Integrating Mass Spectrometry Non-Targeted Analysis and Computational Toxicology to Characterize Chemicals

Antony Williams

Center for Computational Toxicology and Exposure, US-EPA, RTP, NC

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

June 2022

Medical Device and Combination Product Specialty Section webinar
CompTox Chemicals Dashboard
>906k chemicals
The Charge for the Dashboard

• Develop a “first-stop-shop” for environmental chemical data to support EPA and partner decision making:
  – **Centralized location** for relevant chemical data
  – Chemistry, exposure, hazard and dosimetry
  – Combination of existing data and predictive models
  – Publicly accessible, periodically updated, curated

• Easy access to data improves efficiency and ultimately accelerates chemical risk assessment
CompTox Chemicals Dashboard
https://comptox.epa.gov/dashboard
• Chemical page: Wikipedia snippet when available, intrinsic properties, structural identifiers, linked substances
Executive Summary

- Overview of toxicity-related info
  - Quantitative values
  - Physchem. and Fate & Transport
  - Adverse Outcome Pathway links
  - In vitro bioactivity summary plot
Experimental and Predicted Data

- Physchem and Fate & Transport experimental and predicted data
- Data can be downloaded as Excel, TSV and CSV files

<table>
<thead>
<tr>
<th>Property</th>
<th>Experimental average</th>
<th>Predicted average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polarizability</td>
<td>-</td>
<td>35.8 (1)</td>
</tr>
<tr>
<td>Henry's Law</td>
<td>4.57e-7 (1)</td>
<td>4.59e-7 (1)</td>
</tr>
<tr>
<td>Boiling Point</td>
<td>495 (3)</td>
<td>480 (4)</td>
</tr>
<tr>
<td>Flash Point</td>
<td>-</td>
<td>234 (2)</td>
</tr>
<tr>
<td>Melting Point</td>
<td>177 (8)</td>
<td>189 (3)</td>
</tr>
<tr>
<td>Molar Refractivity</td>
<td>-</td>
<td>90.3 (1)</td>
</tr>
<tr>
<td>Molar Volume</td>
<td>-</td>
<td>196 (1)</td>
</tr>
<tr>
<td>Surface Tension</td>
<td>-</td>
<td>53.9 (2)</td>
</tr>
<tr>
<td>Density</td>
<td>-</td>
<td>1.28 (2)</td>
</tr>
<tr>
<td>Vapor Pressure</td>
<td>5.49e-9 (1)</td>
<td>3.61e-9 (3)</td>
</tr>
</tbody>
</table>
### Chemical Hazard Data

**ToxVal Database**

- >50k chemicals
- >770k tox. values
- >30 sources of data
- ~5k journals cited
- ~70k citations

![ToxVal Database Interface](image-url)
**GHS Data**

CID 2336

**Benzo[a]pyrene**

**GHS Classification**

Showing 6 of 6

**Pictogram(s)**

- 🚨 Immediate
- 📊 Health Hazard
- 🌿 Environmental Hazard

**Signal**

- **Danger**

<table>
<thead>
<tr>
<th>Signal</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>H317:</td>
<td>May cause an allergic skin reaction [Warning Sensitization, Skin]</td>
</tr>
<tr>
<td>H340:</td>
<td>May cause genetic defects [Danger Germ cell mutagenicity]</td>
</tr>
<tr>
<td>H350:</td>
<td>May cause cancer [Danger Carcinogenicity]</td>
</tr>
</tbody>
</table>

**Hazard Statements**

- H360FD: May damage fertility; May damage the unborn child [Danger Reproductive toxicity]
- H400: Very toxic to aquatic life [Warning Hazardous to the aquatic environment, acute hazard]
- H410: Very toxic to aquatic life with long lasting effects [Warning Hazardous to the aquatic environment, long-term hazard]

**Precautionary Statement Codes**


(The corresponding statement to each P-code can be found at the GHS Classification page.)
### Chemical Weight Fractions (CWF)

<table>
<thead>
<tr>
<th>Product Name</th>
<th>Product Use Category</th>
<th>Categorization Subtype</th>
<th>Minimum Weight Fraction</th>
<th>Maximum Weight Fraction</th>
<th>Data Type</th>
<th>Source</th>
<th>Product Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>48741 exh mixture</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>asphalt cement penetration 60-70</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>base-neutral 4.1ml methylene chl.</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>base neutral calibration check co.</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>benzo (a) pyrene_ 98% b100%</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>benzo (a) pyrene_ md-1956</td>
<td>Not yet Categorized</td>
<td></td>
<td>0.990</td>
<td>1.00</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>blastout 2000 c/ 9 co 875</td>
<td>Not yet Categorized</td>
<td></td>
<td>0.00</td>
<td>1.00e+3</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>blastout 2000 universal_ 873</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>blastout 2000 universal_ 870</td>
<td>Not yet Categorized</td>
<td></td>
<td>0.00</td>
<td>1.00e-3</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>blastout 4000 strong_ 872</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>blastout 4000 universal_ 872</td>
<td>Not yet Categorized</td>
<td></td>
<td>0.00</td>
<td>1.00e-3</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
<tr>
<td>clo-011a clp base/neutrals check ...</td>
<td>Not yet Categorized</td>
<td></td>
<td>-</td>
<td>-</td>
<td>reported</td>
<td>SIRI</td>
<td>1</td>
</tr>
</tbody>
</table>

Rows: 29
Searching Literature and the Internet
Real-time retrieval of data from PubMed (~30 million abstracts and growing)
Choose from set of pre-defined queries
Adjust and fine tune queries based on interests
**Determination of atrazine and its metabolites in mouse urine and plasma by LC-MS analysis.**

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

• We know how complex identifiers are…
  – CASRN(s)
  – Hundreds of names (maybe)
  – SMILES
  – InChIs
  – EINECS, EC numbers

• What can WE do to help navigate the internet?
Benzo(a)pyrene
50-32-8 | DTXSID2020139
Searched by DSSTox Substance Id.

General
- EPA Substance Registry Service
- PubChem
- Chemsider
- CPCat
- DrugBank
- Wikipedia
- MSDS Lookup
- ChEMBL
- ToxPlanet
- ACS Reagents: Chemicals
- Wolfram Alpha
- ECHA Infocard
- ChemAgora
- Consumer Product Information Database
- ChEBI
- NIST Chemistry Webbook
- WEBWISER
- PubChem Safety Sheet
- Consumer Product Information Database
- PubChem Chemical Vendors

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

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

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

Prediction
- 2D NMR HSQC/HMBC Prediction
- Carbon-13 NMR Prediction
- Proton NMR Prediction
- ChemRIP NMR Predictor
- LSERD
• Links to ~90 websites providing access to additional data on the chemical of interest
NIST WebBook
https://webbook.nist.gov/chemistry/
MassBank of North America
https://mona.fiehnlab.ucdavis.edu
Chemical Lists of Interest...
Lists of Extractables and Leachables
https://comptox.epa.gov/dashboard/chemical-lists?filtered=&search=extractables

- Chemical lists with extractables & leachables
- Expands with literature extraction
EXTRACTABLES: Extractables & Leachables Safety Information Exchange (ELSIDE)

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey

Start typing to search.

Identifier substring search

List Details

Description: The Extractables and Leachables Safety Information Exchange (ELSIDE: https://www.elsiedata.org/) was established by scientists in pharmaceuticals companies to advance the concept of sharing pre-competitive safety information on extractables and leachables, among industry. The vision was that such a collaborative effort would reduce duplicative safety studies across companies, streamline development projects, and allow industry and other stakeholders to share experiences and information to help advance the practice and science of extractables, leachables and materials evaluation.

Number of Chemicals: 457
Extractables

EXTRACTABLES: Leachable Chemical Substances from Common Drinking Water Piping Materials

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey
Start typing to search.

Identifier substring search

List Details

Number of Chemicals: 151

EXTRACTABLES: Chemical migrants in plastic food contact products

Search for chemical by systematic name, synonym, CAS number, DTXSID or InChIKey
Start typing to search.

Identifier substring search

List Details

Description: List of chemical migrants in plastic food contact products listed in the article “Detection and quantification analysis of chemical migrants in plastic food contact products” by Qian et al [https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0208467]
Number of Chemicals: 81
Download data, send to batch search for data harvesting
“MS-ready” structures

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

Andrew D. McEachran, Kamel Mansouri, Chris Grulke, Emma L. Schymanski, Christoph Rutkies, and Antony J. Williams
Overview of MS-Ready Structures

• All structure-based chemical substances are algorithmically processed to
  – Split multicomponent chemicals into individual structures
  – Desalt and neutralize individual structures
  – Remove stereochemical bonds from all chemicals

• MS-Ready structures are then mapped to original substances to provide a path between chemicals detected by mass spectrometry to original substances
Nicotine
CN1CCC[C@H]1C1=CN=CC=C1
DTXSID1020930 | SNICXCGAKADSCV
54-11-5 | 162.1157 | 0.929 | 72
Tox: yes | Expo: yes | Bioassay: yes

D-Nicotine
CN1CC[C@H]1C1=CN=CC=C1
DTXSID004635 | SNICXCGAKADSCV
25162-00-9 | 162.1157 | 0.929 | 20
Tox: no | Expo: yes | Bioassay: yes

Benzoic acid, 2-hydroxy-, compd. with
3-[(2S)-1-methyl-2-pyrrolidinyl]pyridine (1:1)
OC=(O)C1=C(O)C=CC=C1.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID075319 | AIBWPBUEUAKMKNS
29790-52-1 | 300.1474 | 0.929 | 6
Tox: no | Expo: yes | Bioassay: no

Nicotine hydrochloride
CI.CN1CCC[C@H]1C1=CN=CC=C1
DTXSID602093 | HDJBTCAJIMNWXE
2820-51-1 | 198.0924 | 0.929 | 9
Tox: no | Expo: yes | Bioassay: yes

DL-Nicotine
CN1CC1C1=CN=CC=C1
DTXSID3048154 | SNICXCGAKADSCV
22083-74-5 | 162.1157 | 0.953 | 9
Tox: yes | Expo: no | Bioassay: yes

DL-Nicotine-d3
[2H][2H][2H]CN1CC1C1=CN=CC=C1
DTXSID80442666 | SNICXCGAKADSCV
69980-24-1 | 165.1345 | 0.929 | 1
Tox: no | Expo: no | Bioassay: no
Mass and Formula Searching
Advanced Searches: Mass Search

Mass Search

± Min/Max

Adduct All Adducts
Neutral

Choose adduct from dropdown

215.093 Da ± 5 Da ppm

Search
Advanced Searches: Mass Search

Search Results
Searched by Mass: 215.093 +/- 5.0 ppm.

27 of 30 chemicals visible

Select all  Download  Send to Batch Search  Mass Difference

<table>
<thead>
<tr>
<th>Chemical Name</th>
<th>CAS Number</th>
<th>Structure</th>
<th>Molecular Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1H-Pyrazole-4-diazonium, 2,3-dihydro-1,3,5-triazin...</td>
<td>DTXSID:DTXSID201480103</td>
<td><img src="image1" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>4-[{2-Hydroxyethyl}(4-oxobut-2-en-3-yl)methyl]-2...</td>
<td>CASRN:11044-04-8</td>
<td><img src="image2" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>6-Ethoxy-5,5-dimethyl-4,6-dioxoheptanoate</td>
<td>CASRN:0000503</td>
<td><img src="image3" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>Atrazine</td>
<td>DTXSID:DTXSID28900112</td>
<td><img src="image4" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>6-Chloro-N^2-ethyl-N^6-propyl-1,3,5-triazin...</td>
<td>CASRN:9952-64-0</td>
<td><img src="image5" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>N-Butyl-3-chloro-6-hydrazinopyridazine</td>
<td>CASRN:51261-46-9</td>
<td><img src="image6" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>N-Butyl-6-chloro-3-hydrazinopyridazine</td>
<td>CASRN:51261-44-7</td>
<td><img src="image7" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>N^2-(Butan-2-yl)-6-chloro-N^6-methyl-1,3,5-triazin...</td>
<td>CASRN:9953-35-0</td>
<td><img src="image8" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>4-(Butylimino)-5-chloro-3-imino-1,2,3,4-tetrahydro...</td>
<td>CASRN:6270-25-3</td>
<td><img src="image9" alt="Structure" /></td>
<td>215.093</td>
</tr>
<tr>
<td>2-(3-Trifluoromethylphenyl)pyrrolidine</td>
<td>CASRN:109836-17-1</td>
<td><img src="image10" alt="Structure" /></td>
<td>215.093</td>
</tr>
</tbody>
</table>
Advanced Searches: Formula Search

Search Results
Searched by Exact Molecular Formula: C8H14CIN5.

7 chemicals

- Atrazine
  - DTXSID:DTXSID096200112
  - CASRN:1119-92-4
  - TOXCAS:62/1024

- GS 18183
  - DTXSID:DTXSID000187506
  - CASRN:343833-27-2
  - TOXCAS:-

- 6-Chloro-N²-ethyl-N⁴-propyl-1,3,5-triazine-2,4-diamine
  - DTXSID:DTXSID00536638
  - CASRN:90952-64-0
  - TOXCAS:-

- N-Butyl-6-chloro-2-hydroxy-3-pyridazinecarboxamide
  - DTXSID:DTXSID00615440
  - CASRN:61361-46-9
  - TOXCAS:-

- N-Butyl-6-chloro-3-hydroxy-2-pyridazinecarboxamide
  - DTXSID:DTXSID00615443
  - CASRN:61361-46-7
  - TOXCAS:-

- N²-(Butyl)imidazo[6,2-b]-1,2,3,4-tetrahydroidopyrimidin-2(3H)-one-9,10-diol
  - DTXSID:DTXSID09976320
  - CASRN:6270-25-3
  - TOXCAS:-

- 4-(Butoxymethyl)-6-chloro-2-imino-1,2,3,4-tetrahydropyrimidin-2(3H)-one-9,10-diol
  - DTXSID:DTXSID09976320
  - CASRN:6270-25-3
  - TOXCAS:-
Batch Searching
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 Search

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

2. Enter Identifiers to Search
   (Please enter one identifier per line. Processing time increases with number of inputs.)
   - DTXSID9020374
   - DTXSID9020827
   - DTXSID2022678
   - DTXSID4023381
   - DTXSID9044164
   - DTXSID7032004
   - DTXSID4022361
   - DTXSID8021771

3. Display All Chemicals

45% batched

45 Chemicals Found from 110 Input(s)

Replace Identifiers with Selected Chemicals

<table>
<thead>
<tr>
<th>Structure</th>
<th>DTXSID</th>
<th>Preferred Name</th>
<th>CASRN</th>
<th>Mono. Mass</th>
<th>Mol. Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DTXSID30022678</td>
<td>Boltaamidine</td>
<td>90357-06-5</td>
<td>430.061041</td>
<td>C18H20F8N2O5</td>
</tr>
<tr>
<td></td>
<td>DTXSID3002621</td>
<td>PR-8-Fenoxinate</td>
<td>67614-33-9</td>
<td>498.128821</td>
<td>C25H22ClNO3</td>
</tr>
</tbody>
</table>
Batch Search – Excel, CSV, SDF file

### Customize Export Results

Your file will be exported in Microsoft Excel Format (.xlsx)

#### Chemical Identifiers

- DTXSID
- CASRN
- INCHIKey
- IUPAC Name

#### Structures

- MolFile
- SMILES
- IUPAC Name
- Chemical Names
- SMILES

#### Intrinsic and Predicted Properties

- Molecular Formula
- Average Molecular Weight
- Monoisotopic Mass
- TEST Model Predictions
- CERES Model Predictions

#### Metadata

- Citation Level Details
- Safety Data
- CASRN/Predicted Exposure
- Data Sources
- Include Toxic Data Availability
- Assay Hit Count
- Number of Enriched Actives
- PubChem Data Sources
- CASRN Product Occurrence Count
- IRIS
- PREDICT
- Wikipedia Article
- QC Notes
- Include links to ASTOR reports

#### Presence in Lists

<table>
<thead>
<tr>
<th>Title</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>40CFR1164 CF</td>
<td>40 CFR 1164 Designation of Hazardous Substances (Above Ground Storage Tanks)</td>
</tr>
<tr>
<td>ACEReadCF</td>
<td>LIST: ACEReadCF, Aromatic Chlorinated Hydrocarbons</td>
</tr>
<tr>
<td>AECOLVALUES CF</td>
<td>AECOL VALUES, Aqueous Equilibrium, Solubility Levels</td>
</tr>
<tr>
<td>AGALTON CF</td>
<td>LIST: AGALTON, Aqueous Equilibrium, Solubility Levels</td>
</tr>
<tr>
<td>ALLSURFACTANTS CF</td>
<td>CATEGORY: Surfactants</td>
</tr>
<tr>
<td>AMILOGAECOS CF</td>
<td>CATEGORY: Alkylglycosides</td>
</tr>
<tr>
<td>AMPHIBOLES CF</td>
<td>Amphibole minerals</td>
</tr>
<tr>
<td>ANTIMICROBIALS CF</td>
<td>CATEGORY: ANTIMICROBIALS, Antimicrobials from HHS.gov</td>
</tr>
<tr>
<td>AOPSTRESSORS CF</td>
<td>List of Advanced Oxidation Process Stressors</td>
</tr>
<tr>
<td>APOCARETRO CF</td>
<td>LIST: APOMIXIS, Chemicals for Reproductive Analysis</td>
</tr>
</tbody>
</table>

Rows: 315

#### Download Export file for the chemicals selected

Download Export file for the chemicals selected.
Work in Progress
New lists added regularly

• New chemicals registered daily and released with each new version of the Dashboard
Predicted Mass Spectra
http://cfmid.wishartlab.com/

- MS/MS spectra prediction for ESI+, ESI-, and EI
- Predictions generated and stored for >800,000 structures, to be accessible via Dashboard
Spectral Viewer Comparison
Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-019-02351-7

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

Alex Chao¹,² • Hussein Al-Ghoul¹,² • Andrew D. McEachran¹,³ • Ilya Balabin⁴ • Tom Transue⁴ • Tommy Cathey⁴ • Jarod N. Grossman²,³ • Randolph Singh¹,⁵ • Elin M. Ulrich² • Antony J. Williams⁶ • Jon R. Sobus²

Received: 4 October 2019 / Revised: 27 November 2019 / Accepted: 11 December 2019
© The Author(s) 2019
Critical Assessment of Small Molecule Identification

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

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

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

Mailing lists

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

<table>
<thead>
<tr>
<th>CASMI Year</th>
<th>CFM-ID Only</th>
<th>CFM-ID + DS</th>
<th>Winners’ Results</th>
<th>Total in DB/Total in Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>36%</td>
<td>64%</td>
<td>36%</td>
<td>14/14</td>
</tr>
<tr>
<td>2013</td>
<td>81%</td>
<td>88%</td>
<td>88%</td>
<td>16/16</td>
</tr>
<tr>
<td>2014</td>
<td>57%</td>
<td>76%</td>
<td>71%</td>
<td>42/42</td>
</tr>
<tr>
<td>2016-training</td>
<td>63%</td>
<td>96%</td>
<td>81%</td>
<td>312/312</td>
</tr>
<tr>
<td>2016-challenge</td>
<td>66%</td>
<td>94%</td>
<td>74%</td>
<td>208/208</td>
</tr>
<tr>
<td>2017</td>
<td>59%</td>
<td>53%</td>
<td></td>
<td>227/243</td>
</tr>
</tbody>
</table>
## Hazard Module

### Human Health Effects

<table>
<thead>
<tr>
<th>Toxicity: VH - Very High  H - High  M - Medium  L - Low  I - Inconclusive  N/A - Not Applicable  Authority: Authority</th>
<th>Acute Mammalian Toxicity</th>
<th>Neurotoxicity</th>
<th>Systemic Toxicity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Oral</td>
<td>Inhalation</td>
<td>Dermal</td>
</tr>
<tr>
<td>60-35-5</td>
<td>L</td>
<td>I</td>
<td>I</td>
</tr>
<tr>
<td>Acetamide</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>107-13-1</td>
<td>H</td>
<td>H</td>
<td>H</td>
</tr>
<tr>
<td>Acrylonitrile</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1912-24-9</td>
<td>M</td>
<td>H</td>
<td>L</td>
</tr>
</tbody>
</table>
• Metadata ranking of candidates based on mass/formula searching was a starting point

• Searching experimental spectra against *in silico* predictions was next.

• Searching experimental spectra against experimental spectra is the next phase.

• Assembling experimental mass spec data from the internet, homogenize formats and database – includes curation
Database of MS Methods

• There are hundreds of methods distributed across the EPA website – that can be enabled by cheminformatics
  – Focus: Aggregate MS method documents and extract chemicals to make methods structure searchable

  – Vision: Search by structure/substructure/similarity to find existing method(s) as a starting point
Presently working with scientists from NIST

Dataset to be released as a list

Spectra will be added to our experimental spectral database for searching
We have identified phenoxyethanol, several phthalates and phthalate metabolites, and bisphenol A in pooled human urine samples (NIST/SRMs), together with other contaminants in need of further analysis. A similar study was conducted for extracts from orthopedic casts using 16 polymer extracts.

The NIST Mass Spectrometry Data Center is working on a comprehensive approach to the chemical analysis of plastics-related compounds (PRC) using mass spectrometry as a contribution to the NIST circular economy program (https://www.nist.gov/circular-economy).

Due to the versatility of plastics, many manufactured products and environmental pollutants are associated with PRC. 10,547 compounds including monomers, additives, and processing aids of the plastics industry (Environ. Sci. Technol. 2021, 55,9339-9351) and three types of ionizations has been used in this work to obtain mass spectra and build standard libraries of PRC.
Request for participation

- Please point us to relevant datasets and articles
- Nothing is perfect – so please flag issues
Conclusion

• Dashboard access to data for ~900,000 chemicals, next release will update to **1.2 million** substances
• Extractable and leachables lists continue to expand
• “MS-Ready data” facilitates structure identification
• Data continues to grow with ongoing curation activities
• Proof-of-concept developments will release in future versions
• **Do you want to learn more?**
Antony Williams
CCTE, US EPA Office of Research and Development,
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821

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