Using the US EPA’s CompTox Chemistry Dashboard for structure identification and non-targeted analyses

Antony Williams¹, Andrew D. McEachran³, Seth Newton², Kristin Isaacs², Katherine Phillips², Nancy Baker¹, Chris Grulke¹ and Jon R. Sobus²

1) National Center for Computational Toxicology, U.S. Environmental Protection Agency, RTP, NC
2) National Exposure Research Laboratory, U.S. Environmental Protection Agency, RTP, NC
3) Oak Ridge Institute of Science and Education (ORISE) Research Participant, Research Triangle Park, NC

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA
• A publicly accessible website delivering access:
  – ~760,000 chemicals with related property data
  – Experimental and predicted physicochemical property data
  – Experimental Human and Ecological hazard data
  – Integration to “biological assay data” for 1000s of chemicals
  – Information regarding consumer products containing chemicals
  – Links to other agency websites and public data resources
  – “Literature” searches for chemicals using public resources
  – “Batch searching” for thousands of chemicals

  – DOWNLOADABLE Open Data for reuse and repurposing
CompTox Chemistry Dashboard
https://comptox.epa.gov/dashboard

Chemistry Dashboard

761 Thousand Chemicals

Search a chemical by systematic name, synonym, CAS number, or InChIKey

Identifier substring search

See what people are saying, read the dashboard comments!

Latest News
Read more news

A Movie Regarding how to Identify "Known Unknowns" Using the CompTox Dashboard
March 28th, 2017 at 7:26:41 PM

Recently we published a paper regarding identifying known unknowns using the US EPA's CompTox Chemistry Dashboard. Analytical and Bioanalytical Chemistry, March 2017, Volume 409, Issue 7, pp 1720–1730. A movie explaining the paper in full animated detail has been put on YouTube. Enjoy the movie interlude here.
1,2-Propylene glycol
57-55-6 | DTXSID0021206

Wikipedia

Propylene glycol (IUPAC name: propane-1,2-diol) is a synthetic organic compound with the chemical formula C3H8O2. It is a viscous colorless liquid which is nearly odorless but possesses a faintly sweet taste. Chemically it is classed as a diol and is miscible with a broad range of solvents, including water, acetone, and chloroform. It is produced on a large scale and is primarily used in the production of polymers, but also sees use in food.

Intrinsic Properties

Structural Identifiers

Linked Substances

Presence in Lists

Record Information

Quality Control Notes

Chemical Properties  Env. Fate/Transport  Hazard  AOME (beta)  Exposure  Bioassays  Similar Compounds  Related Substances  Synonyms  Literature  Links  Comments
### Access to Chemical Hazard Data

#### Chemistry Dashboard | EPAHFR

<table>
<thead>
<tr>
<th>Priority</th>
<th>Type</th>
<th>Subtype</th>
<th>Risk Assessment Class</th>
<th>Units</th>
<th>Study Type</th>
<th>Exposure Route</th>
<th>Species</th>
<th>Subsource</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>NOEL</td>
<td>Cardiovascular subchronic</td>
<td>5000.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NOEL</td>
<td>Endocrine subchronic</td>
<td>5000.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LOEL</td>
<td>Hematological subchronic</td>
<td>2500.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LOEL</td>
<td>Hepatic subchronic</td>
<td>2500.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NOEL</td>
<td>Immune system immunotoxic</td>
<td>5000.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NOEL</td>
<td>Renal subchronic</td>
<td>5000.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>LOEL</td>
<td>Systemic subchronic</td>
<td>2500.0 mg/kg-day</td>
<td>subchronic oral rat</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NOEL</td>
<td>Hematological subchronic</td>
<td>1500.0 mg/kg-day</td>
<td>subchronic oral rabbit</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>NOEL</td>
<td>Systemic subchronic</td>
<td>1500.0 mg/kg-day</td>
<td>subchronic oral rabbit</td>
<td>Vaille et al. PPRTV (...)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Structure Identification using the dashboard
  – Formula/mass-based searching – 1 chemical at a time
Advanced Searches

Advanced Search

Mass Search

Mass Search

Molecular Formula Search

Generate Molecular Formula(e)

Include Halogens: F[0-20] Cl[0-20] Br[0-20] I[0-20]
Advanced Searches
Mass Based Search

Mass Search

± Min/Max

M

191.131 Da

± 5 Da ppm

Search
Formula Searches

Molecular Formula Search

C12H17NO

Search
Exact Formula Search: C12H17NO
298 Chemicals
• Structure Identification using the dashboard
  – Formula/mass-based searching – 1 chemical at a time
  – Distilling structures into “MS-Ready form”
Specific Data-Mappings “MS-Ready Structures”

A) Molecular Ion

\[ +H^+ \]

\[ m/z = 256.1702 \]

B) MS-Ready Form

\[ \text{monoisotopic mass} = 255.1623 \]

\[ C_{17}H_{21}NO \]

\[ \text{DTXCID802949} \]

C) Mappings from MS-Ready

- Diphenhydramine
  \[ C_{17}H_{23}NO \]
  \[ 255.1623 \]
  \[ \text{DTXSID4022949} \]

- Diphenhydramine citrate
  \[ C_{23}H_{29}NO_8 \]
  \[ 447.1893 \]
  \[ \text{DTXSID80237211} \]

- Diphenhydramine hydrochloride
  \[ C_{17}H_{21}CINO \]
  \[ 291.1390 \]
  \[ \text{DTXSID4020537} \]

- Diphenhydramine salicylate
  \[ C_{24}H_{27}NO_4 \]
  \[ 393.1940 \]
  \[ \text{DTXSID10225883} \]
Diphenhydramine
15 Total MS-Ready Mappings
“MS Ready” Formula Search C12H17NO
354 Chemicals
Dashboard for Structure ID

- Structure Identification using the dashboard
  - Formula/mass-based searching – 1 chemical at a time
  - Distilling structures into “MS-Ready form”
  - Ranking based on metadata
### Table 1

Searching ChemSpider by Elemental Composition then Sorting by Number of Associated References

<table>
<thead>
<tr>
<th>Class of compounds</th>
<th>Number compounds in class</th>
<th>Position of compound sorted in descending order by number of references</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>#1</td>
</tr>
<tr>
<td>Drugs</td>
<td>45</td>
<td>43</td>
</tr>
<tr>
<td>Pesticides</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Toxins</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Polymer antioxidants</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Polymer UV stabilizers</td>
<td>10</td>
<td>8</td>
</tr>
</tbody>
</table>
Identifying known unknowns using the US EPA’s CompTox Chemistry Dashboard

Andrew D. McEachran¹ · Jon R. Sobus² · Antony J. Williams³

<table>
<thead>
<tr>
<th></th>
<th>Mass-based searching</th>
<th>Formula-based searching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dashboard</td>
<td>ChemSpider</td>
</tr>
<tr>
<td>Average rank position</td>
<td>1.3</td>
<td>2.2a</td>
</tr>
<tr>
<td>Percent in #1 position</td>
<td>85%</td>
<td>70%</td>
</tr>
</tbody>
</table>

a Average rank in ChemSpider shown here does not include an outlier where the rank was 201, when added the average rank position is 3.5
Additional Metadata Ranking

- US EPA CompTox Chemistry Dashboard Data Sources
- “CPDat” Consumer Product Database
- PubChem Data Source Count
- PubMed Reference Count
Additional Metadata Ranking
C12H17NO: 354 Chemicals

<table>
<thead>
<tr>
<th>Structure</th>
<th>DTXSID</th>
<th>Preferred Name</th>
<th>CASRN</th>
<th>QC Level</th>
<th>CPDat Count</th>
<th>Number of Sources</th>
<th>PubChem Data Source</th>
<th>PubMed Data Source</th>
<th>Monoisotopic Mass</th>
</tr>
</thead>
<tbody>
<tr>
<td>DTXSID2021995</td>
<td>DEET</td>
<td>134-60-3</td>
<td>Level 1</td>
<td>111</td>
<td>104</td>
<td>155</td>
<td>753</td>
<td>191.131014</td>
<td></td>
</tr>
<tr>
<td>DTXSID1023447</td>
<td>Phenindione</td>
<td>604-03-7</td>
<td>Level 2</td>
<td>12</td>
<td>27</td>
<td>35</td>
<td>50</td>
<td>191.131014</td>
<td></td>
</tr>
<tr>
<td>DTXSID1024197</td>
<td>N-Benzylacetamide</td>
<td>91-49-6</td>
<td>Level 2</td>
<td>1</td>
<td>26</td>
<td>50</td>
<td>1</td>
<td>191.131014</td>
<td></td>
</tr>
<tr>
<td>DTXSID0179048</td>
<td>N-ethylphenylacetamide</td>
<td>2431-06-1</td>
<td>Level 4</td>
<td>0</td>
<td>18</td>
<td>52</td>
<td>34</td>
<td>191.131014</td>
<td></td>
</tr>
<tr>
<td>DTXSID06965298</td>
<td>N-ethyl-N-methylbenzamide</td>
<td>2728-04-3</td>
<td>Level 1</td>
<td>0</td>
<td>11</td>
<td>48</td>
<td>0</td>
<td>191.131014</td>
<td></td>
</tr>
</tbody>
</table>
### Additional Metadata Ranking

**C12H17NO: 354 Chemicals**

<table>
<thead>
<tr>
<th>Structure</th>
<th>DTXSID</th>
<th>Preferred Name</th>
<th>CPDat Count</th>
<th>Number of Sources</th>
<th>PubChem Data Source</th>
<th>PubMed Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DTXSID2021995</td>
<td>DEET</td>
<td>111</td>
<td>104</td>
<td>155</td>
<td>753</td>
</tr>
<tr>
<td></td>
<td>DTXSID1023447</td>
<td>Phendimetrazine</td>
<td>12</td>
<td>27</td>
<td>35</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>DTXSID2042197</td>
<td>N-Butylacetanilide</td>
<td>1</td>
<td>26</td>
<td>50</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>DTXSID00179048</td>
<td>N,N-Diethylphenylacetamide</td>
<td>0</td>
<td>18</td>
<td>52</td>
<td>34</td>
</tr>
</tbody>
</table>
Top Ranked Chemical

DEET
134-62-3 | DTXSID2021995

Wikipedia

N,N-Diethyl-meta-toluamide, also called DEET () or diethyltoluamide, is the most common active ingredient in insect repellents. It is a slightly yellow oil intended to be applied to the skin or to clothing and provides protection against mosquitoes, ticks, fleas, chiggers, lice, and many biting insects. [Read more]

Intrinsic Properties
Structural Identifiers
Linked Substances
Presence in Lists
Record Information
Quality Control Notes

https://comptos.epa.gov/dashboard/downloads
Additional data streams in development

- US EPA CompTox Chemistry Dashboard Data Sources
- “CPDat” Consumer Product Database
- PubChem Data Source Count
- PubMed Reference Count
- Retention Time Prediction
  \[ SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \cdots \]
- Predicted Environmental Media Occurrence
- Presence in Lists
<table>
<thead>
<tr>
<th>List Name</th>
<th>Number of Chemicals</th>
<th>List Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drinking Water Suspects, KWR Water, Netherlands</td>
<td>136</td>
<td>KWRS/JERPS is a list of prioritized suspects relevant for human health in drinking water from KWR Water in Nieuwegein, The Netherlands. The methods are detailed in Sjerps et al 2016, DOI: 10.1016/j.waters.2016.02.034</td>
</tr>
<tr>
<td>EPA Consumer Products Suspect Screening Results</td>
<td>1705</td>
<td>This is a compiled list of the suspects reported in the supporting information of Phillips et al 2018, DOI: 10.1021/acs.est.7b04781 - Suspect Screening Analysis of Chemicals in Consumer Products with GCxGC-TOF/MS.</td>
</tr>
<tr>
<td>EPA Integrated Risk Information System (IRIS)</td>
<td>510</td>
<td>EPA's IRIS Program identifies and characterizes the health hazards of chemicals found in the environment. Each IRIS assessment can cover a chemical, a group of related chemicals, or a complex mixture.</td>
</tr>
<tr>
<td>EPAHFR - EPA Chemicals associated with hydraulic fracturing</td>
<td>1640</td>
<td>EPAHFR lists chemicals associated with hydraulic fracturing from 2005-20013, as reported in EPA's Hydraulic Fracturing Drinking Water Assessment Final Report (Dec 2016)</td>
</tr>
<tr>
<td>STOFF-IDENT Database of Water-Relevant Substances</td>
<td>8885</td>
<td>STOFF-IDENT is a database of water relevant substances collated from various sources within the STOFF-IDENT and FOR-IDENT projects, hosted by IFU, HSWT and TUM. The database at <a href="https://wwwifu.bayern.de/stoffident/#home">https://wwwifu.bayern.de/stoffident/#home</a> has additional functionality.</td>
</tr>
<tr>
<td>Superfund Chemical Data Matrix</td>
<td>220</td>
<td>The Superfund Chemical Data Matrix (SCDM) generates a list of the corresponding Hazard Ranking System (HRS) factor values, benchmarks, and data elements for a particular chemical.</td>
</tr>
<tr>
<td>Surfactant List Screened in Swiss Wastewater (2014)</td>
<td>122</td>
<td>EAWAGSURF is a list of surfactants screened in Swiss wastewater effluents as part of a 2014 study. Structures/mixtures are being progressively curated and linked (Schymanski/Williams). Further details in Schymanski et al 2014, DOI: 10.1021/es4044374</td>
</tr>
</tbody>
</table>
• Structure Identification using the dashboard
  – Formula/mass-based searching – 1 chemical at a time
  – Distilling structures into “MS-Ready form”
  – Ranking based on metadata
  – Batch searching of formulae and masses
Batch Search

Step One: Select Input

Select Input Type(s)
- Chemical Name
- CASRN
- InChIKey
- Skeleton
- DSSTox Substance ID
- MS-Ready Formula
- Exact Formula(e)
- Monoisotopic Mass

This search is based on what we refer to as ‘Mass Spec Ready Formulae’. All chemicals within the database are treated in a manner that all are desalted and stereochemistry is removed as Mass Spectrometry detects the major components of a salt and is insensitive to stereochemistry. As an example, a search for the formula associated with phenol will return phenol, sodium phenolate and calcium phenoxide.
Batch Search

Select Input Type(s)
- Chemical Name
- CASRN
- InChIKey
- Skeleton
- DSSTox Substance ID
- MS-Ready Formula(e)
- Exact Formula(e)
- Monoisotopic Mass

Enter Identifiers to Search (searches should be limited to <1000 identifiers)
- C6H12O3
- C7H7N3
- C8H11NO
- C7H5NOS
- C9H15NO
- C11H12O
- C9H8O3
- C6H12O5
- C9H15NO2

Metadata
- Curation Level Details
- Data Sources
- Assay Hit Count
- Include links to ACToR reports - SLOWI (BETA)
- NHANES/Predicted Exposure
- Include ToxVal Data Availability
- Number of PubMed Articles
- Abstract Sifter Input File (Beta)
- MetFrag Input File (Beta)
- IRIS
- PPRTV
- PubChem Data Sources
- ToxPrint fingerprints

- NIOSH IDLH Values
- NIOSH International Chemical Safety Cards
- NIOSH Pocket Guide to Chemical Hazards
- NIOSH Skin Notation Profiles
- NORMAN Collaborative Trial 2015 Targets and Suspects
- Norman Network PFAS (KEMI Report)
- NORMAN Network Priority List
- NormaNEWS: Norman Early Warning System
- PFAS list provided by X.Trier et al
- Pharmaceutical List with EU, Swiss and US Consumption Data
- Provisional Peer Reviewed Toxicity Values
- Stockholm Convention on Organic Pollutants
- STOFF-IDENT Database of Water-Related Substances
- Superfund Chemical Data Matrix
- Surfactant List Screened in Swiss Waste Water (2014)
| INPUT  | FOUND_BY | DTXCID | INPUT | DTXCID | DTXCID | DTXCID | DTXCID | DTXCID | DTXCID | DTXCID | DTXCID | DTXCID | DATA | SQL | TOXVAL | TOXCAST | TOXCAST | NUMBER | PUBCHEM | STHO |
|--------|----------|--------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|------|-------|--------|--------|--------|--------|--------|------|
| C6H12O5 | MS Ready | DTXCID7015 | Y | 0.36 | 2/562 | 24 | 83 | Y |
| C6H12O3 | MS Ready | DTXCID003 | 67 | Y | 0.36 | 1/276 | 376 | 80 | Y |
| C6H12O3 | MS Ready | DTXCID106 | 65 | Y | 4.42 | 5/113 | 6 | 77 | Y |
| C6H12O3 | MS Ready | DTXCID105 | 45 | Y | 0.0 | 0/163 | 3 | 94 | Y |
| C6H12O3 | MS Ready | DTXCID9011 | 38 | Y | - | - | 14 | 110 | Y |
| C6H12O3 | MS Ready | DTXCID402 | 34 | Y | 0.0 | 0/113 | - | 53 | Y |
| C6H12O3 | MS Ready | DTXCID202 | 31 | Y | - | - | - | 36 | Y |
| C6H12O3 | MS Ready | DTXCID109 | 30 | - | 2.54 | 7/276 | - | 54 | Y |
| C6H12O3 | MS Ready | DTXCID202 | 24 | Y | 0.0 | 0/113 | - | 47 | Y |
| C6H12O3 | MS Ready | DTXCID303 | 22 | Y | - | - | - | 89 | Y |
| C6H12O3 | MS Ready | DTXCID302 | 20 | Y | - | - | 2 | 25 | Y |
| C6H12O3 | MS Ready | DTXCID407 | 19 | Y | - | - | 12 | 62 | Y |
| C6H12O3 | MS Ready | DTXCID704 | 17 | Y | - | - | - | 64 | Y |
| C6H12O3 | MS Ready | DTXCID704 | 16 | Y | - | - | 3 | 49 | Y |
Batch Search Integration to MetFrag

http://c-ruttkies.github.io/MetFrag/projects/metfragweb/

1. Candidate database
2a. Neutral mass of the precursor molecule
3. Relative mass deviation for candidate selection
4. Neutral molecular formula of the precursor (if available)
2b. Calculate neutral precursor mass from charged ion mass
5. Database specific candidate identifiers
MetFrag Input File

- Curation Level Details
- Data Sources
- Assay Hit Count
- Include links to ACToR reports - SLOW! (BETA)
- NHANES/Predicted Exposure
- Include ToxVal Data Availability
- Number of PubMed Articles
- Abstract Sifter Input File (Beta)
- MetFrag Input File (Beta)
- IRIS
- PPRTV
- PubChem Data Sources
Batch Search Integration to MetFrag

http://c-ruttkies.github.io/MetFrag/projects/metfragweb/
The Dashboard to Support MS-Analysis

MS-Ready Structures Underpin Analysis
Downloadable Data

**DSSTox Identifier to PubChem Identifier Mapping File**

The DSSTox to PubChem Identifiers mapping file is in TXT format and includes the PubChem SID, PubChem CID and DSSTox substance identifier (DTXSID).

<table>
<thead>
<tr>
<th>SID</th>
<th>CID</th>
<th>DTXSID</th>
</tr>
</thead>
<tbody>
<tr>
<td>316388891</td>
<td>20404</td>
<td>DTXSID30873143</td>
</tr>
<tr>
<td>316388890</td>
<td>10412816</td>
<td>DTXSID70873142</td>
</tr>
<tr>
<td>316388889</td>
<td>50742127</td>
<td>DTXSID40873139</td>
</tr>
<tr>
<td>316388888</td>
<td>19073841</td>
<td>DTXSID20873137</td>
</tr>
<tr>
<td>316388887</td>
<td>115050215</td>
<td>DTXSID00873135</td>
</tr>
<tr>
<td>316388886</td>
<td>25021861</td>
<td>DTXSID80873133</td>
</tr>
<tr>
<td>316388885</td>
<td>27884287</td>
<td>DTXSID00873131</td>
</tr>
<tr>
<td>316388884</td>
<td>6731</td>
<td>DTXSID00873130</td>
</tr>
</tbody>
</table>

**DSSTox identifiers mapped to CAS Numbers and Names File**

The DSSTox Identifiers file is in Excel format and includes the CAS Number, DSSTox substance identifier (DTXSID) and the Preferred Name.

<table>
<thead>
<tr>
<th>casn</th>
<th>dsstox_substance_id</th>
<th>preferred_name</th>
</tr>
</thead>
<tbody>
<tr>
<td>26148-68-5</td>
<td>DTXSID7020001</td>
<td>4-alpha-C</td>
</tr>
<tr>
<td>107-29-9</td>
<td>DTXSID2020004</td>
<td>Acetamide oxime</td>
</tr>
<tr>
<td>60-35-5</td>
<td>DTXSID2020005</td>
<td>Acetamide</td>
</tr>
<tr>
<td>101-99-2</td>
<td>DTXSID2020006</td>
<td>Acetaminophen</td>
</tr>
<tr>
<td>908-81-0</td>
<td>DTXSID2020007</td>
<td>Acetohexamide</td>
</tr>
<tr>
<td>18523-69-8</td>
<td>DTXSID2020008</td>
<td>Acetone(4-(5-nitro-2-furyl)-2-thiazolyl) hydrazone</td>
</tr>
<tr>
<td>75-05-8</td>
<td>DTXSID2020009</td>
<td>Acetophenone</td>
</tr>
<tr>
<td>127-96-9</td>
<td>DTXSID2020010</td>
<td>Acetolene</td>
</tr>
<tr>
<td>65734-38-5</td>
<td>DTXSID6020012</td>
<td>N'-Acetyl-4-(hydroxymethyl) phenylhydrazine</td>
</tr>
</tbody>
</table>
Future Work: Combined Substructure/Formula Searching
Future Work: Searching Against Predicted Spectra

Spectra Prediction

Predicts the spectra for a given input molecule. Spectra are computed for low (10V), medium (20V) and high (40V) collision energy levels and are represented by a list of mass intensity pairs, each corresponding to a peak in the spectra.

Input: 
- **Parent Compound Structure**: CCN(C)(C)(O)C1=C(C=C)C=C(C)=C1
- **InChI or SMILES format**
- **Spectra Type**: ESI
- **Ion Mode**: Positive
- **Adduct Type**: [M+H]+

If you wish to run multiple jobs, input larger query molecules, or customize the computation parameters, you can freely download the source code here: [sourceforge.net/projects/cfm-id](http://sourceforge.net/projects/cfm-id).
Future Work: Searching Against Predicted Spectra

- CFM-ID predicted spectra generated for 700,000 chemicals
  - Positive ion, Negative ion, Electron Impact
  - Three energies
Future Work

Scoring scheme into results

\[ SC_{TOTAL} = SC_{DS} + SC_{PM} + SC_{RT} + SC_{MO} + \cdots \]
Conclusion

- The CompTox Chemistry Dashboard provides access to data for ~760,000 chemicals
- High quality curated data and rich metadata facilitates mass spec analysis
- “MS-Ready” processed data enables structure identification
Acknowledgments

• The CompTox Chemistry Dashboard team
• NERL colleagues:
  – Jon Sobus, Elin Ulrich, Mark Strynar, Seth Newton (NTA Analysis)
  – Katherine Phillips, Kathie Dionisio, Kristin Isaacs (Consumer Products Database)
• Emma Schymanski – Luxembourg Center for Systems Biomedicine (MS-ready/NTA)
Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology (NCCT)
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821