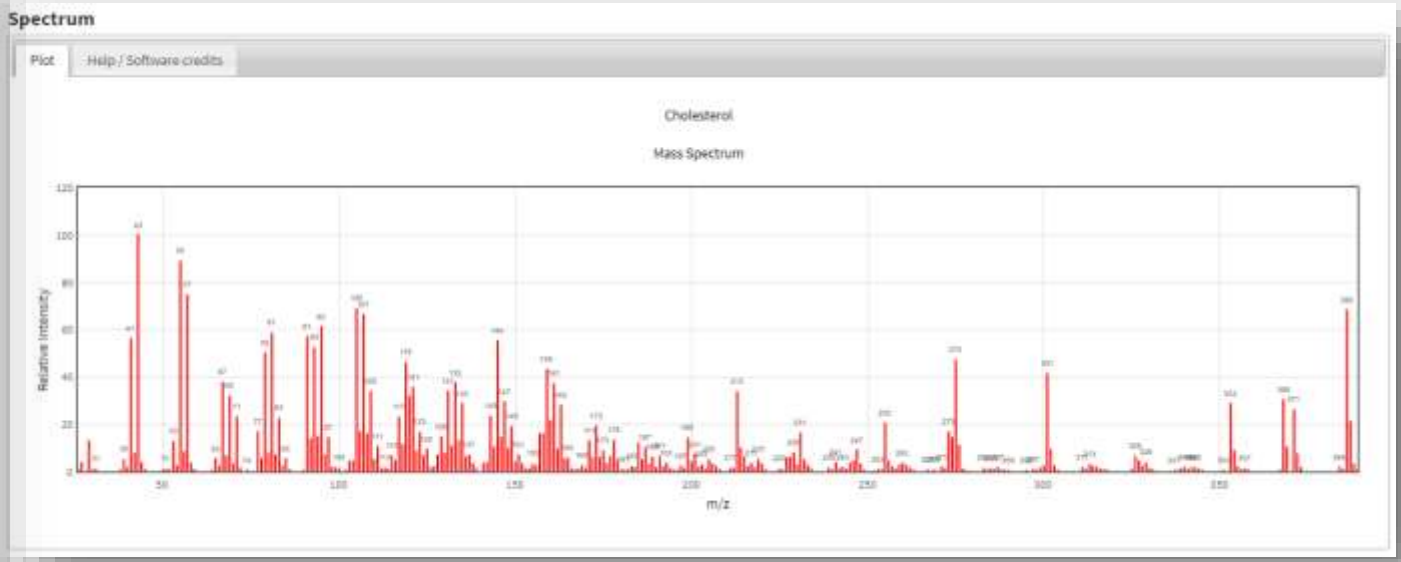


# Link Access

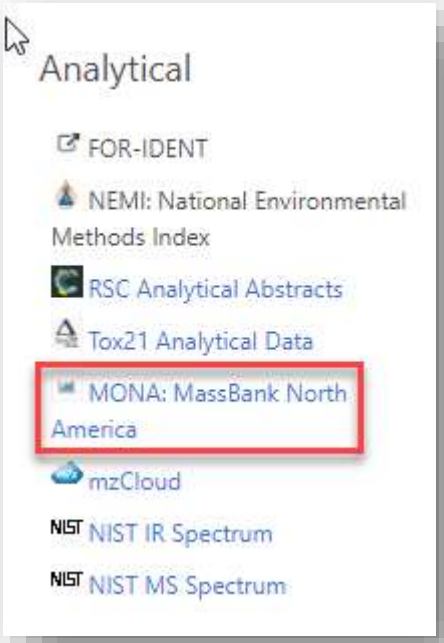

Bisphenol A
80-05-7 | DTXSID7020182
Searched by Approved Name.
DETAILS
EXECUTIVE SUMMARY
PROPERTIES
ENV. FATE/TRANSPORT
HAZARD
ADME
EXPOSURE
BIOACTIVITY
SIMILAR COMPOUNDS
GENRA (BETA)
RELATED SUBSTANCES
SYNONYMS
LITERATURE
LINKS
COMMENTS
General
EPA Substance Registry Service
Household Products Database
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

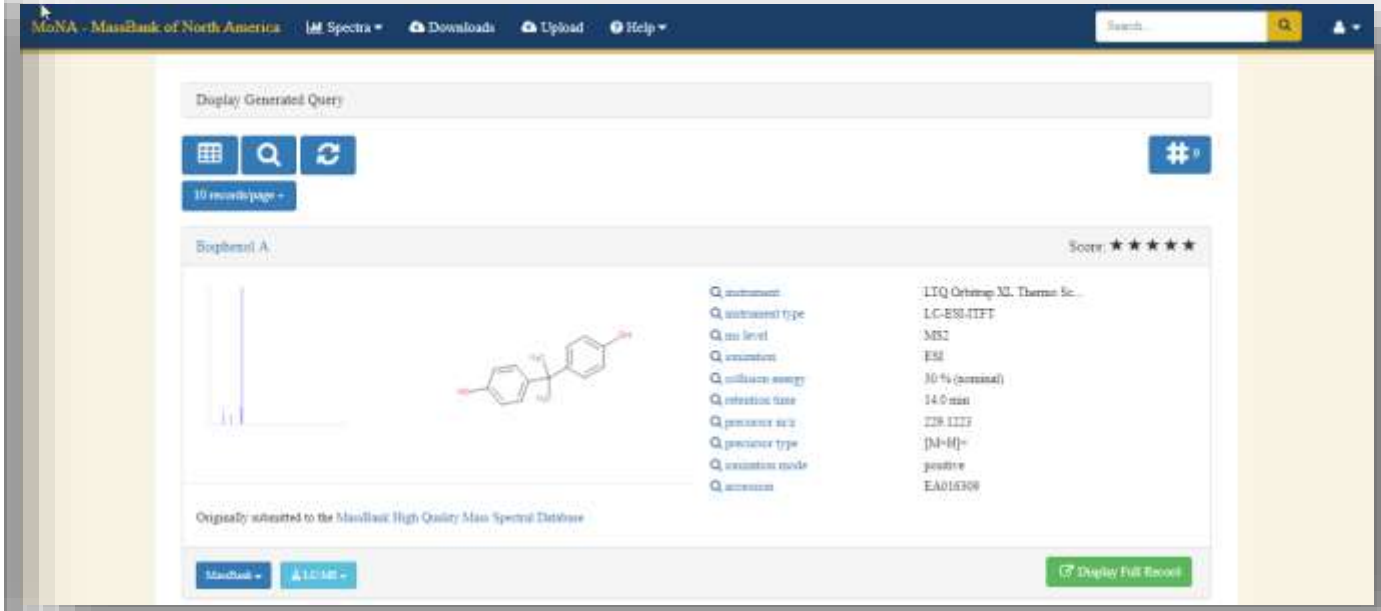
# NIST WebBook
# https://webbook.nist.gov/chemistry/



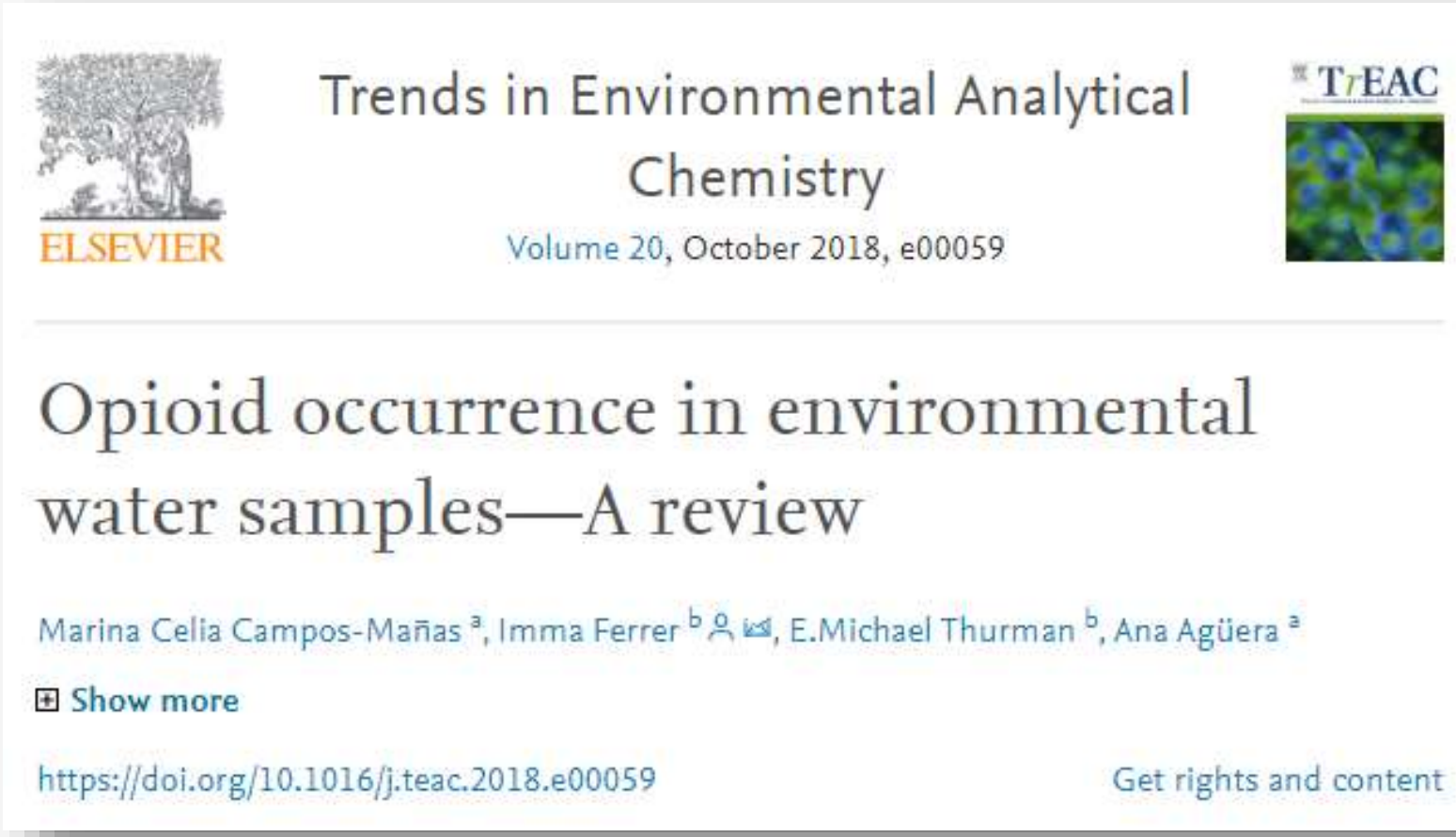
# MassBank of North America
# https://mona.fiehnlab.ucdavis.edu



# 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]

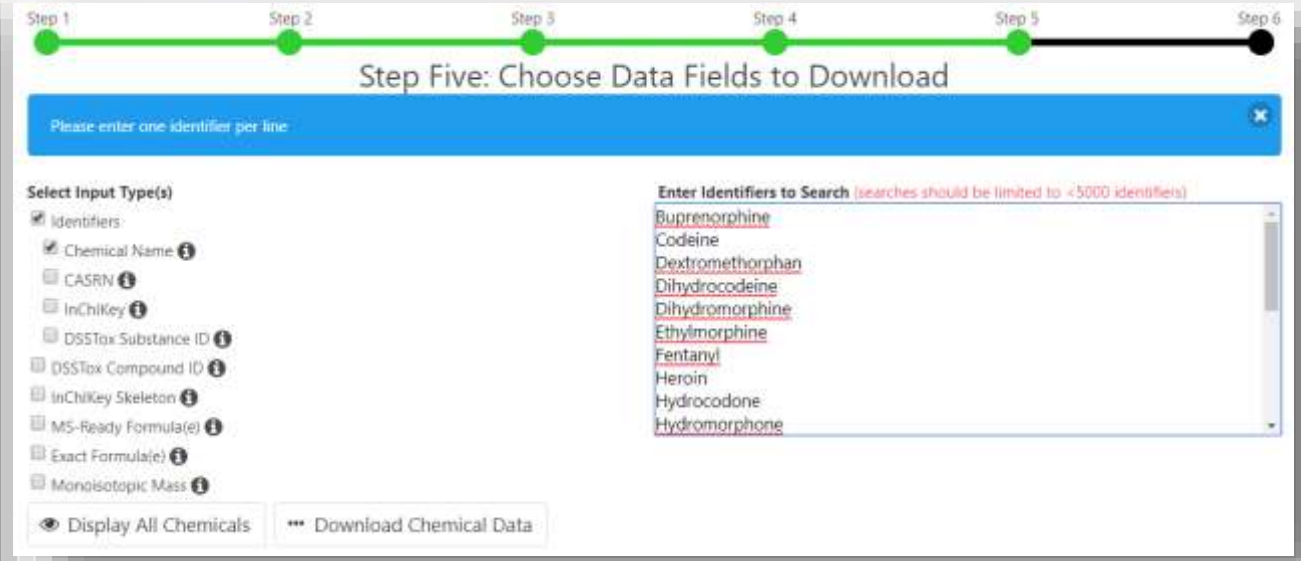
Show more

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

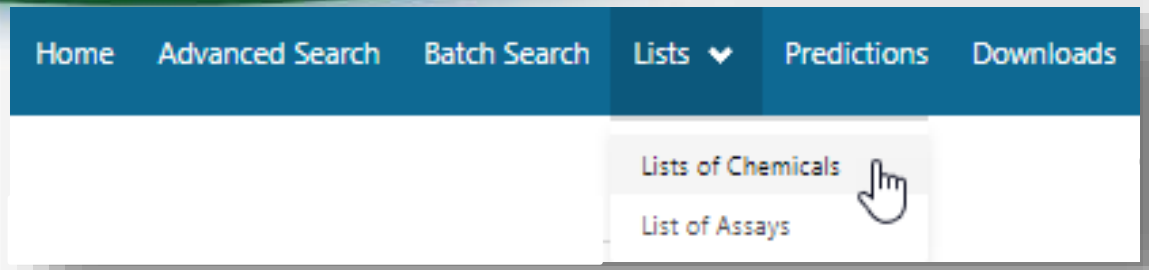
**Intrinsic And Predicted Properties**

- [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

# 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) |

# EPAHFR: Hydraulic Fracturing

## WATER|EPA; Chemicals associated with hydraulic fracturing

Search EPAHFR Chemicals

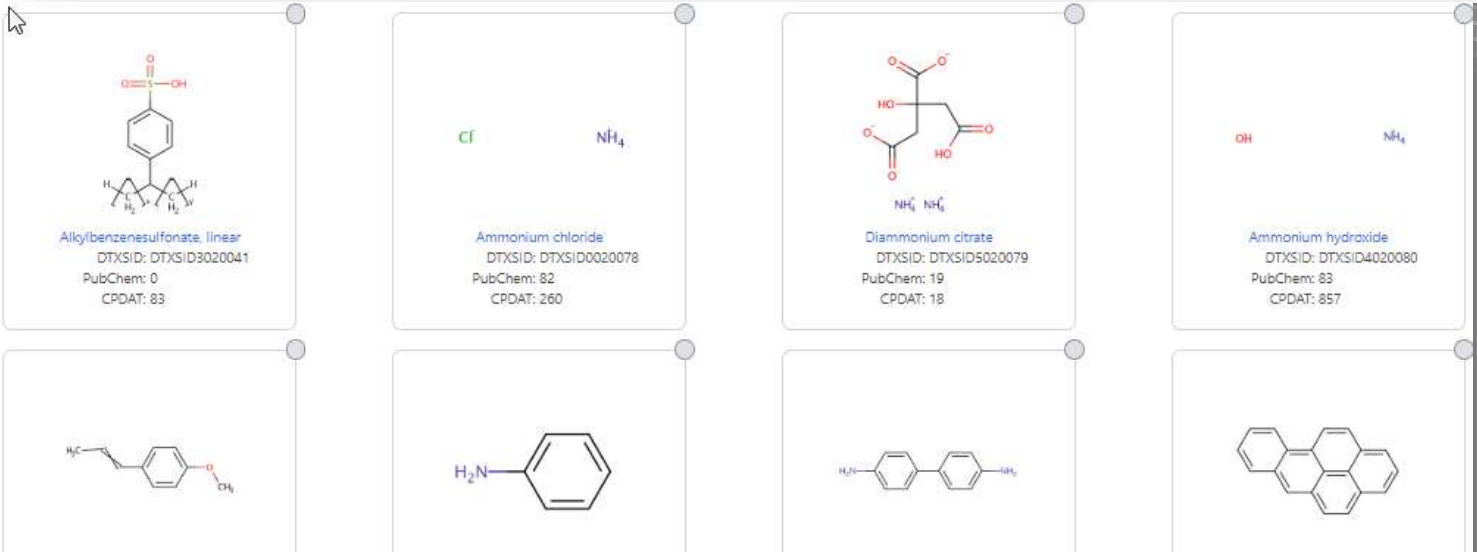
Identifier substring search

**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



# 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

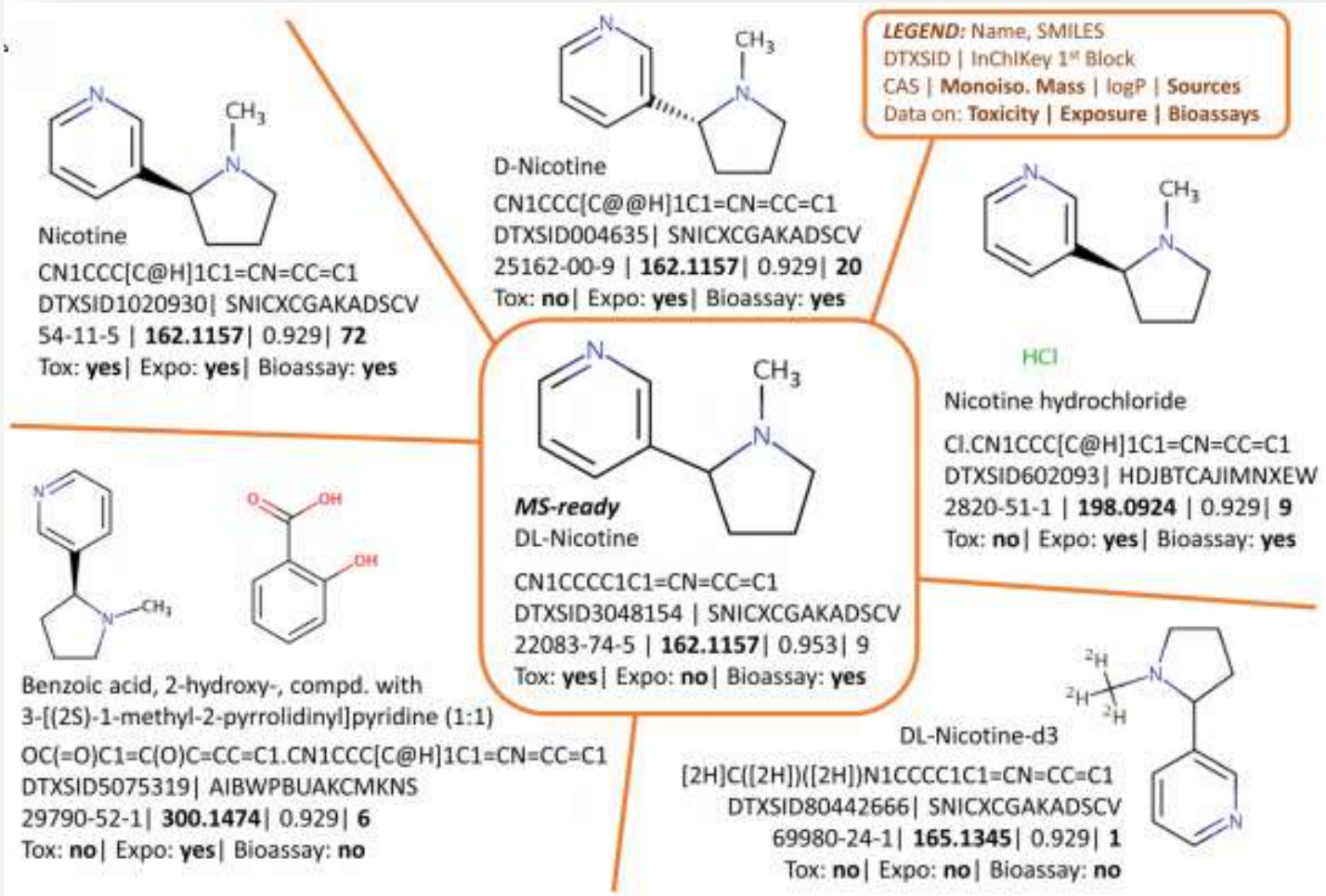
Environmental Science & Technology

Viewpoint

pubs.acs.org/est

# Open Science for Identifying "Known Unknown" Chemicals

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



# “MS-Ready Structures”

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

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

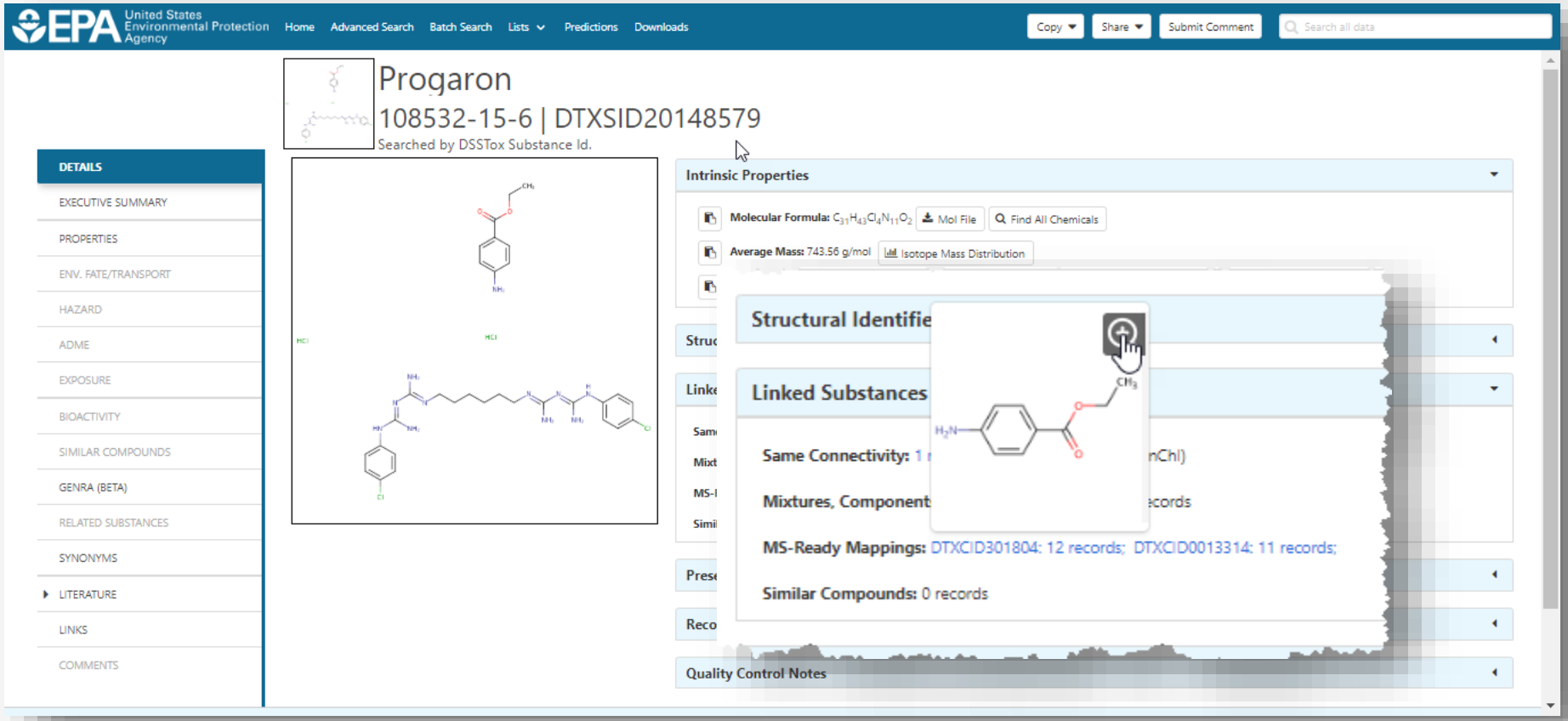
Journal of Cheminformatics

METHODOLOGY Open Access

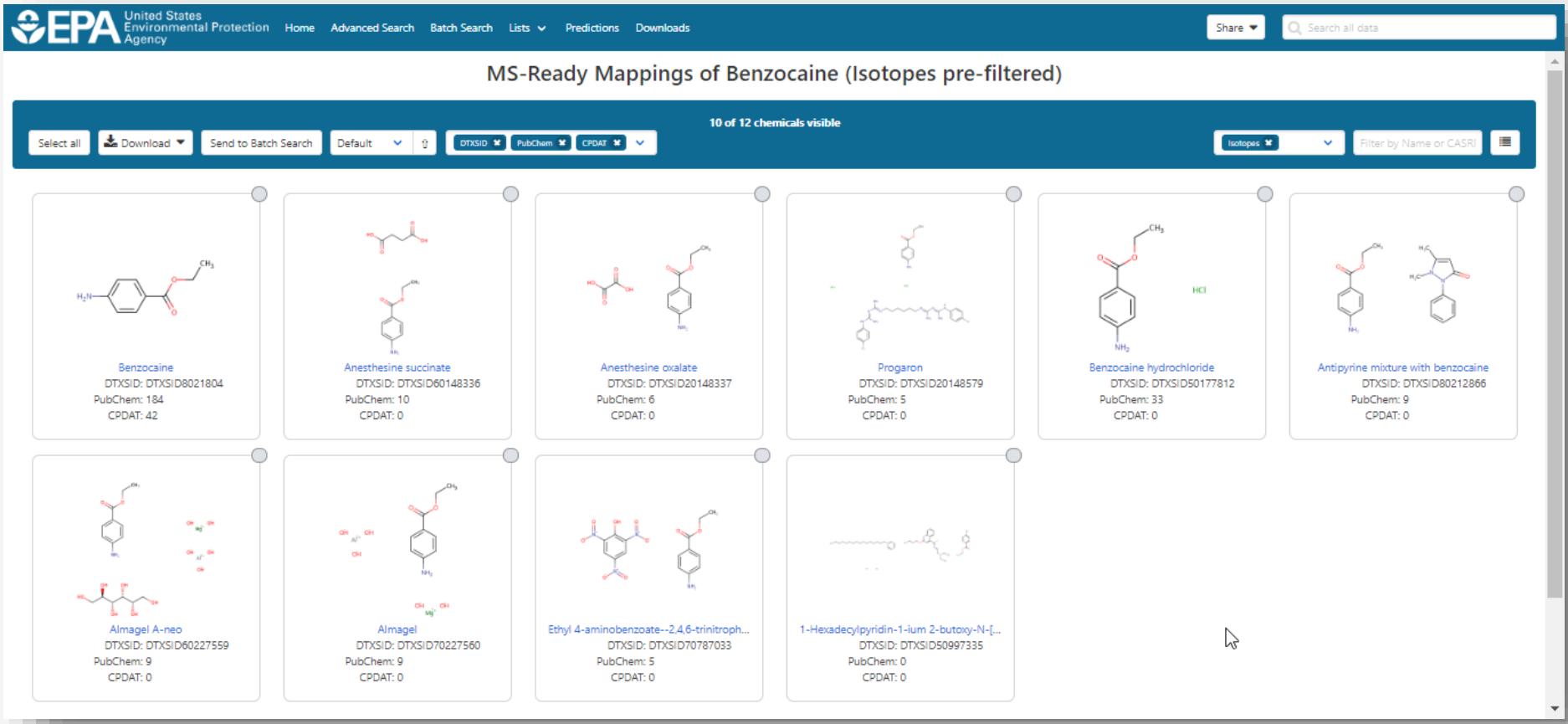
## “MS-Ready” structures for non-targeted high-resolution mass spectrometry screening studies

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*]

# MS-Ready Mappings

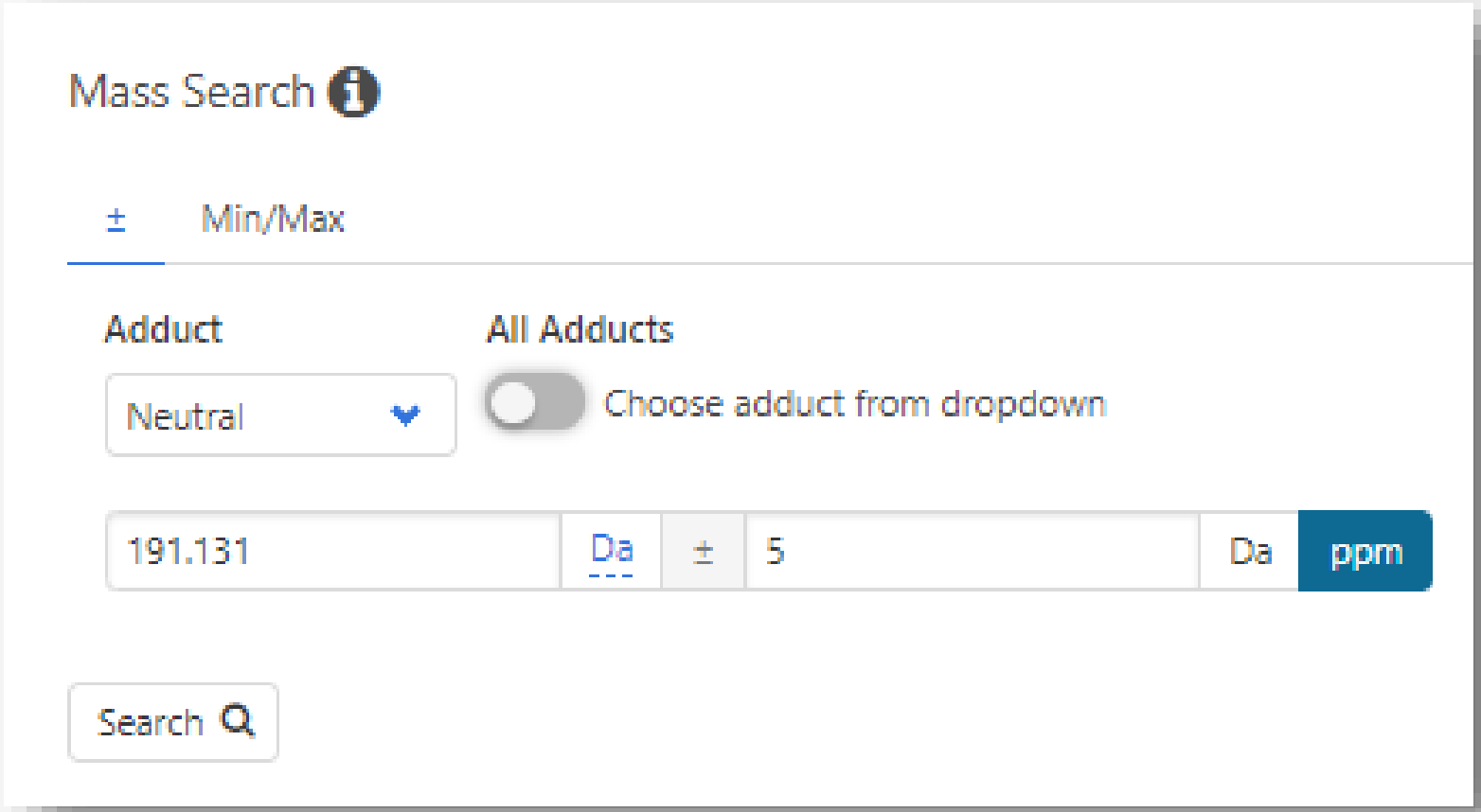


# MS-Ready Mappings Set

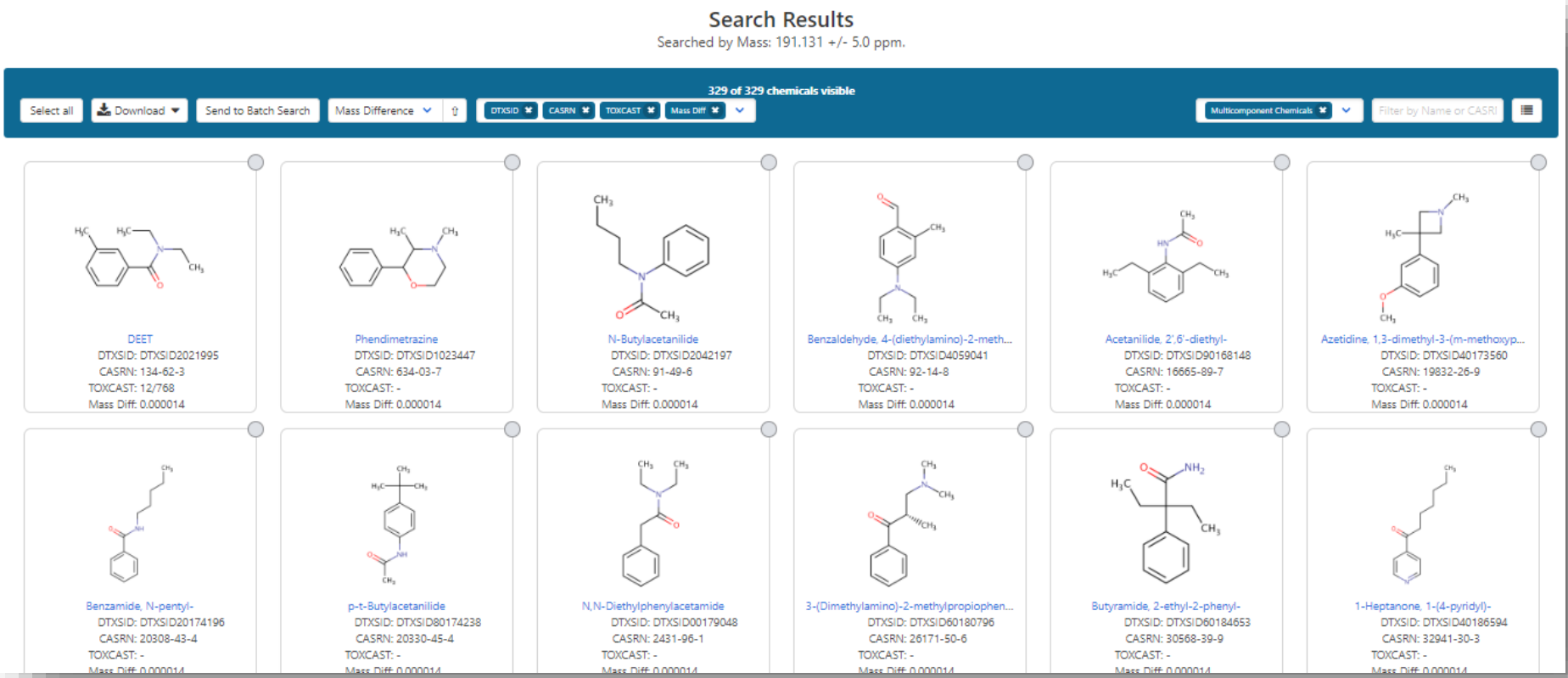


# MASS AND FORMULA SEARCHING

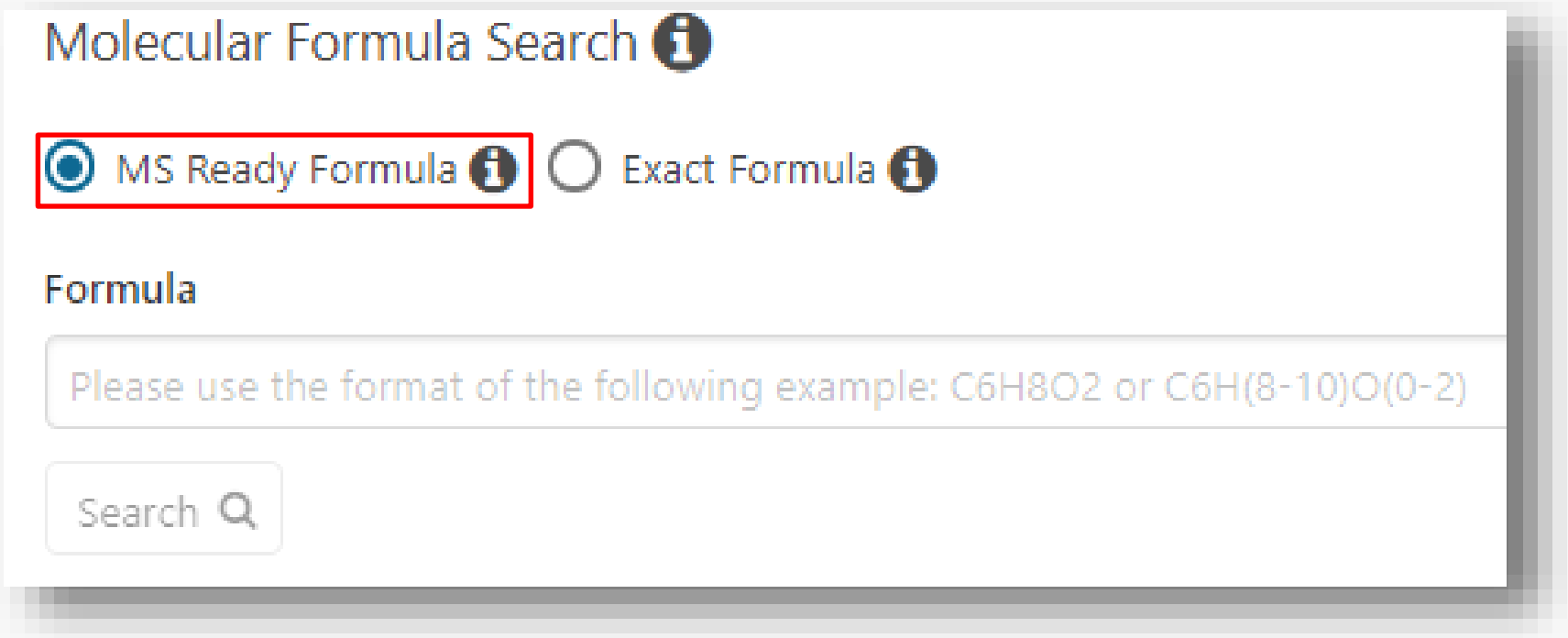
# Advanced Searches
# **Mass** Search



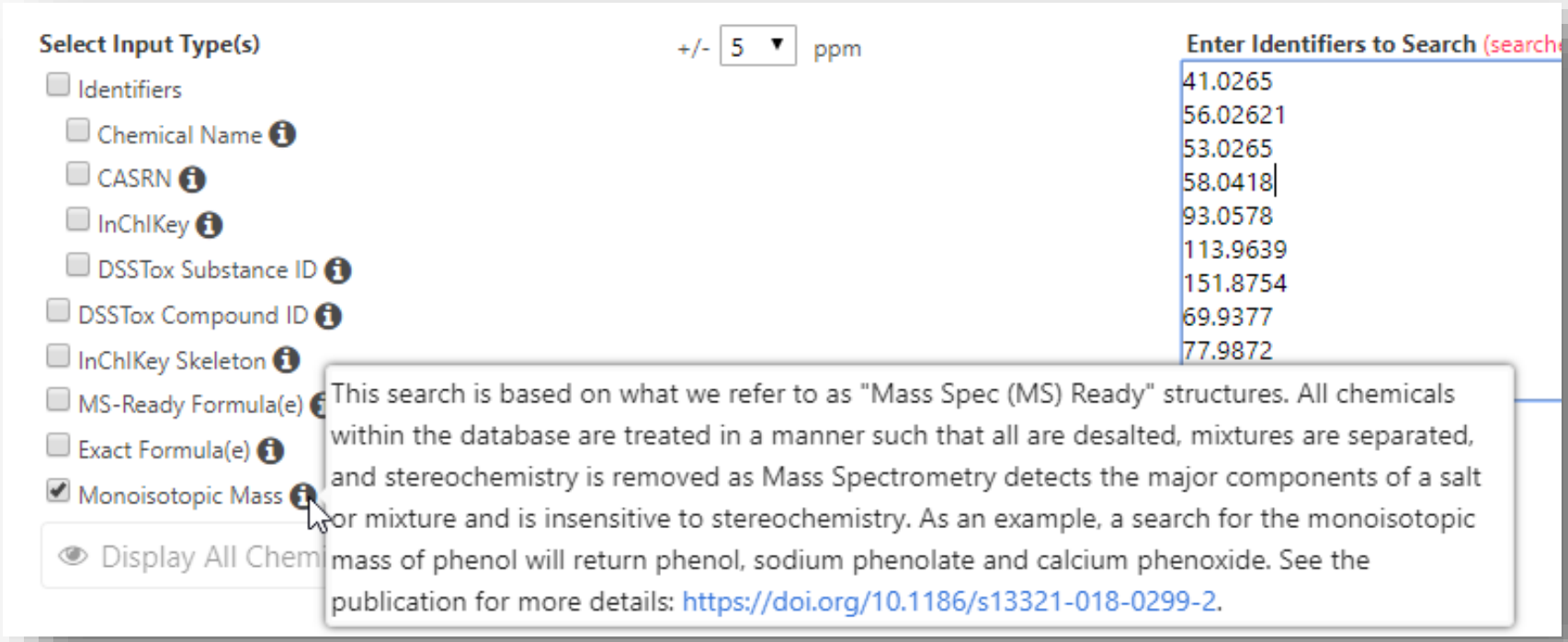
# Advanced Searches
# **Mass** Search



# MS-Ready Structures for **Formula Search**

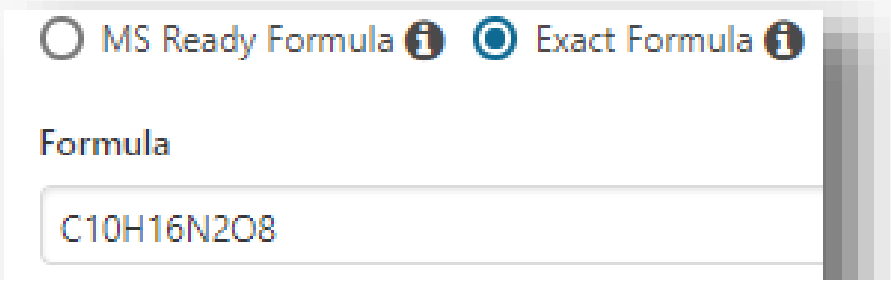


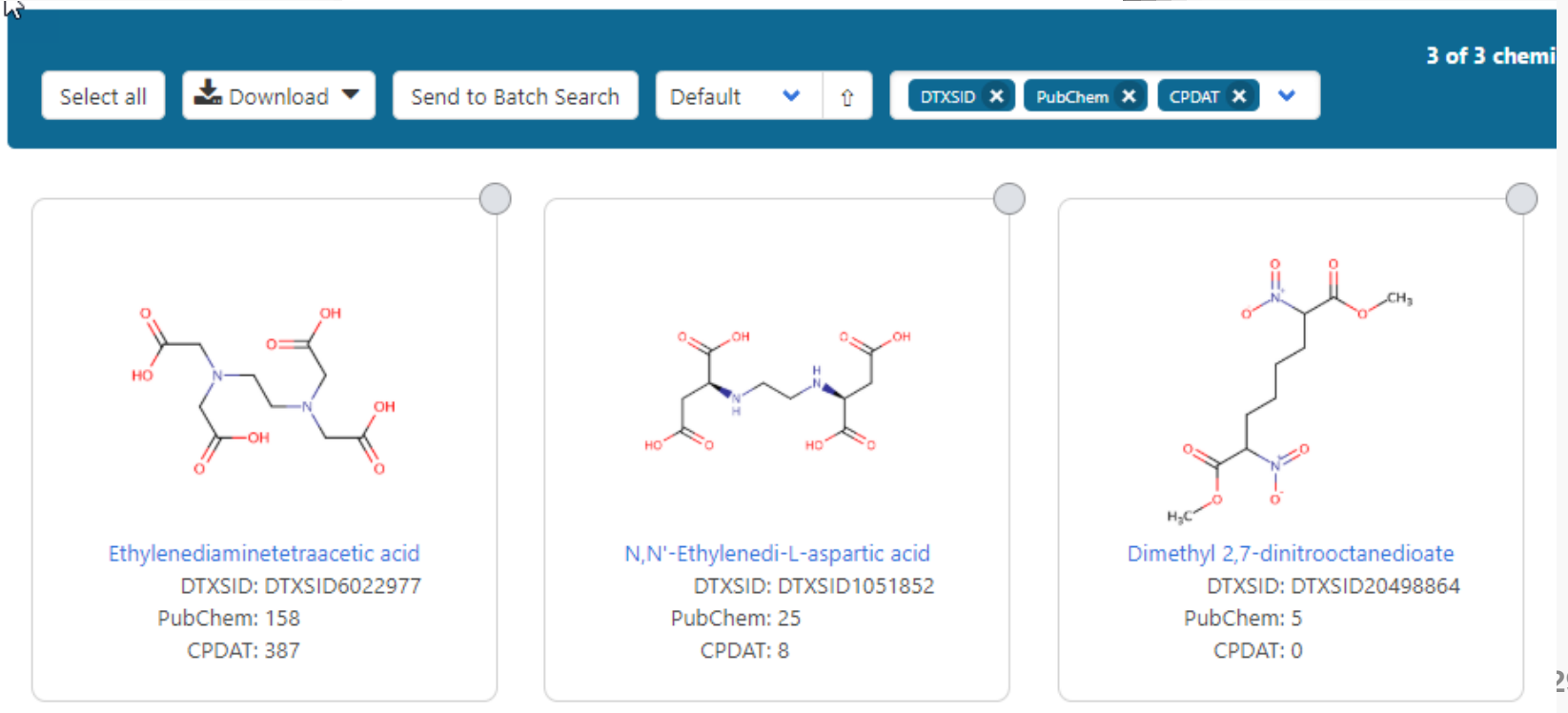
# MS-Ready Structures Batch Searches



# MS-Ready Mappings

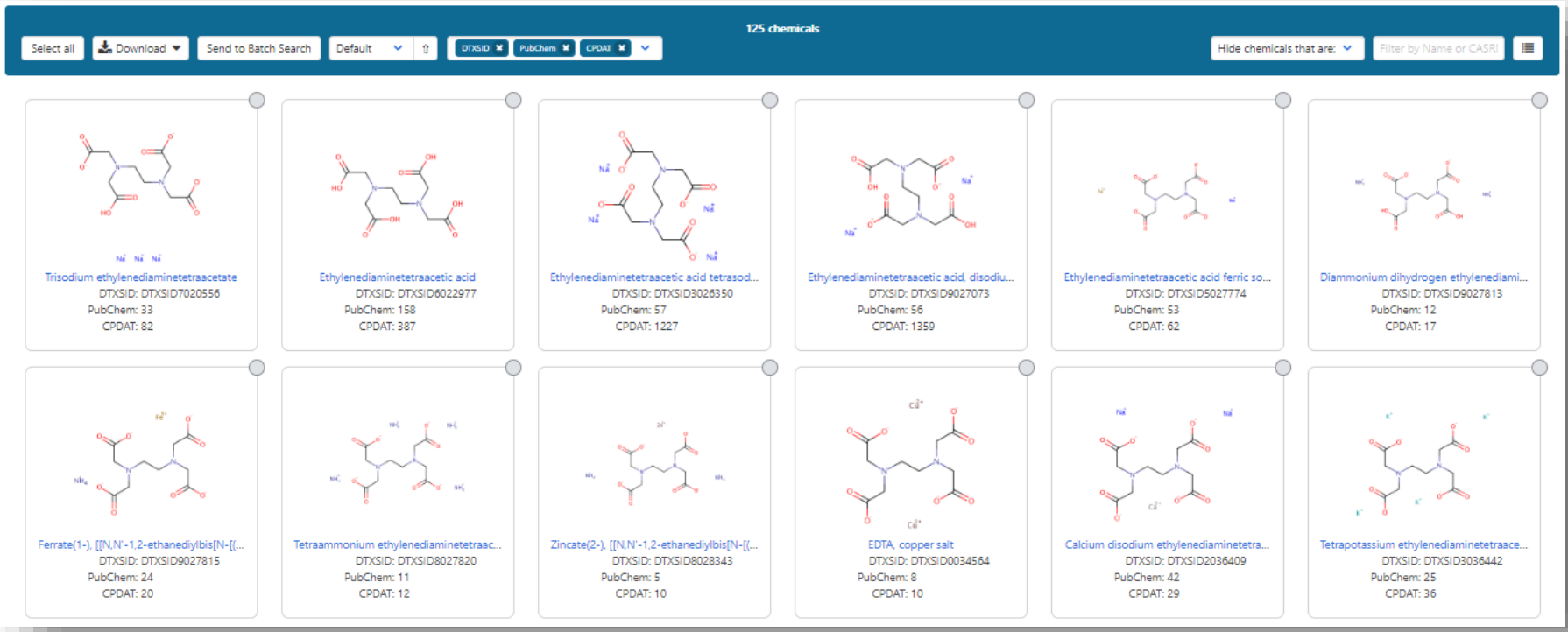
- **EXACT Formula**: C10H16N2O8: 3 Hits



# MS-Ready Mappings

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



# MS-Ready Mappings

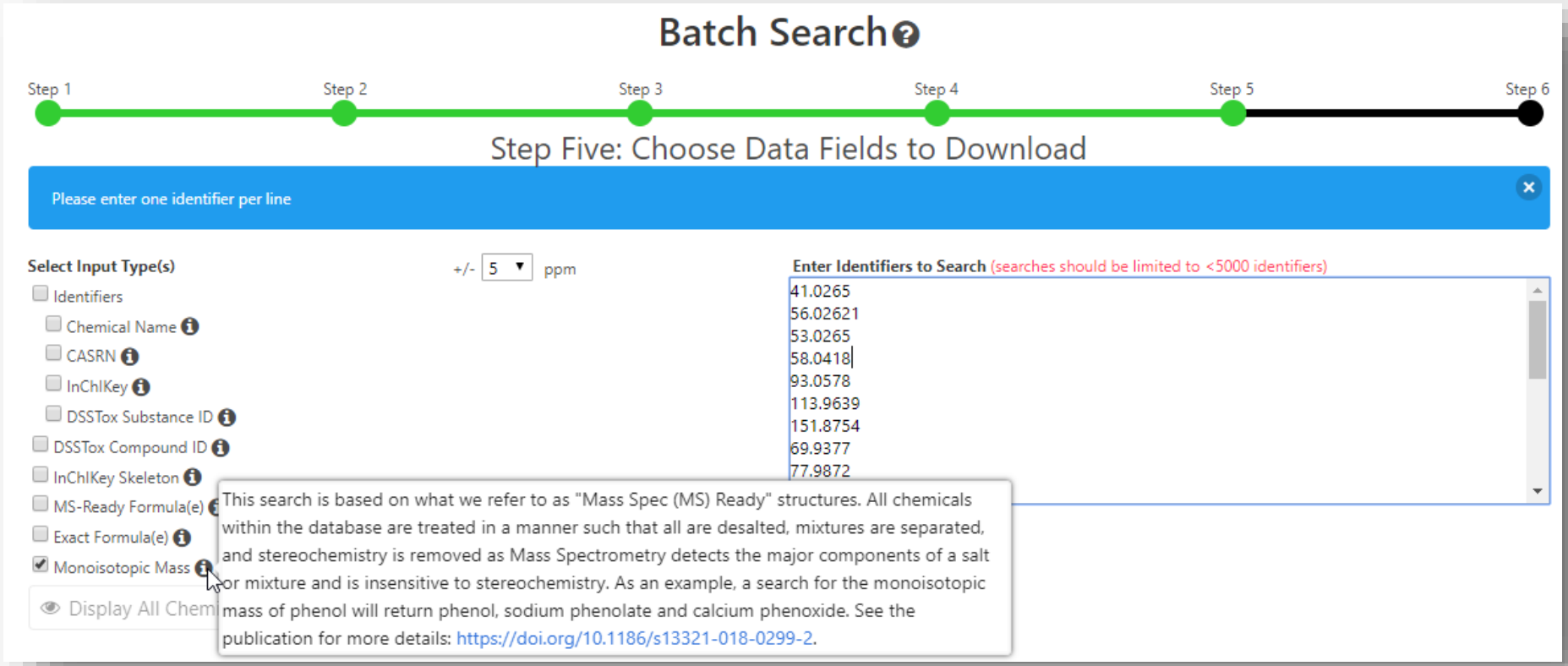
- 125 chemicals returned in total
  - 8 of the 125 are **single component** chemicals
  - 3 of the 8 are **isotope-labeled**
  - 3 are **neutral compounds** and 2 are **charged**

# Batch Searching

- 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



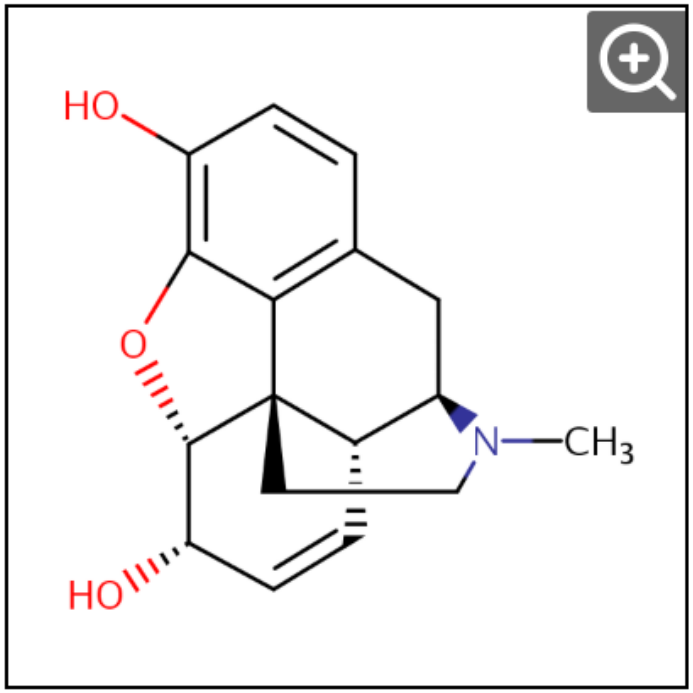
# Searching batches using MS-Ready Formula (or mass) searching

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 |

# SUPPORTING FUNCTIONALITY FOR MASS-SPEC

# 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**

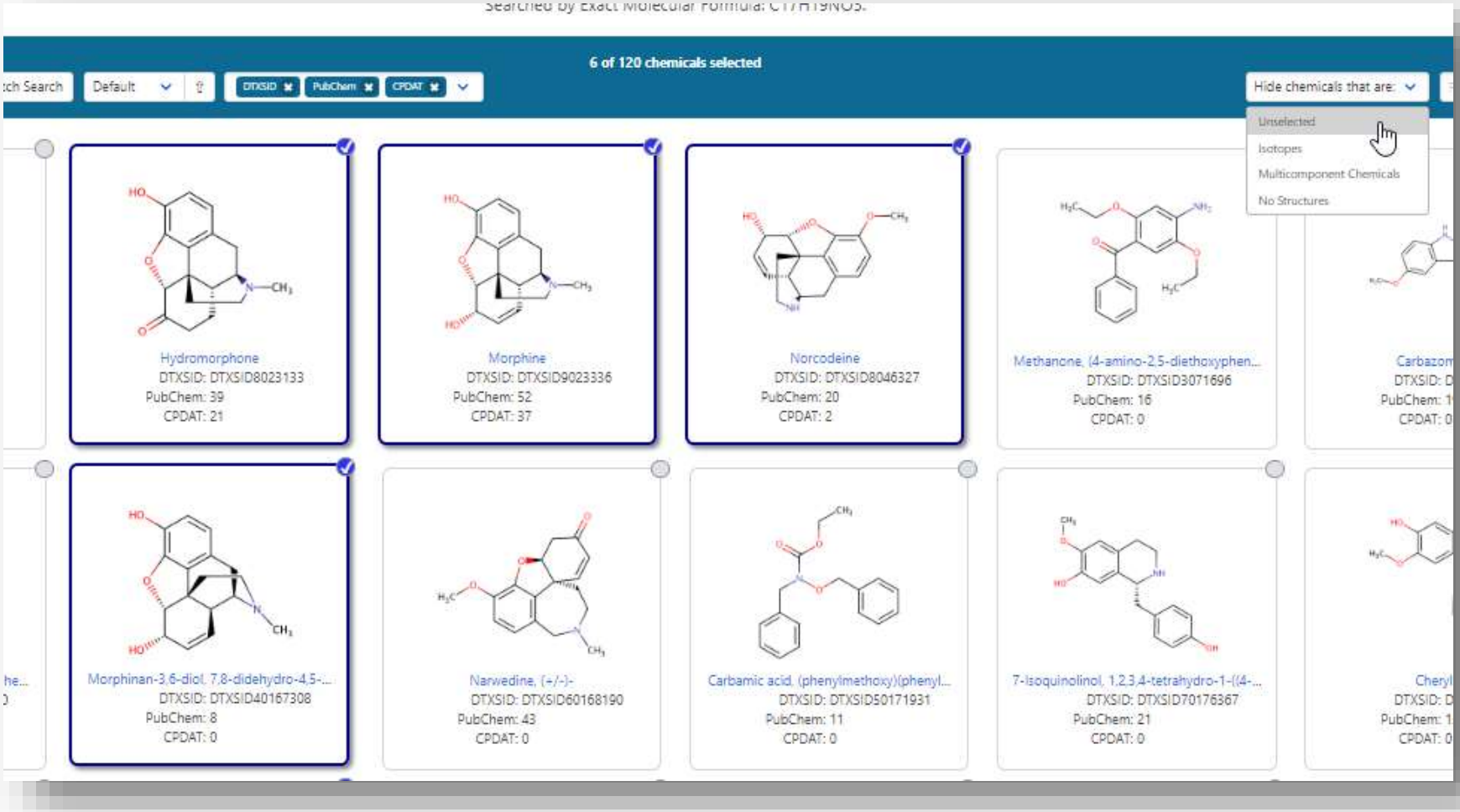
**Molecular Formula:** $C_{17}H_{19}NO_3$ Mol File Find All Chemicals

**Average Mass:** 285.343 g/mol Isotope Mass Distribution

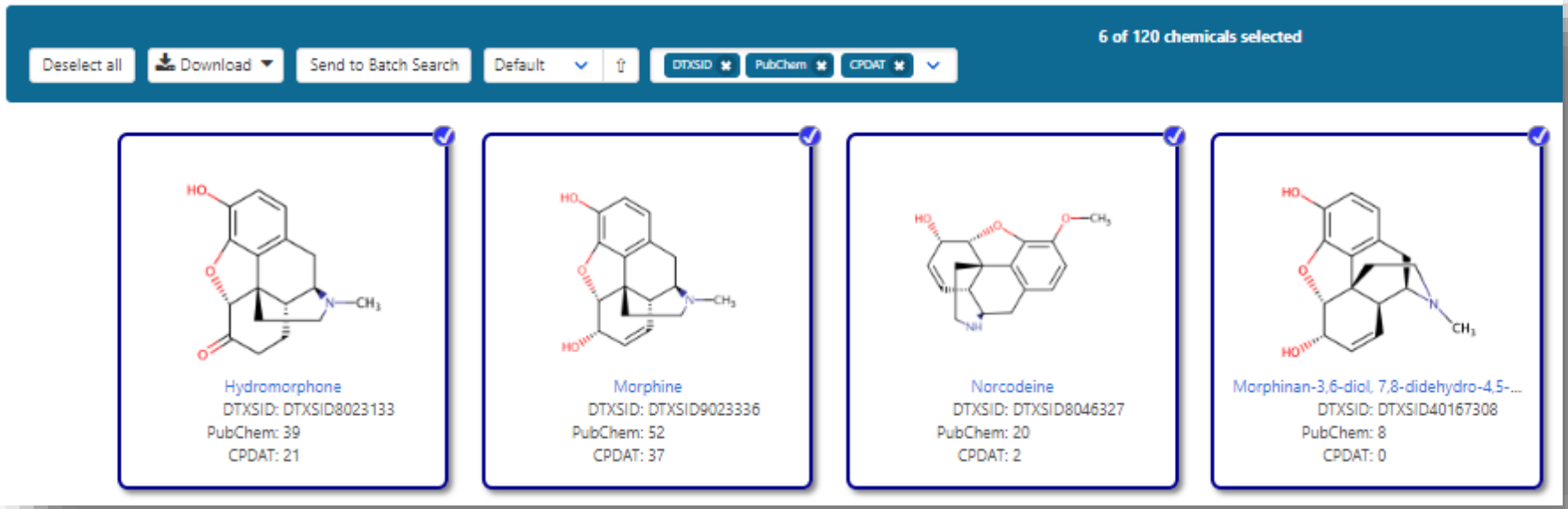
**Monoisotopic Mass:** 285.136493 g/mol

**Structural Identifiers**

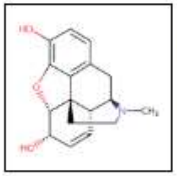
# Select Chemicals of Interest



# Prune to list of interest



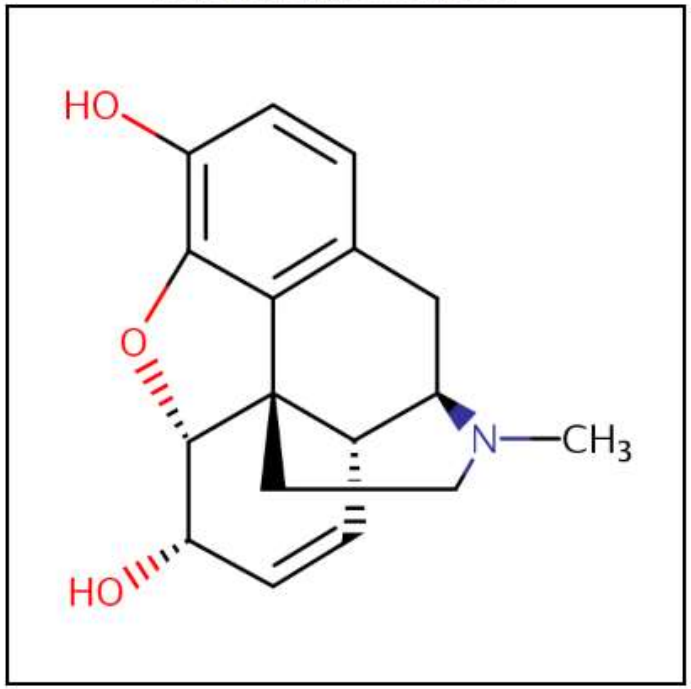
# Structure Similarity Searches



## Morphine

57-27-2 | DTXSID9023336

Searched by Approved Name.




**Wikipedia**

**Intrinsic Properties**

**Structural Identifiers**

**Linked Substances**

**Same Connectivity:** 3 records (based on first layer of InChI)

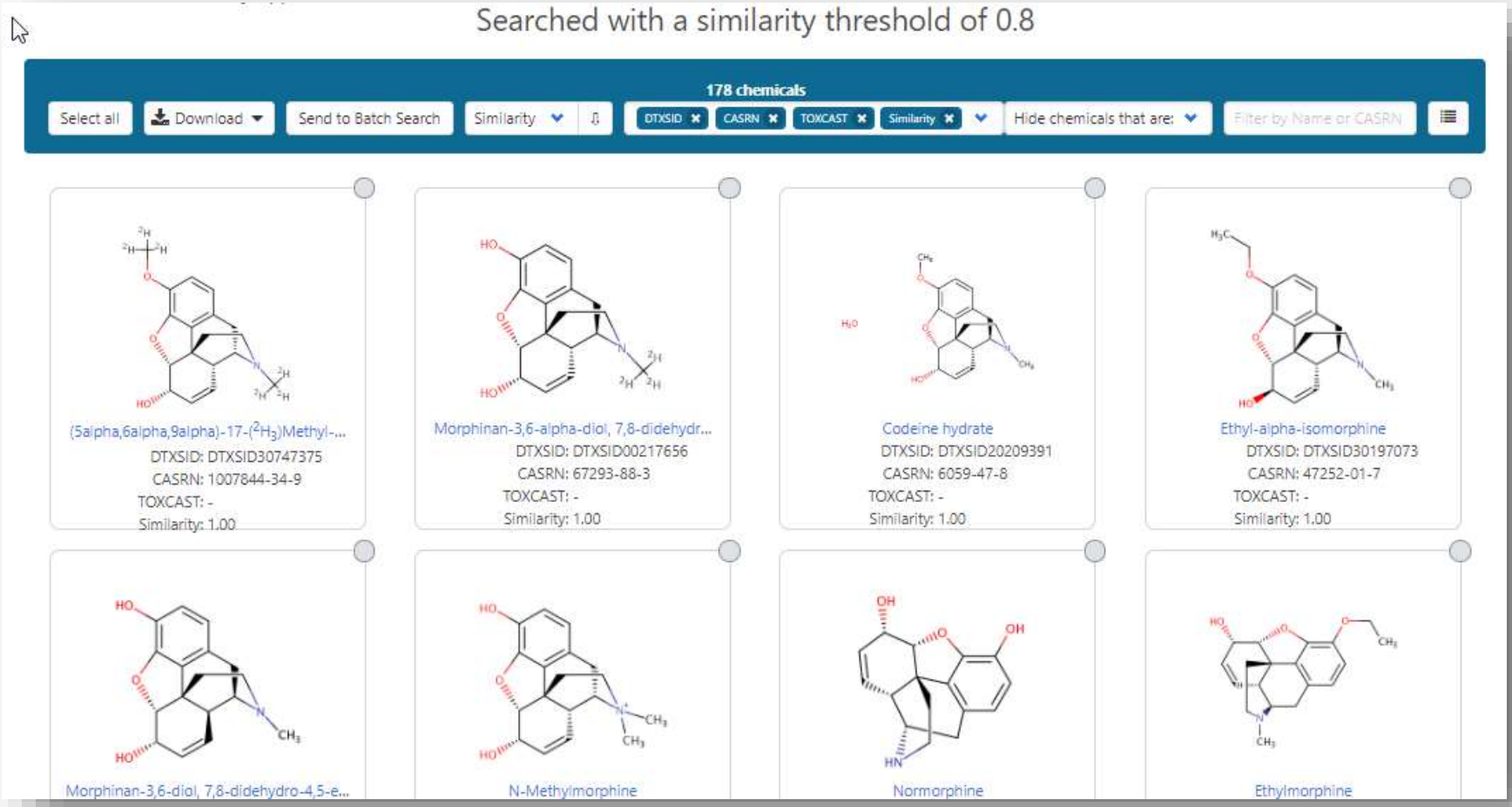
**Mixtures, Components and Neutralized Forms:** 13 records (based on QSAR ready mappings and with the compound as a component of a mixture)

**MS-Ready Mappings:** DTXCID60196731: 21 records;

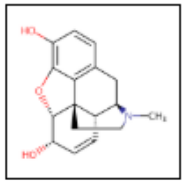
**Similar Compounds:** 178 records (based on Tanimoto coefficient >0.8)

**Presence in Lists**

# Structure Similarity Searches



# 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

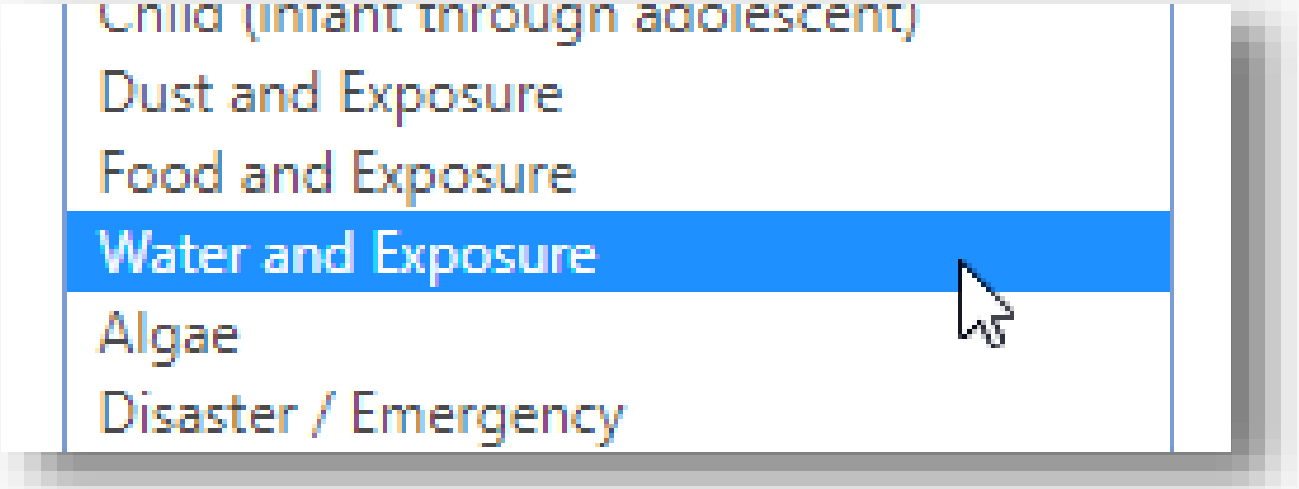
Retrieve Articles

- 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

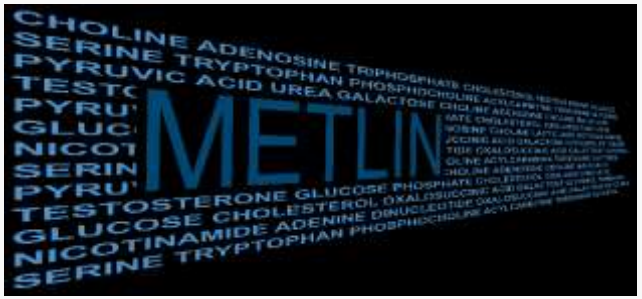

To find articles quickly, enter terms to sift abstracts.

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 | |

# Example Online Resources for MS

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




PubChem

GNPS

MassBank

ChemSpider
Search and share chemistry

# DO WE REALLY NEED ANOTHER DATABASE?

# Is a bigger database better?

Journal of The American Society for Mass Spectrometry
January 2012, Volume 23, Issue 1, pp 179–185 | Cite as

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

Authors Authors and affiliations

James L. Little, Antony J. Williams, Alexey Pshenichnov, Valery Tkachenko

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

71 Million chemical structures

# 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?

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!

# 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

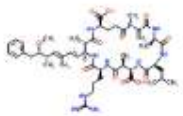
*Systematic Name, Synonym, Trade Name, Registry Number, SMILES, InChI or CSID*

FILTER

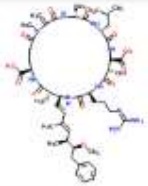
Search Hits Limit: 100

Found 9 results

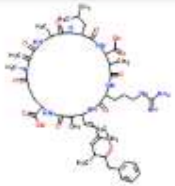
Search term: **ZYZCGGRZINLQBL** (Found by InChIKey (skeleton match))



Cyanoginosin

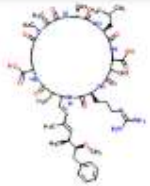


MCYST-LR



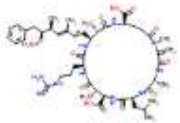
15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E)-6-methoxy-3,5-dimethyl-7-phenyl-1,3-heptadien-1-

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

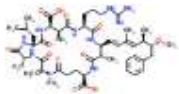


(5R,8S,15S)-15-(3-[(Diaminomethylene)amino]propyl)-8-isobutyl-18-[(1E,3E,5S,6S)-6-methoxy-

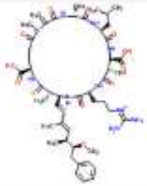
15-(3-[(Diaminomethylene)amino]propyl)-8-isobutyl-18-(6-methoxy-3,5-dimethyl-7-



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

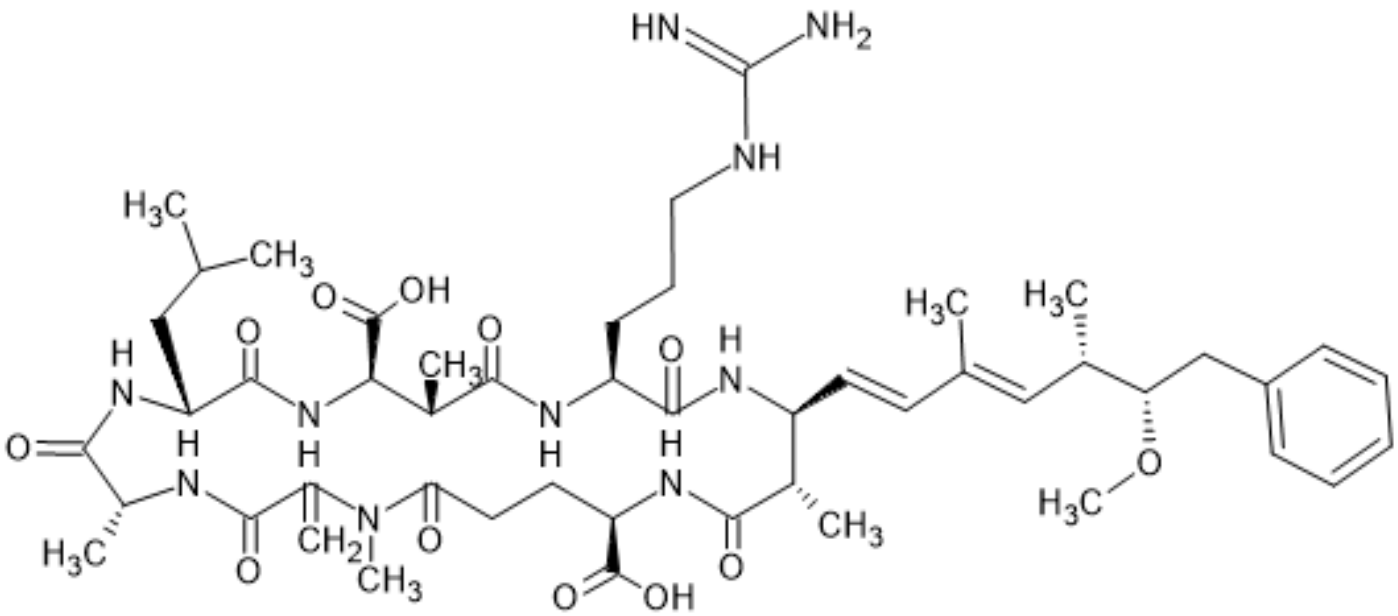


(5R,8R,11R,12S,15S,18S,22R)-15-(3-Carbamimidamidopropyl)-8-isobutyl-18-[(1E,3E,5R,6R)-6-methoxy-3,5-dimethyl-



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

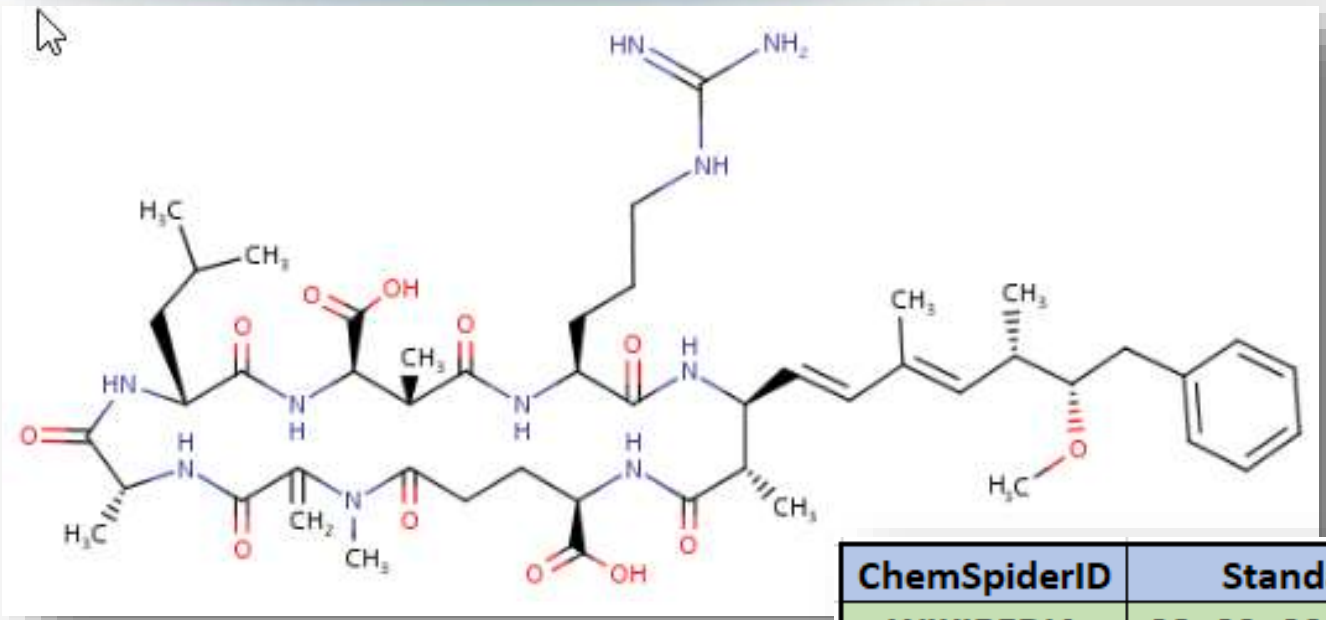
# International Chemical Identifier



InChI=1S/C49H74N10O12/c1-26(2)23-37-46(66)58-40(48(69)70)30(6)42(62)55-35(17-14-22-52-49(50)51)45(65)54-34(19-18-27(3)24-28(4)38(71-10)25-33-15-12-11-13-16-33)29(5)41(61)56-36(47(67)68)20-21-39(60)59(9)32(8)44(64)53-31(7)43(63)57-37/h11-13,15-16,18-19,24,26,28-31,34-38,40H,8,14,17,20-23,25H2,1-7,9-10H3,(H,53,64)(H,54,65)(H,55,62)(H,56,61)(H,57,63)(H,58,66)(H,67,68)(H,69,70)(H4,50,51,52)/b19-18+,27-24+/t28-,29-,30-,31+,34-,35-,36+,37+,38-,40+/m0/s1

ZYZCGGRZINLQBL-JCGNTXOTSA-N

# 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

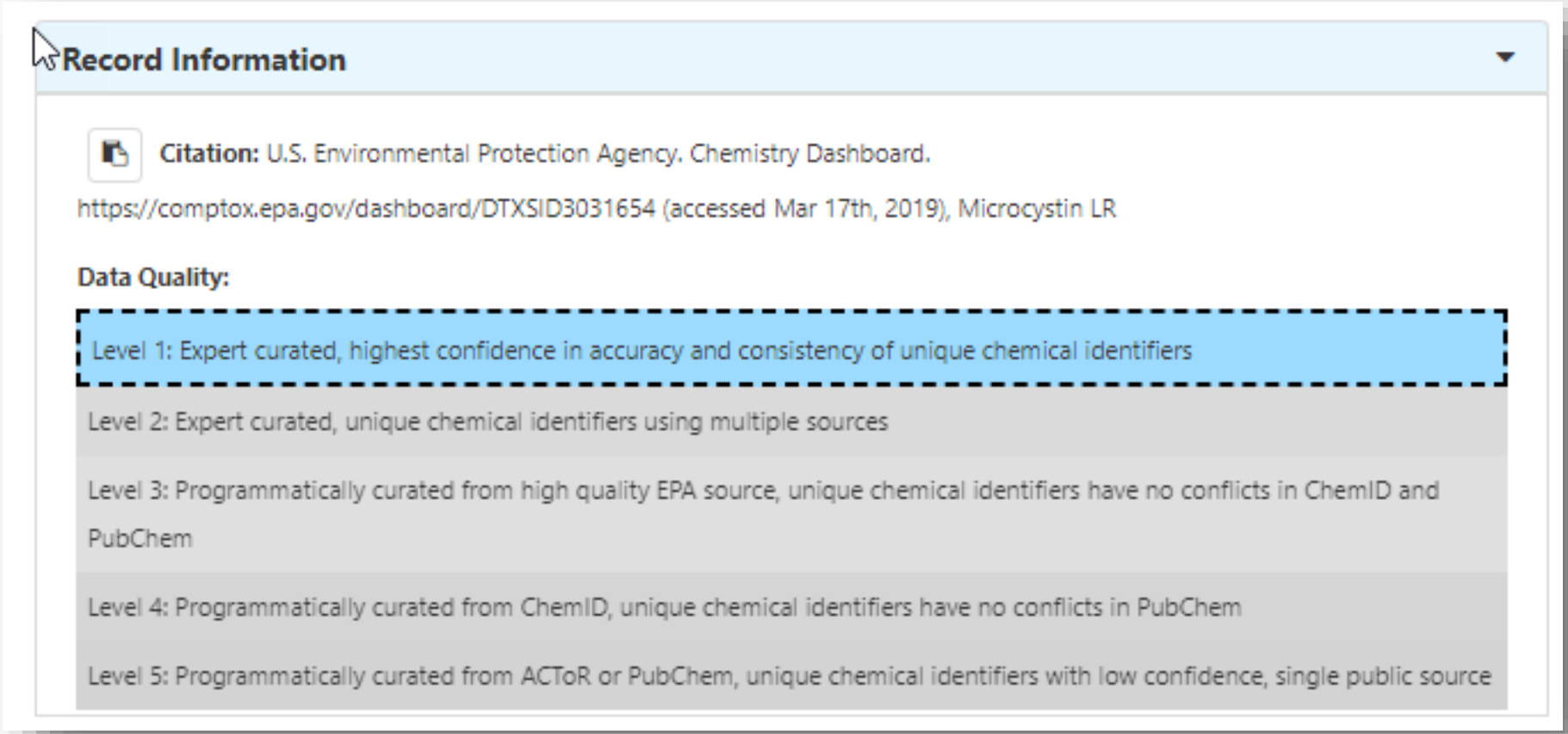

UniChem

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 |

# Delivering a Better Database

- We have full time curators checking data



# UVCB CHEMICAL SUBSTANCES

# UVCB Chemicals

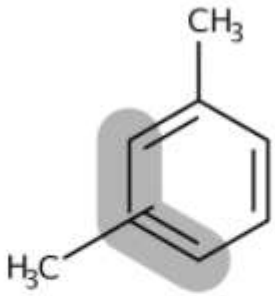
EPA United States Environmental Protection Agency

## Chemical Substances of Unknown or Variable Composition, Complex Reaction Products and Biological Materials (UVCB Substance) on the TSCA Inventory

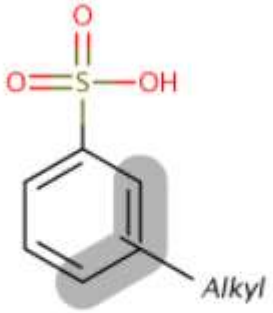
This paper is a compendium of information related to the broad class of chemical substances referred to as UVCBs for the Toxic Substances Control Act (TSCA) Chemical Substance Inventory. These chemical substances cannot be represented by unique structures and molecular formulas.

# “Markush Structures”

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

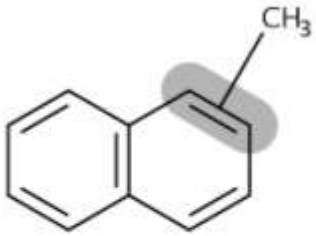



Xylenes
1330-20-7

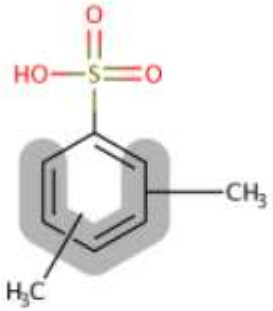



(C10-C16) Alkylbenzenesulfonic acid
68584-22-5

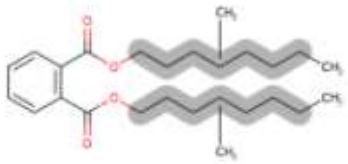

n-Nonylphenol
25154-52-3




Methylnaphthalene
1321-94-4

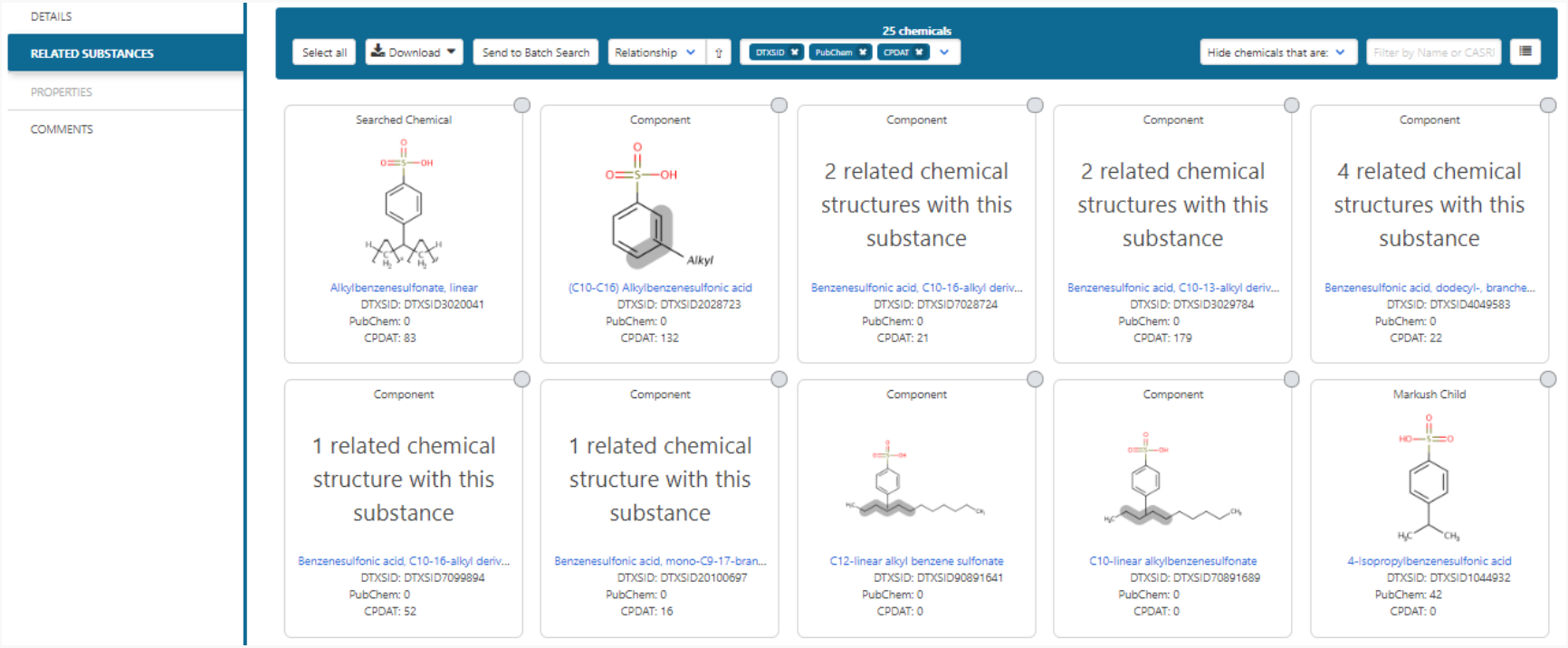



Sodium xylenesulfonate
1300-72-7



Diisononyl phthalate
28553-12-0

# UVCB: Complex Surfactants



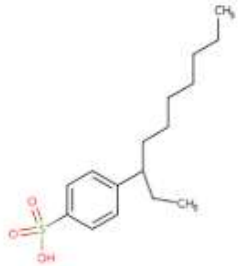
# UVCB: Complex Surfactants

## C10-linear alkylbenzenesulfonate

NOCAS_891689 | DTXSID70891689

Searched by DSSTox Substance Id.

Markush Child



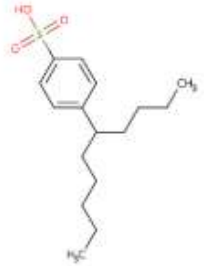
4-(Decan-3-yl)benzene-1-sulfonic acid

DTXSID: DTXSID20859618

PubChem: 0

CPDAT: 0

Markush Child



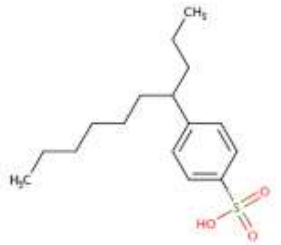
4-(decan-5-yl)benzene-1-sulfonic acid

DTXSID: DTXSID70881146

PubChem: 0

CPDAT: 0

Markush Child



4-(decan-4-yl)benzenesulfonic acid

DTXSID: DTXSID40891333

PubChem: 0

CPDAT: 0

EPA United States Environmental Protection Agency

# WORK IN PROGRESS

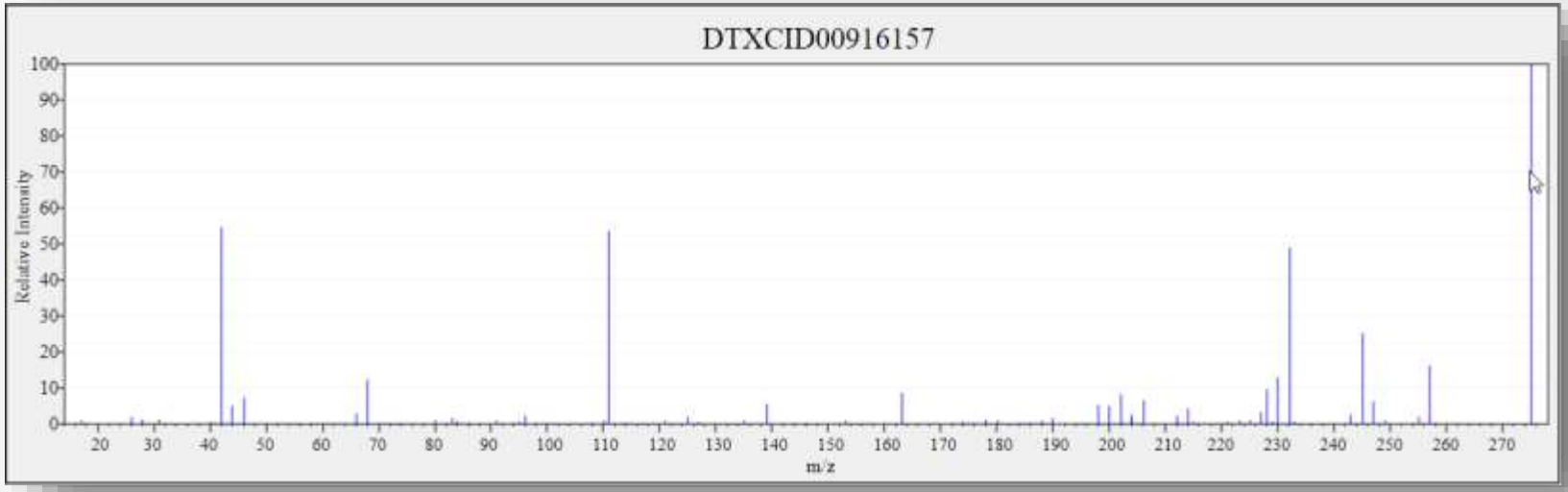
# Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database

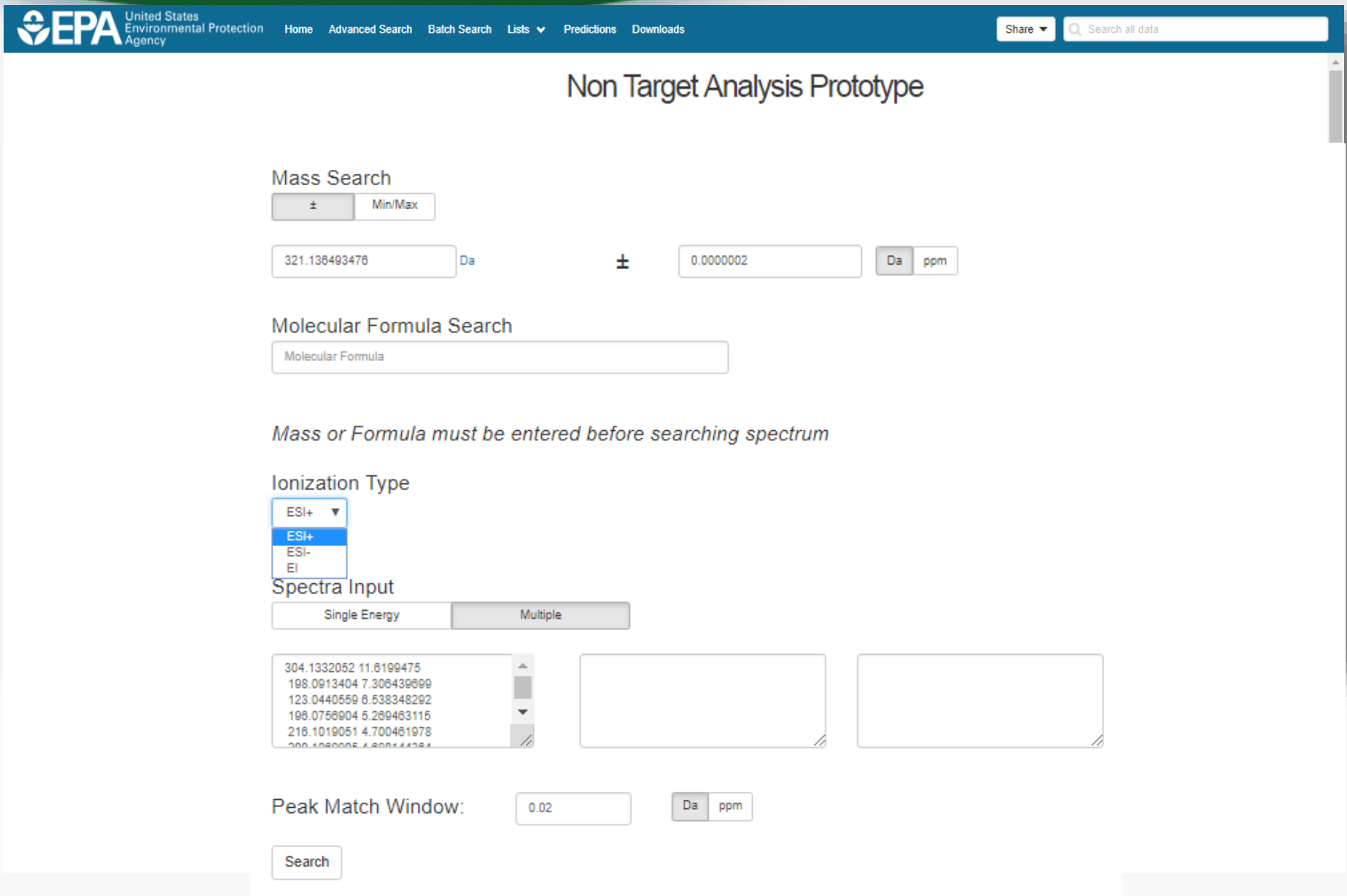
# Predicted Mass Spectra

http://cfmid.wishartlab.com/

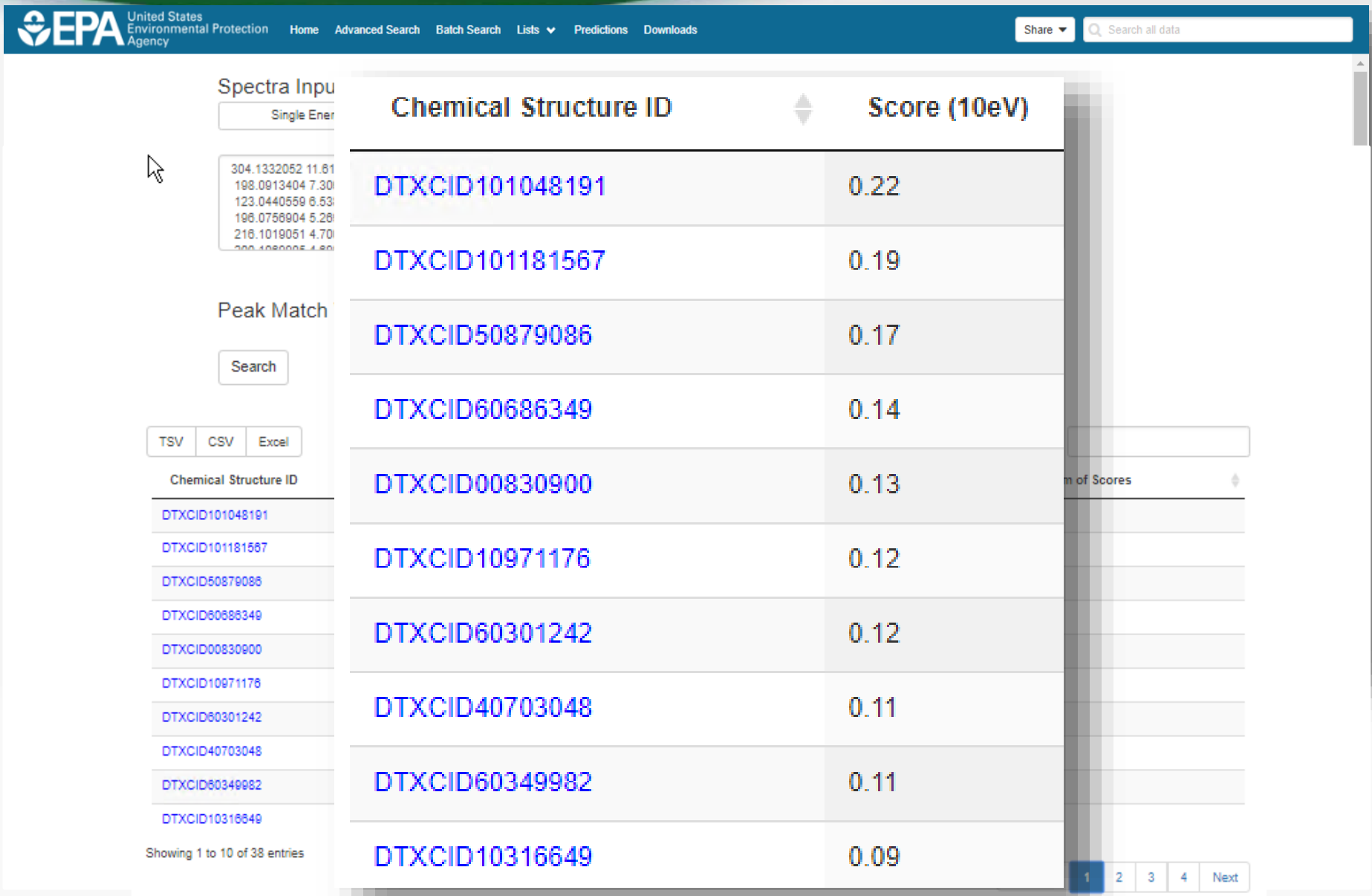

- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard



# Search Expt. vs. Predicted Spectra

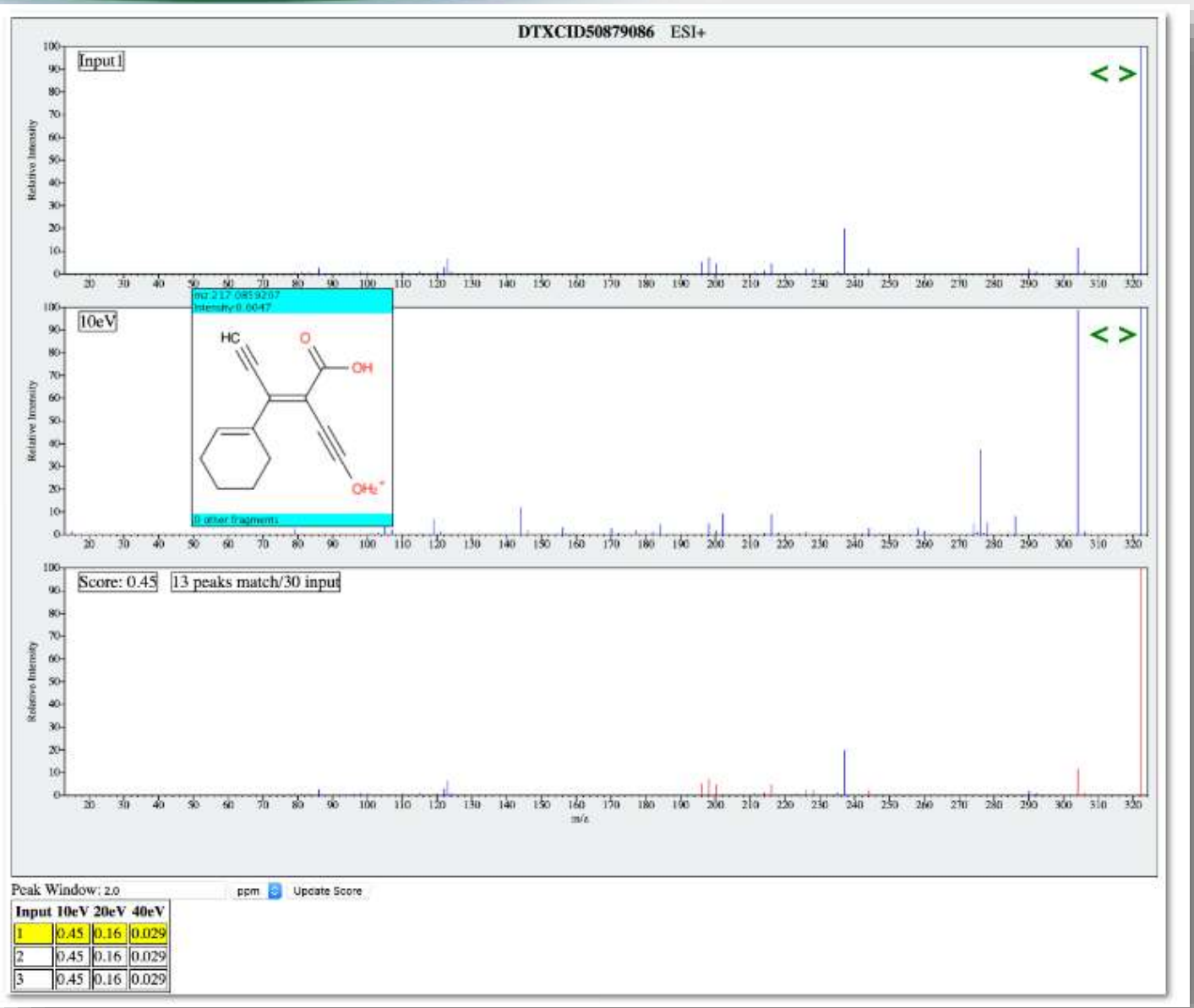


# Search Expt. vs. Predicted Spectra




| 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



# Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction

# Moving to Relative Retention Times

ELSEVIER

Journal of Hazardous Materials

Volume 363, 5 February 2019, Pages 277-285


## Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants

Reza Aalizadeh, Maria-Christina Nika, Nikolaos S. Thomaidis

Show more

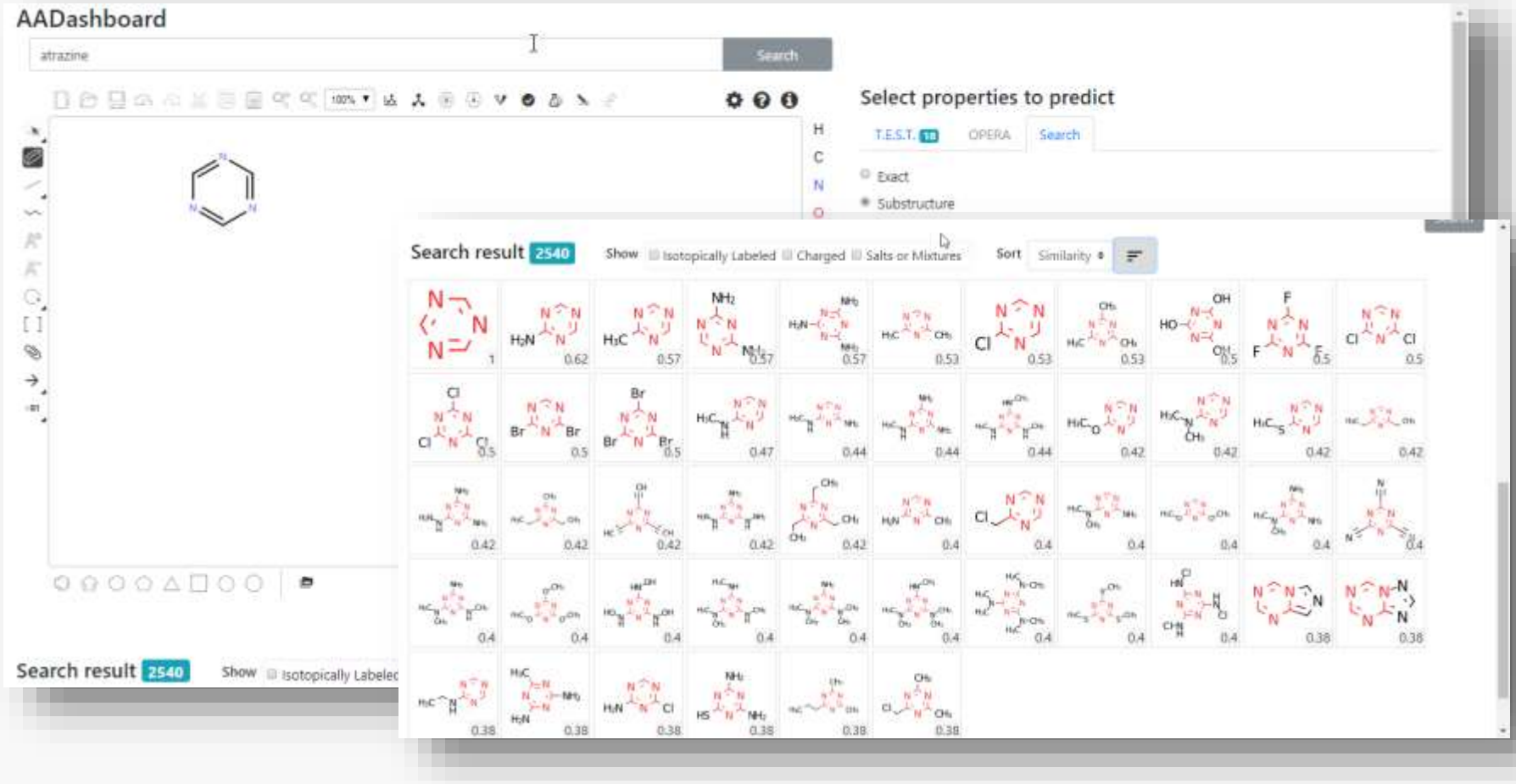
https://doi.org/10.1016/j.jhazmat.2018.09.047

Get rights and content

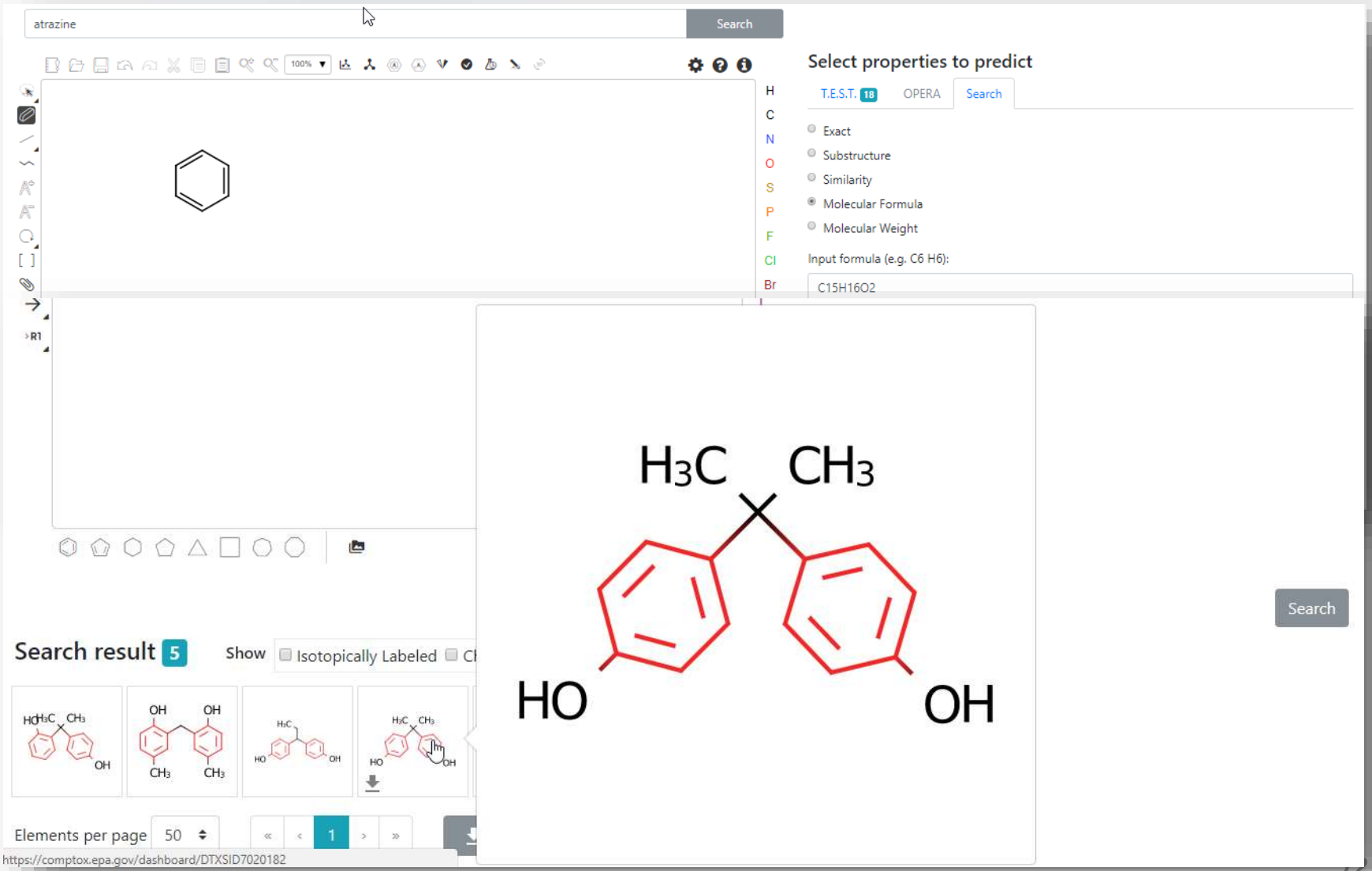
# Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search

# Prototype Development



# Prototype Development



# Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search
- Integration of predicted ion mobility data

# Collision Cross Section Prediction

EPA United States Environmental Protection Agency

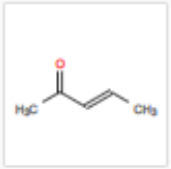
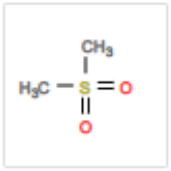
## PNNL Collision Cross Section Database

Pacific Northwest NATIONAL LABORATORY

Search

Showing 1 to 25 of 1000 rows 25 rows per page

‹ 1 2 3 4 5 ... 40 ›

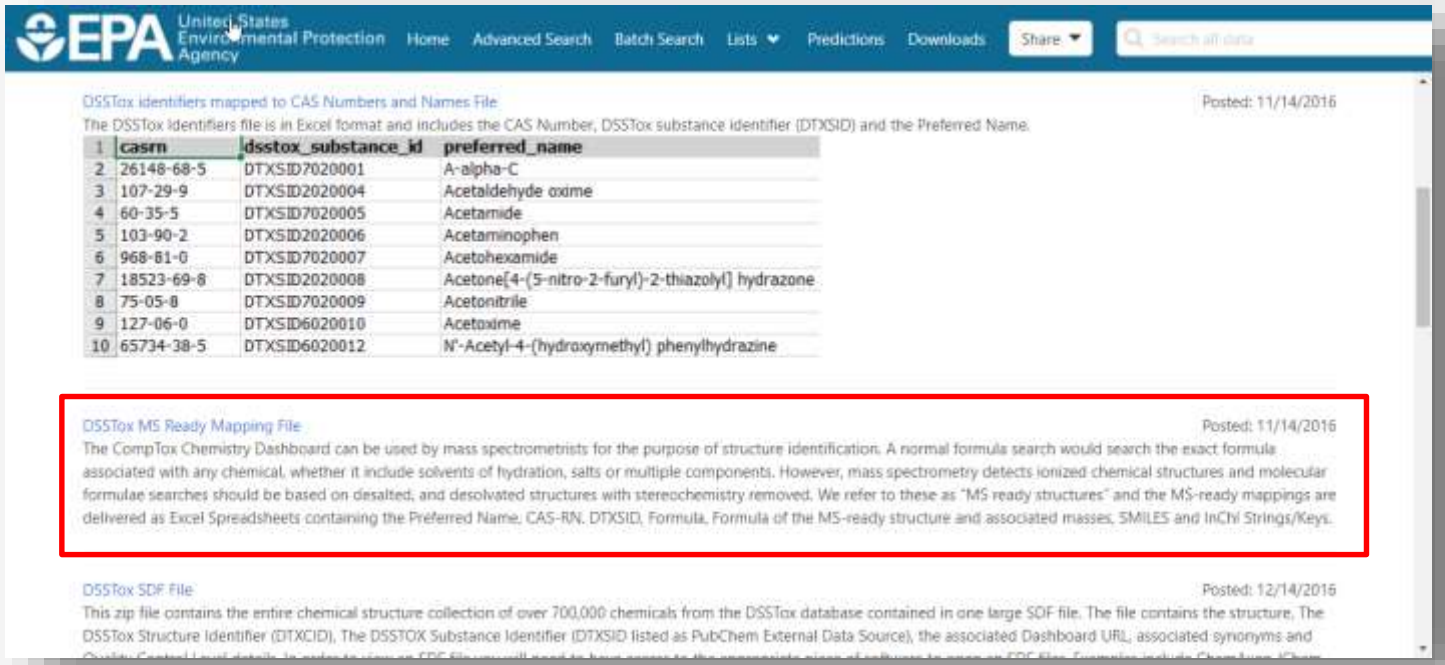
| Chemical | SMILES | InChI | Formula | Mass | CCS (Å$^2$) |
|---|---|---|---|---|---|
|  (3E)-pent-3-en-2-one | * | * | $C_5H_8O$ | 84.0575 | [M-H]$^-$ 112.1 ISiCLE Lite v0.1.0<br>[M+Na]$^+$ 112.6 ISiCLE Lite v0.1.0<br>[M+H]$^+$ 113.1 ISiCLE Lite v0.1.0 |
|  Dimethyl sulfone | * | * | $C_2H_6O_2S$ |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

# Work in Progress

- CFM-ID
  - Viewing and Downloading pre-predicted spectra
  - Search spectra against the database
- Retention Time Index Prediction
- Structure/substructure/similarity search
- Integration of predicted ion mobility data
- Access to API and web services for programmatic access

# API services and Open Data

- Groups waiting on our API and web services
- Mass Spec companies instrument integration
- Release will be in iterations but for now our data are available



# SIDE EFFECTS OF SHARING 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**

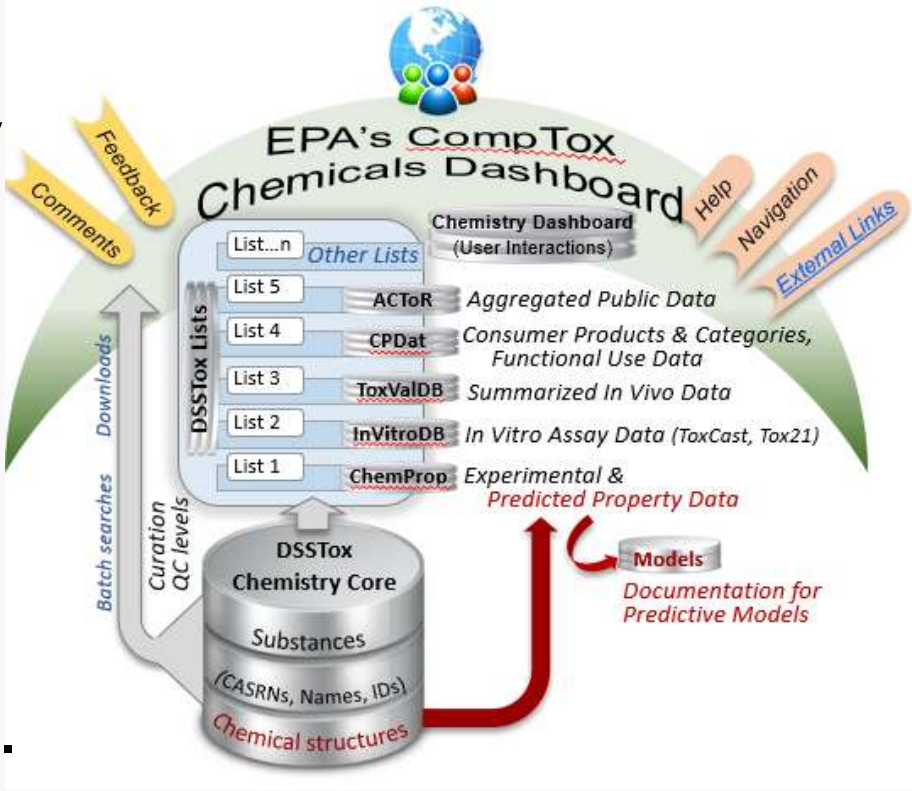
| | | | |
|---|---|---|---|
| **Wastewater Suspect List based on Swedish Product Data** | Wastewater Suspect List **Original File with Mapped DTXSIDs** (12/02/2019) | KEMIWWSUS **InChIKeys** (12/02/2019) | A prioritized list of 1,123 substances relevant for wastewater based on Swedish product registry data, including scores. Provided by Stellan Fischer, KEMI. |
| **Algal toxins list from CompTox** | ALGALTOX **XLSX**, **CSV** (14/02/2019)<br>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)<br>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)<br>CompTox **HDXNOEX List**<br>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)<br>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)<br>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)<br>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

# 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
- Future releases will offer even more utility
- We are committed to open API development with time..



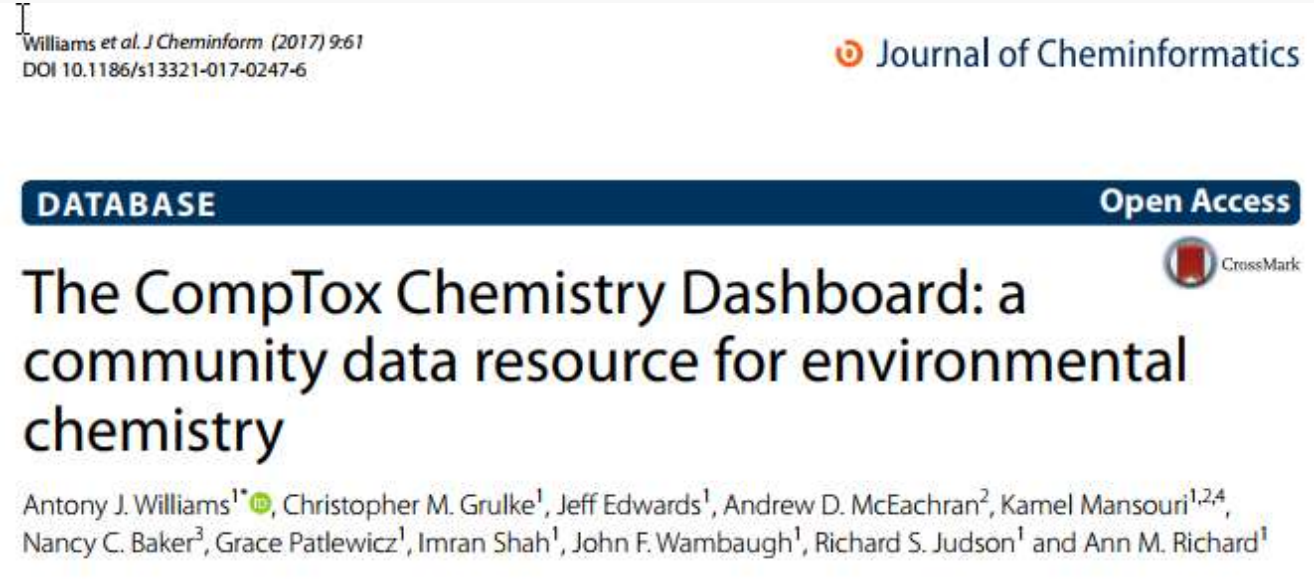
# Acknowledgements


**EPA-RTP**

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

# Contact

**Antony Williams**
NCCT, 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