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

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

Spring 2019
ACS Spring Meeting, Orlando
### Suspect Screening and Non-Targeted Analysis Workflows

#### 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
- **Green** = Informatics & Web Services
- **Purple** = Mathematical & QSPR Modeling
CompTox Chemicals Dashboard
https://comptox.epa.gov/dashboard

875 Thousand Chemicals

Latest News

Journal of Cheminformatics article regarding “MS-Ready structures”
March 9th, 2019 at 1:09:45 PM
A recent article describes “MS-Ready structures”, what they are, how they are generated and details regarding the benefits of these structures in navigating structure relationships across the dashboard. The article is published in the Journal of Cheminformatics here.
Basic Search

Search results for "Bisphenol" include:

- Bisphenol A (DTXSID7020182)
- Bisphenol A bis(2-hydroxyethyl ether) diacrylate (DTXSID6066991)
- Bisphenol A bis(2-hydroxyethyl ether) dimethacrylate (DTXSID1066992)
- Bisphenol A bis(2-hydroxypropyl) ether (DTXSID8051592)
- Bisphenol A carbonate polymer (DTXSID6027840)
- Bisphenol A diglycidyl ether (DTXSID6024624)
- Bisphenol A glycidyl methacrylate (DTXSID7044841)
Sources of Exposure to Chemicals
Identifiers to Support Searches

**Bisphenol A**

80-05-7 | DTXSID7020182

*Searched by Approved Name.*

### Synonyms

<table>
<thead>
<tr>
<th>Synonym</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bisphenol A</td>
<td>Valid</td>
</tr>
<tr>
<td>4,4'-(Propane-2,2-diylyldiphenol)</td>
<td>Valid</td>
</tr>
<tr>
<td>Phenol, 4,4'-(1-methylethylene)bis-</td>
<td>Valid</td>
</tr>
<tr>
<td>80-05-7 <em>Active CAS-RN</em></td>
<td>Valid</td>
</tr>
<tr>
<td><strong>BPA</strong></td>
<td>Valid</td>
</tr>
<tr>
<td>4,4'-Propane-2,2-diylphenol</td>
<td>Valid</td>
</tr>
<tr>
<td>Phenol, 4,4'-(1-methylethylene)bis-</td>
<td>Valid</td>
</tr>
<tr>
<td>4-05-50-05717 <em>Biotulin Registry Number</em></td>
<td>Biotulin</td>
</tr>
<tr>
<td>(4,4'-Dihydroxydiphenyl)(dimethyl)methane</td>
<td>Good</td>
</tr>
<tr>
<td>2,2-Bis(4-hydroxyphenyl) propane</td>
<td>Good</td>
</tr>
<tr>
<td>2,2-Bis(4-hydroxyphenyl)propane</td>
<td>Good</td>
</tr>
<tr>
<td>2,2-Bis(4-hydroxyphenyl)-propylene</td>
<td>Good</td>
</tr>
<tr>
<td>2,2-Bis(4-hydroxyphenyl)propylene</td>
<td>Good</td>
</tr>
<tr>
<td>2,2-Di(4-Hydroxyphenyl)Propane</td>
<td>Good</td>
</tr>
</tbody>
</table>
Bisphenol A
80-05-7 | DTXSID7020182
Search by Approved Name.

General
- EPA Substances Registry Service
- Household Products Database
- Chemical Entities of Biological Interest (ChEBI)
- PubChem
- ChemSpider
- CPCat
- DrugBank
- HMDB
- NTP
- Wikipedia
- MSDS Lookup
- CHEMBL
- Chemical Vendors
- CAS EPA Office of Environmental Health Hazard Assessment
- NIOSH Chemical Safety Cards
- Toxicology Data
- ACS Reactant Chemicals
- Wolfram Alpha
- SciFinder
- CAS REGISTRY Number

Toxicology
- ACToR
- DrugPortal
- TOPS
- ChemView
- CTD
- eChemPortal
- GastroTox
- NIOSH
- ToxCast Dashboard 2
- ToxRefDB
- International Toxicity Estimates for Risk
- ATSDR Toxic Substances Portal
- Supporting Chemical Data Matrix
- NIOSH CROH Values
- ACToR PDF Report
- Toxicology Release Inventory
- CREST
- National Air Toxics Assessment

Publications
- Google Scholar
- PubMed
- PubMed Central
- Environmental Health Perspectives
- NLM BioLJ
- National Toxicology Program
- Google Books
- Google Scholar
- Google Patents
- AHRFWEB
- EMBL

Analytical
- EPA OAE
- NLM: National Environmental Methods Index
- TOXNET
- RSC Analytical Abstracts
- Tox21 Analytical Data
- NLM: Mass Spectrometry North America
- msCloud
- NIST: FIES IR Spectrum
- NIST: FIES MS Spectrum

Prediction
- 2D NMR HSQC NMR Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRTRP Predictor
- LEEDR
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
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
## Add Other Data of Interest

### Chemical Identifiers
- DTXSID
- Chemical Name
- DTXCID
- CAS-RN
- InChIKey
- IUPAC Name

### Structures
- Mol File
- SMILES
- InChI String
- MS-Ready SMILES
- QSAR-Ready SMILES

### Intrinsic And Predicted Properties
- Molecular Formula
- Average Mass
- Monoisotopic Mass
- TEST Model Predictions
- OPERA Model Predictions

### Table

<table>
<thead>
<tr>
<th>INPUT</th>
<th>DTXSID</th>
<th>CASRN</th>
<th>MOLECULAR FORMULA</th>
<th>MONOISOTOPIC</th>
<th>MS READY_SMILES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buprenorphine</td>
<td>DTXSID202:</td>
<td>52485-79-7</td>
<td>C29H41NO4</td>
<td>467.3035588</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Codeine</td>
<td>DTXSID202:</td>
<td>76-57-3</td>
<td>C18H21NO3</td>
<td>299.1521435</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Dextrometh</td>
<td>DTXSID302:</td>
<td>125-71-3</td>
<td>C18H25NO</td>
<td>271.1936144</td>
<td>[H]C12CC3=C(C=</td>
</tr>
<tr>
<td>Dihydrocode</td>
<td>DTXSID502:</td>
<td>125-28-0</td>
<td>C18H23NO3</td>
<td>301.1677936</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Dihydromor</td>
<td>DTXSID704:</td>
<td>509-60-4</td>
<td>C17H21NO3</td>
<td>287.1521435</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Ethylmorph</td>
<td>DTXSID104:</td>
<td>76-58-4</td>
<td>C19H23NO3</td>
<td>313.1677936</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Fentanyl</td>
<td>DTXSID902:</td>
<td>437-38-7</td>
<td>C22H28N2O</td>
<td>336.2201635</td>
<td>C(CC(=O)N(C(=C</td>
</tr>
<tr>
<td>Heroin</td>
<td>DTXSID604:</td>
<td>561-27-3</td>
<td>C21H23NO5</td>
<td>369.1576228</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Hydrocodone</td>
<td>DTXSID802:</td>
<td>125-29-1</td>
<td>C18H21NO3</td>
<td>299.1521435</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Hydromorphone</td>
<td>DTXSID802:</td>
<td>466-99-9</td>
<td>C17H19NO3</td>
<td>285.1364935</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Ketamine</td>
<td>DTXSID802:</td>
<td>6740-88-1</td>
<td>C13H16CINO</td>
<td>237.0920418</td>
<td>CNC1(CC=CC1=C(C</td>
</tr>
<tr>
<td>Meperidine</td>
<td>DTXSID902:</td>
<td>57-42-1</td>
<td>C15H21NO2</td>
<td>247.1572289</td>
<td>CCC=C(=O)C1(C(C</td>
</tr>
<tr>
<td>Methadone</td>
<td>DTXSID702:</td>
<td>76-99-3</td>
<td>C21H27NO</td>
<td>309.2092645</td>
<td>CCC(=O)C(C(C(C</td>
</tr>
<tr>
<td>Morphine</td>
<td>DTXSID902:</td>
<td>57-27-2</td>
<td>C17H19NO3</td>
<td>285.1364935</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Morphinone</td>
<td>DTXSID501:</td>
<td>467-02-7</td>
<td>C17H17NO3</td>
<td>283.1208434</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Naloxone</td>
<td>DTXSID802:</td>
<td>465-65-6</td>
<td>C19H21NO4</td>
<td>327.1470582</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Naltrexone</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Oxycodone</td>
<td>DTXSID502:</td>
<td>76-42-6</td>
<td>C18H21NO4</td>
<td>315.1470582</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Oxymorphone</td>
<td>DTXSID502:</td>
<td>76-41-5</td>
<td>C17H19NO4</td>
<td>301.1314081</td>
<td>[H]C12CC3=C4C</td>
</tr>
<tr>
<td>Propoxyphene</td>
<td>DTXSID102:</td>
<td>469-62-5</td>
<td>C22H29NO2</td>
<td>339.2198292</td>
<td>CCC(=O)OC(CC1=C</td>
</tr>
<tr>
<td>Sufentanil</td>
<td>DTXSID602:</td>
<td>56030-54-7</td>
<td>C22H30N2O2S</td>
<td>386.2072994</td>
<td>CCC(=O)N(C(=CC(=</td>
</tr>
<tr>
<td>Tramadol</td>
<td>DTXSID908:</td>
<td>27203-92-5</td>
<td>C16H25NO2</td>
<td>263.188529</td>
<td>CCC=CC=CC(=C</td>
</tr>
</tbody>
</table>
CHEMICAL LISTS
<table>
<thead>
<tr>
<th>List Acronym</th>
<th>List Name</th>
<th>Last Updated</th>
<th>Number of Chemicals</th>
<th>List Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HDXEXCH</td>
<td>MASSPECDB: Hydrogen Deuterium Exchange Standard Set · Under HDX Conditions</td>
<td>2018-11-07</td>
<td>592</td>
<td>Observed species (deuterated and undeuterated) from the HDXNOEX list under hydrogen deuterium exchange conditions (Ruttkies, Schymanski et al. in prep.)</td>
</tr>
<tr>
<td>HDXNOEX</td>
<td>MASSPECDB: Hydrogen Deuterium Exchange Standard Set · NO Exchange</td>
<td>2018-11-07</td>
<td>765</td>
<td>Environmental standard set used to investigate hydrogen deuterium exchange in small molecule high resolution mass spectrometry (Ruttkies, Schymanski et al. in prep.)</td>
</tr>
<tr>
<td>MASSBANKEUSP</td>
<td>MASSPECDB: MassBank.EU Collection: Special Cases</td>
<td>2017-07-16</td>
<td>263</td>
<td>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.</td>
</tr>
<tr>
<td>MYCOTOXINS</td>
<td>MASSPECDB: Mycotoxins from MassBank.EU</td>
<td>2017-03-02</td>
<td>88</td>
<td>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</td>
</tr>
<tr>
<td>List Acronym</td>
<td>List Name</td>
<td>Last Updated</td>
<td>Number of Chemicals</td>
<td>List Description</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------------------------------------</td>
<td>-------------</td>
<td>---------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>EPAPFAS75S1</td>
<td>PFAS</td>
<td>EPA: List of 75 Test Samples (Set 1)</td>
<td>2018-06-29</td>
<td>74</td>
</tr>
<tr>
<td>EPAPFAS75S2</td>
<td>PFAS</td>
<td>EPA: List of 75 Test Samples (Set 2)</td>
<td>2019-02-21</td>
<td>75</td>
</tr>
<tr>
<td>EPAPFASI NSOL</td>
<td>PFAS</td>
<td>EPA: Chemical Inventory Insoluble in DMSO</td>
<td>2018-06-29</td>
<td>43</td>
</tr>
<tr>
<td>EPAPFASNV</td>
<td>PFAS</td>
<td>EPA: ToxCast Chemical Inventory</td>
<td>2018-06-29</td>
<td>430</td>
</tr>
<tr>
<td>EPAPFASRL</td>
<td>PFAS</td>
<td>EPA: Cross-Agency Research List</td>
<td>2017-11-16</td>
<td>199</td>
</tr>
<tr>
<td>PFASM AS T E R</td>
<td>PFAS Master List of PFAS Substances</td>
<td>2018-07-26</td>
<td>5061</td>
<td>PFASM AS T E R is a consolidated list of PFAS substances spanning and bounded by the below lists of current interest to researchers and regulators worldwide.</td>
</tr>
<tr>
<td>PFASOECD</td>
<td>PFAS: Listed in OECD Global Database</td>
<td>2018-05-16</td>
<td>4729</td>
<td>OECD released a New Comprehensive Global Database of Per- and Polyfluoroalkyl Substances, (PFASs) listing more than 4700 new PFAS</td>
</tr>
<tr>
<td>PFASTRIER</td>
<td>PFAS Community-Compiled List (Trier et al., 2015)</td>
<td>2017-07-16</td>
<td>597</td>
<td>PFASTRIER community-compiled public listing of PFAS (Trier et al., 2015)</td>
</tr>
</tbody>
</table>
EPAHFR: Hydraulic Fracturing

WATER|EPA; Chemicals associated with hydraulic fracturing

**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 Fracturing 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](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

<table>
<thead>
<tr>
<th>INPUT</th>
<th>DTXSID</th>
<th>MASSBANKREF</th>
<th>NEMILIST</th>
<th>WRTMSD</th>
<th>NORMANPRI</th>
<th>SUSDAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buprenorph</td>
<td>DTXSID202</td>
<td>-</td>
<td>Y</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Codeine</td>
<td>DTXSID202</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Dextrometh</td>
<td>DTXSID302</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Dihydrocod</td>
<td>DTXSID502</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Dihydromor</td>
<td>DTXSID704</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Ethylmorph</td>
<td>DTXSID104</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Fentanyl</td>
<td>DTXSID902</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Heroin</td>
<td>DTXSID604</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Hydrocodor</td>
<td>DTXSID802</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Hydromor</td>
<td>DTXSID802</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Ketamine</td>
<td>DTXSID802</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Meperidine</td>
<td>DTXSID902</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Methadone</td>
<td>DTXSID702</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Morphine</td>
<td>DTXSID902</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Morphinone</td>
<td>DTXSID501</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Naloxone</td>
<td>DTXSID802</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Naltriben</td>
<td></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>Oxycodone</td>
<td>DTXSID502</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Oxymorpho</td>
<td>DTXSID502</td>
<td>-</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Propoxyph</td>
<td>DTXSID102</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Sufentanil</td>
<td>DTXSID602</td>
<td>Y</td>
<td>Y</td>
<td>-</td>
<td></td>
<td>Y</td>
</tr>
<tr>
<td>Tramadol</td>
<td>DTXSID908</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td></td>
<td>Y</td>
</tr>
</tbody>
</table>
“MS-READY” STRUCTURES
Open Science for Identifying “Known Unknown” Chemicals

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

Mass Search

± Min/Max

Adduct  All Adducts
Neutral  Choose adduct from dropdown

191.131  Da ± 5

Search
MS-Ready Structures for Formula Search

Molecular Formula Search

- MS Ready Formula
- Exact Formula

Formula

Please use the format of the following example: C6H8O2 or C6H(8-10)O(0-2)

Search
This search is based on what we refer to as "Mass Spec (MS) Ready" structures. All chemicals within the database are treated in a manner such that all are desalted, mixtures are separated, and stereochemistry is removed as Mass Spectrometry detects the major components of a salt or mixture and is insensitive to stereochemistry. As an example, a search for the monoisotopic mass of phenol will return phenol, sodium phenolate and calcium phenoxide. See the publication for more details: https://doi.org/10.1186/s13321-018-0299-2.
**MS-Ready Mappings**

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

![Chemical structures and metadata](image_url)
MS-Ready Mappings

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

Step Five: Choose Data Fields to Download

Please enter one identifier per line

Select Input Type(s)

- Identifiers
- Chemical Name
- CASRN
- InChIKey
- DSSTox Substance ID
- DSSTox Compound ID
- InChIKey Skeleton
- MS-Ready Formula(s)
- Exact Formula(s)
- Monoisotopic Mass
- Display All Chem

Enter Identifiers to Search (searches should be limited to < 5000 identifiers)

41.0265
56.02621
53.0265
58.0418
93.0578
113.9639
151.8754
69.9377
77.9872

This search is based on what we refer to as "Mass Spec (MS) Ready" structures. All chemicals within the database are treated in a manner such that all are desalted, mixtures are separated, and stereochemistry is removed as Mass Spectrometry detects the major components of a salt or mixture and is insensitive to stereochemistry. As an example, a search for the monoisotopic mass of phenol will return phenol, sodium phenolate and calcium phenoxide. See the publication for more details: https://doi.org/10.1186/s13321-018-0299-2.
## Searching batches using MS-Ready Formula (or mass) searching

<table>
<thead>
<tr>
<th>A</th>
<th>DTXSID</th>
<th>CASRN</th>
<th>PREFERRED NAME</th>
<th>DTXSID</th>
<th>CASRN</th>
<th>PREFERRED NAME</th>
<th>DTXSID</th>
<th>CASRN</th>
<th>PREFERRED NAME</th>
<th>DTXSID</th>
<th>CASRN</th>
<th>PREFERRED NAME</th>
<th>DTXSID</th>
<th>CASRN</th>
<th>PREFERRED NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>C14H22N2O3</td>
<td>29122-68-7</td>
<td>Atenolol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>C14H22N2O3</td>
<td>6673-35-4</td>
<td>Practolol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>C14H22N2O3</td>
<td>40488544</td>
<td>Bucolone</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>C14H22N2O3</td>
<td>1045407</td>
<td>Trimeprazine dihydrochloride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>C14H22N2O3</td>
<td>6715-13-0</td>
<td>R+(-)-Atenolol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>C14H22N2O3</td>
<td>2048531</td>
<td>Trimeprazine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>C14H22N2O3</td>
<td>10239405</td>
<td>Esatenol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>C14H22N2O3</td>
<td>50200634</td>
<td>N-Diethylaminomethyl)-2-(4-hydroxyphenoxoyacetamide</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>C14H22N2O3</td>
<td>4020111</td>
<td>dl-Atenolol hydrochloride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>C14H22N2O3</td>
<td>1068693</td>
<td>Benzenamidine, 2,5-diethoxy-4-(4-morpholinyl)-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>C18H34N2O6S</td>
<td>3023216</td>
<td>Lincomycin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>C18H34N2O6S</td>
<td>7047803</td>
<td>Lincomycin hydrochloride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>C18H34N2O6S</td>
<td>20849438</td>
<td>PUBCHEM 74132749</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>C10H12N2O</td>
<td>1047576</td>
<td>Cotinine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>C10H12N2O</td>
<td>8075330</td>
<td>Serotonin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>C10H12N2O</td>
<td>8044412</td>
<td>4-Methyl-1-phenylpyrazolidin-3-one</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>C10H12N2O</td>
<td>80165166</td>
<td>Serotonin hydrochloride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>C10H12N2O</td>
<td>2048870</td>
<td>4R,5S)-4-methyl-5-phenyl-4,5-dihydro-1,3-oxazol-2-amine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>C10H12N2O</td>
<td>10196105</td>
<td>6-Hydroxytryptamine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>C10H12N2O</td>
<td>90185693</td>
<td>1,4,5,6-Tetrahydro-5-phenoxypridimine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>C10H12N2O</td>
<td>40178777</td>
<td>2-Benzimidazolepropanol</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>C10H12N2O</td>
<td>80157026</td>
<td>N-Cyclopropyl-N'-phenylurea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>C10H12N2O</td>
<td>30205607</td>
<td>4-Hydroxytryptamine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>C14H18N4O3</td>
<td>5023900</td>
<td>Benomyl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>C14H18N4O3</td>
<td>3023712</td>
<td>Trimethoprim</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>C14H18N4O3</td>
<td>40209671</td>
<td>Trimethoprim hydrochloride</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>C14H18N4O3</td>
<td>70204210</td>
<td>Benzenemethanol, 4-(2,4-diamino-5-pyrimidine)methyl)-2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>C14H18N4O3</td>
<td>12052671</td>
<td>6-Methoxy-4-(3-((N,N-dimethylamino)propylamino)-5-8-quini</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>C14H18N4O3</td>
<td>30213742</td>
<td>1H,1,2,4-Benzotriazine-3-carboxylic acid, 4,5-dihydro-4-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>C14H18N4O3</td>
<td>30219608</td>
<td>2,4-Pyrimidinediameine, 5-((3,4,5-trimethoxyphenyl)methyl)-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>C14H18N4O3</td>
<td>20241165</td>
<td>L-Aspartic acid, compound with 5-((3,4,5-trimethoxyphenyl)methyl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>C14H18N4O3</td>
<td>80241166</td>
<td>L-Glutamic acid, compound with 5-((3,4,5-trimethoxyphenyl)methyl</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>C14H18N4O3</td>
<td>20243761</td>
<td>1H-Pyridine(2,3-e)-1,4-diazepine, 2,3,5-trione, 4-2(diethylamin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>C12H11N7</td>
<td>6021373</td>
<td>Triamterene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>C12H11N7</td>
<td>60204446</td>
<td>Ampyrimidines</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>C12H11N7</td>
<td>5056421</td>
<td>Benzenemine, 4-azido-N-(4-azidophenyl)-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>C12H11N7</td>
<td>50848255</td>
<td>Sulfuric acid-5-phenylteridine-2,4,7-trimine (1/1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>C12H11N7</td>
<td>50575263</td>
<td>(1E)-N-Phenyl-1,2-bis(1H,1,2,4-triazol-1-yl)ethan-1-imine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>C8H9N2O</td>
<td>20200006</td>
<td>Acetaminophen</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>C8H9N2O</td>
<td>60256355</td>
<td>Methyl 2-aminobenzoate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Notes
- DTXSID: Database record ID
- CASRN: Chemical Abstracts Service Registry Number
- Preferred Name: Chemical name as preferred by the database
- MOL FORMULA: Molecular formula
- MONOISOTOTPIC MASS: Monoisotopic mass
- DATA SOURCES: Sources of data for the chemical entry
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

Intrinsic Properties

- **Molecular Formula:** C_{17}H_{19}NO_{3}
- **Average Mass:** 285.343 g/mol
- **Monoisotopic Mass:** 265.136493 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

Searched with a similarity threshold of 0.8

178 chemicals

Select all  Download  Send to Batch Search  Similarity  $  DTXSID  CASRN  TOXCAST  Similarity  Hide chemicals that are:  Filter by Name or CASRN
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
- 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

<table>
<thead>
<tr>
<th>Child (infant through adolescent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dust and Exposure</td>
</tr>
<tr>
<td>Food and Exposure</td>
</tr>
<tr>
<td>Water and Exposure</td>
</tr>
<tr>
<td>Algae</td>
</tr>
<tr>
<td>Disaster / Emergency</td>
</tr>
</tbody>
</table>

Optionally, edit the query before retrieving.

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

37 of 37 articles loaded...

To find articles quickly, enter terms to sift abstracts.

- wastewater
- Spectrometry
- EPA

<table>
<thead>
<tr>
<th>PMID</th>
<th>Year</th>
<th>Title</th>
<th>Authors</th>
<th>Journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>25874731</td>
<td>2017</td>
<td>Simultaneous analysis of opioid analgesics and their metabolites...</td>
<td>Krizman, Matasic, Kostarijevski, Aliev, Terzic</td>
<td>Journal of chromatography: A</td>
</tr>
<tr>
<td>25769872</td>
<td>2015</td>
<td>Evaluating external contamination of polybrominated...</td>
<td>Picco, Aleksa, Carnevale, Kapur, Goodyer, Koren</td>
<td>Therapeutic drug monitoring</td>
</tr>
<tr>
<td>22544551</td>
<td>2012</td>
<td>Spatial distribution of illicit drugs in surface waters...</td>
<td>Vazquez-Rojas, Andreu, Blasco, Morillas, Picó</td>
<td>Environmental science and pollution research information</td>
</tr>
<tr>
<td>20861487</td>
<td>2010</td>
<td>Analysis of illicit and illicit drugs in waste, surface and...</td>
<td>Bersel, Brenneisen, Mathieu</td>
<td>Chemosphere</td>
</tr>
<tr>
<td>17935751</td>
<td>2007</td>
<td>Illicit drugs: a novel group of environmental contaminants...</td>
<td>Zuccato, Castiglioni, Ragnati, Chiabrando, Grassi;</td>
<td>Water research</td>
</tr>
<tr>
<td>17607391</td>
<td>2007</td>
<td>Using environmental analytical data to estimate lev...</td>
<td>Bones, Thomas, Paul</td>
<td>Journal of environmental monitoring: JEM</td>
</tr>
<tr>
<td>17469084</td>
<td>2006</td>
<td>Simultaneous determination of psychoactive drugs...</td>
<td>Hummel, Löffler, Fink, Ternes</td>
<td>Environmental science &amp; technology</td>
</tr>
<tr>
<td>30883159</td>
<td>2018</td>
<td>Assessment of drugs of abuse in a wastewater treatment...</td>
<td>Kumar, Techarke, O'Brien, Mueller, Wilkins, Padhre</td>
<td>The Science of the total environment</td>
</tr>
<tr>
<td>30484821</td>
<td>2015</td>
<td>Effect of enriched environment during adolescence...</td>
<td>Mohammadnejad, Najat, Miladi-Gorji</td>
<td>Developmental psychology</td>
</tr>
<tr>
<td>28574368</td>
<td>2015</td>
<td>Estimation of the consumption of illicit drugs during...</td>
<td>Poppe, Hammond-Weinberger, Subedi</td>
<td>The Science of the total environment</td>
</tr>
<tr>
<td>28577791</td>
<td>2017</td>
<td>Evaluation of in-solution transformation of selected illic...</td>
<td>Gei, Banko, Li, Jiang, Lai, Mueller, Thai</td>
<td>The Science of the total environment</td>
</tr>
<tr>
<td>28472597</td>
<td>2017</td>
<td>Occurrence and fate of illicit drugs and pharmaceuticals...</td>
<td>Cavailler, Ruiz, Ibañez, Emiko, Hennedy; d'Elia; Gracioli, M.</td>
<td>The Science of the total environment</td>
</tr>
<tr>
<td>27746911</td>
<td>2015</td>
<td>Effects of voluntary exercise on the viability, proliferation...</td>
<td>Hoyer, Salehi, Zarbakhsh, Bordej, Miladi-Gorji</td>
<td>Neuroscience letters</td>
</tr>
<tr>
<td>27261079</td>
<td>2016</td>
<td>Genotoxic effects induced by the exposure to an...</td>
<td>Perlini, Magni, Castiglioni, Binelli</td>
<td>Ecotoxicology and environmental safety</td>
</tr>
<tr>
<td>27179320</td>
<td>2016</td>
<td>Temporal trends in drug use in Adelaide, South Aus...</td>
<td>Toscharke, Chen, Gerber, White</td>
<td>The Science of the total environment</td>
</tr>
</tbody>
</table>
Example Online Resources for MS

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

- **PubChem**

- **MassBank**

- **METLIN**

- **GNPS**

- **ChemSpider**
  - Search and share chemistry
DO WE REALLY NEED ANOTHER DATABASE?
Is a bigger database better?

ChemSpider was 26 million chemicals then

Much BIGGER today

Is bigger better??
Comparing Search Performance

- Dashboard content was 720k chemicals
- Only 3% of ChemSpider size
- What was the comparison in performance?
SAME dataset for comparison

<table>
<thead>
<tr>
<th>Compound class</th>
<th>Number in class</th>
<th>Average rank</th>
<th>Number of compounds in each position rank-ordered</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>#1</td>
</tr>
<tr>
<td>Pharmaceutical drug</td>
<td>72</td>
<td>1.4</td>
<td>55</td>
</tr>
<tr>
<td>Industrial chemicals</td>
<td>42</td>
<td>5.5</td>
<td>28</td>
</tr>
<tr>
<td>Personal care products</td>
<td>8</td>
<td>6.1</td>
<td>3</td>
</tr>
<tr>
<td>Steroid hormones</td>
<td>7</td>
<td>1.0</td>
<td>7</td>
</tr>
<tr>
<td>Perfluorochemicals</td>
<td>6</td>
<td>1.2</td>
<td>5</td>
</tr>
<tr>
<td>Pesticides</td>
<td>12</td>
<td>2.3</td>
<td>6</td>
</tr>
<tr>
<td>Veterinary drugs</td>
<td>3</td>
<td>1.3</td>
<td>2</td>
</tr>
<tr>
<td>Dyes</td>
<td>2</td>
<td>1.0</td>
<td>2</td>
</tr>
<tr>
<td>Food product/natural compounds</td>
<td>4</td>
<td>3.8</td>
<td>2</td>
</tr>
<tr>
<td>Illicit drugs</td>
<td>2</td>
<td>2.0</td>
<td>1</td>
</tr>
<tr>
<td>Misc. molecules</td>
<td>3 &quot;</td>
<td>1.3</td>
<td>2</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Average rank (±SD)</th>
<th>Number in each position rank-ordered</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#1</td>
<td>#2</td>
</tr>
<tr>
<td>Mass-based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dashboard</td>
<td>1.2 ± 0.7</td>
<td>77a</td>
<td>5</td>
</tr>
<tr>
<td>ChemSpider</td>
<td>2.2 ± 6.1b</td>
<td>68</td>
<td>8</td>
</tr>
<tr>
<td>Formula-based</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dashboard</td>
<td>1.1 ± 0.4</td>
<td>78a</td>
<td>8</td>
</tr>
<tr>
<td>ChemSpider</td>
<td>1.3 ± 1.0</td>
<td>77</td>
<td>8</td>
</tr>
</tbody>
</table>

aOne 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
International Chemical Identifier


ZYZCGGRZINLQBL-JCGNTXOTSA-N
Comparing ChemSpider Structures

![Chemical Structure Image]

<table>
<thead>
<tr>
<th>ChemSpiderID</th>
<th>Standard InChIKey Stereolayer</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIKIPEDIA</td>
<td>t28,-29,-30,-31+,34,-35,-36+,37,+38,-40+</td>
</tr>
<tr>
<td>CompTox</td>
<td>t28,-29,-30,-31+,34,-35,-36+,37,+38,-40+</td>
</tr>
<tr>
<td>4941647</td>
<td>t28,-29,-30,-31+,34,-35,-36+,37,+38,-40+</td>
</tr>
<tr>
<td>393078</td>
<td>t28,-29,-30,-31+,34,-35,-36+,\textbf{37},-38,-40+</td>
</tr>
<tr>
<td>57618348</td>
<td>t28,-29,-30,-31+,34,-35,-36+,\textbf{37},-38,-40+</td>
</tr>
<tr>
<td>29342071</td>
<td>t28,-29,-30,-31+,\textbf{34}+,35,-36+,\textbf{37},-38,-40+</td>
</tr>
<tr>
<td>7987594</td>
<td>t28,-\textbf{29},?,-30?,31+,\textbf{34}?,-35,-36?,\textbf{37},-38,-40?</td>
</tr>
<tr>
<td>22900854</td>
<td>t28,-\textbf{29},?,30+,31-,\textbf{34}+,35+,36-,\textbf{37},-38,-40-</td>
</tr>
<tr>
<td>19692240</td>
<td>NONE</td>
</tr>
<tr>
<td>2831283</td>
<td>NONE</td>
</tr>
</tbody>
</table>
## Comparing ChemSpider Structures

<table>
<thead>
<tr>
<th>ChemSpiderID</th>
<th>InChIKey</th>
<th># Stereocenters</th>
<th># Different</th>
</tr>
</thead>
<tbody>
<tr>
<td>WIKIPEDIA</td>
<td>ZYDZCGGRZINLQBL-JCGNTXOTSA-N</td>
<td>10/10</td>
<td>0</td>
</tr>
<tr>
<td>CompTox</td>
<td>ZYDZCGGRZINLQBL-JCGNTXOTSA-N</td>
<td>10/10</td>
<td>0</td>
</tr>
<tr>
<td>4941647</td>
<td>ZYDZCGGRZINLQBL-JCGNTXOTSA-N</td>
<td>10/10</td>
<td>0</td>
</tr>
<tr>
<td>393078</td>
<td>ZYDZCGGRZINLQBL-GWRQVWKTSA-N</td>
<td>10/10</td>
<td>1</td>
</tr>
<tr>
<td>57618348</td>
<td>ZYDZCGGRZINLQBL-UPPCHHEJSA-N</td>
<td>10/10</td>
<td>1</td>
</tr>
<tr>
<td>29342071</td>
<td>ZYDZCGGRZINLQBL-IIJTUTQBSA-N</td>
<td>10/10</td>
<td>2</td>
</tr>
<tr>
<td>7987594</td>
<td>ZYDZCGGRZINLQBL-BESLYTPASA-N</td>
<td>5/10</td>
<td>6</td>
</tr>
<tr>
<td>22900854</td>
<td>ZYDZCGGRZINLQBL-QAXSDTKVSA-N</td>
<td>9/10</td>
<td>8</td>
</tr>
<tr>
<td>19692240</td>
<td>ZYDZCGGRZINLQBL-ORZJCNZSA-N</td>
<td>0/10</td>
<td>10</td>
</tr>
<tr>
<td>2831283</td>
<td>ZYDZCGGRZINLQBL-UHFFFAOYSA-N</td>
<td>0/10</td>
<td>10</td>
</tr>
</tbody>
</table>
Other Searches

### UniChem Search Results

<table>
<thead>
<tr>
<th>CMR Query</th>
<th>src_id</th>
<th>Source</th>
<th>src_compound_id</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>ChEMBL</td>
<td>CHEMBL444092</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>Guide to Pharmacology</td>
<td>4735</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>KEGG Ligand</td>
<td>C05371</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>ChEBI</td>
<td>6925</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000169715525</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000255288110</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000255288111</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000255288112</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000255288113</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000255288114</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000255288115</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000583653042</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>ZINC</td>
<td>ZINC000669680403</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>eMolecules</td>
<td>26754757</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>eMolecules</td>
<td>31239828</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td>IBM Patent System</td>
<td>DA3C2F25F29692734272194ED0E2C009</td>
</tr>
<tr>
<td>14</td>
<td>14</td>
<td>FDA SRS</td>
<td>EQ8332842Y</td>
</tr>
</tbody>
</table>
Delivering a Better Database

- We have full time curators checking data

**Record Information**

- **Citation:** U.S. Environmental Protection Agency. Chemistry Dashboard.
  https://comptox.epa.gov/dashboard/DTXSID3031654 (accessed Mar 17th, 2019), Microcystin LR

**Data Quality:**

- **Level 1:** Expert curated, highest confidence in accuracy and consistency of unique chemical identifiers
- **Level 2:** Expert curated, unique chemical identifiers using multiple sources
- **Level 3:** Programmatically curated from high quality EPA source, unique chemical identifiers have no conflicts in ChemID and PubChem
- **Level 4:** Programmatically curated from ChemID, unique chemical identifiers have no conflicts in PubChem
- **Level 5:** Programmatically curated from ACToR or PubChem, unique chemical identifiers with low confidence, single public source
UVCB
CHEMICAL SUBSTANCES
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
UVCB: Complex Surfactants
UVCB: Complex Surfactants

C10-linear alkylbenzenesulfonate

NOCAS_891689 | DTXSID70891689

Searches by DSSTox Substance Id.

4-(Decan-3-yl)benzene-1-sulfonic acid
DTXSID: DTXSID20859618
PubChem: 0
CPDAT: 0

4-(Decan-5-yl)benzene-1-sulfonic acid
DTXSID: DTXSID70881146
PubChem: 0
CPDAT: 0

4-(Decan-4-yl)benzenesulfonic acid
DTXSID: DTXSID40891333
PubChem: 0
CPDAT: 0
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

Non Target Analysis Prototype

Mass Search

\[ \pm \text{ Min/Max} \]

321.138493478 Da \[ \pm \] 0.0000002 Da ppm

Molecular Formula Search

Molecular Formula

Mass or Formula must be entered before searching spectrum

Ionization Type

\[
\begin{array}{c}
\text{ESI+} \\
\text{ESI+} \\
\text{ESI-} \\
\text{EI} \\
\end{array}
\]

Spectra Input

Single Energy  Multiple

304.1332052 11.6199475
downarrow
104.0014047 7.308438600
downarrow
123.0440556 6.538343822
downarrow
195.0756004 6.269438115
downarrow
210.1019051 7.400410782

Peak Match Window:

0.02 Da ppm

Search
### Search Expt. vs. Predicted Spectra

<table>
<thead>
<tr>
<th>Chemical Structure ID</th>
<th>Score (10eV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTXCID101048191</td>
<td>0.22</td>
</tr>
<tr>
<td>DTXCID101181567</td>
<td>0.19</td>
</tr>
<tr>
<td>DTXCID50879086</td>
<td>0.17</td>
</tr>
<tr>
<td>DTXCID60686349</td>
<td>0.14</td>
</tr>
<tr>
<td>DTXCID00830900</td>
<td>0.13</td>
</tr>
<tr>
<td>DTXCID10971176</td>
<td>0.12</td>
</tr>
<tr>
<td>DTXCID60301242</td>
<td>0.12</td>
</tr>
<tr>
<td>DTXCID40703048</td>
<td>0.11</td>
</tr>
<tr>
<td>DTXCID60349982</td>
<td>0.11</td>
</tr>
<tr>
<td>DTXCID10316649</td>
<td>0.09</td>
</tr>
</tbody>
</table>
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

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
• CFM-ID
  – Viewing and Downloading pre-predicted spectra
  – Search spectra against the database
• Retention Time Index Prediction
• Structure/substructure/similarity search
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
## PNNL Collision Cross Section Database

### Chemicals

<table>
<thead>
<tr>
<th>Chemical</th>
<th>SMILES</th>
<th>InChI</th>
<th>Formula</th>
<th>Mass</th>
<th>CCS (Å²)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3E)-pent-3-en-2-one</td>
<td><img src="image1" alt="Chemical Structure" /></td>
<td><img src="image2" alt="Chemical Structure" /></td>
<td>C₅H₈O</td>
<td>84.0575</td>
<td>[M-H]⁻ 112.1 ISiCLE Lite v0.1.0, [M+Na]⁺ 112.6 ISiCLE Lite v0.1.0, [M+H]⁺ 113.1 ISiCLE Lite v0.1.0</td>
</tr>
<tr>
<td>Dimethyl sulfone</td>
<td><img src="image3" alt="Chemical Structure" /></td>
<td><img src="image4" alt="Chemical Structure" /></td>
<td>C₂H₆O₂S</td>
<td>94.0089</td>
<td>[M-H]⁻ 106.9 ISiCLE Lite v0.1.0, [M+Na]⁺ 107.3 ISiCLE Lite v0.1.0, [M+H]⁺ 108.1 ISiCLE Lite v0.1.0</td>
</tr>
<tr>
<td>Isothiocyanatocyclopropane</td>
<td><img src="image5" alt="Chemical Structure" /></td>
<td><img src="image6" alt="Chemical Structure" /></td>
<td>C₄H₅NS</td>
<td>99.0143</td>
<td>[M-H]⁻ 111.9 ISiCLE Lite v0.1.0, [M+Na]⁺ 112.1 ISiCLE Lite v0.1.0, [M+H]⁺ 110.0 ISiCLE Lite v0.1.0</td>
</tr>
</tbody>
</table>
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

[Visit the NORMAN webpage](https://www.norman-network.com/?q=node/236)

---

## NORMAN

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

---

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