

ORCID: 0000-0002-2668-4821 Antony Williams I williams.antony@epa.gov I 919-541-1033

Problem Definition and Goals

Problem: Non-targeted analysis (NTA) and suspect screening workflows routinely identify <10% of chemicals observed in environmental samples.

Goals: To develop tools within the freely available and highly curated EPA CompTox Chemistry Dashboard to assist in structure identification to improve NTA and suspect screening workflows. Details regarding methodology and access will be standardized and open to mass spectrometry users.

Abstract

The iCSS CompTox Chemistry Dashboard (https://comptox.epa.gov) is a publicly accessible web-based application provided by the National Center for Computational Toxicology at the US-EPA. It serves a number of purposes, including providing a chemistry database underpinning many of our public-facing projects (e.g. ToxCast and ExpoCast). The available data and searches provide a valuable path to structure identification using mass spectrometry as the source data. With an underlying database of ~720,000 chemicals, the dashboard has already been used to assist in identifying chemicals present in house dust. This poster reviews the benefits of the EPA's platform and underlying algorithms used for the purpose of compound identification using high-resolution mass spectrometry data. Standard approaches for both mass and formula lookup are available but the dashboard delivers a novel approach for hit ranking based on functional use of the chemicals. The focus on highquality data, novel ranking approaches and integration to other resources of value to mass spectrometrists makes the Dashboard a valuable resource for the identification of environmental chemicals.

EPA's CompTox Chemistry Dashboard

Bisphenol A

80-05-7 | DTXSID7020182 0

| Q 🔟 🖪 🕹 - Q - | | Intrinsic Pr | operties | | | | | |
|----------------------------------|----------------|---------------------------|---|-----------------------|----------|---------|----------|----------|
| H ₃ C CH ₃ | | Average | ar Formula: C15H16O2 • Mass: 228.291 g/mol • topic Mass: 228.115030 g/mol | | | | | ľs ľs |
| НО ОН | | Structural Record Infe | | | | | | |
| Chemical Properties | External Links | Synonyms | Product Composition | ToxCast in Vitro Data | Exposure | PubChem | Comments | |

- The US EPA's CompTox Chemistry Dashboard is a public chemistry resource developed to support computational toxicology research efforts within the National Center for Computational Toxicology.
- The data underlying the Dashboard was assembled over a period of more than a decade into the DSSTox database.
- Chemical specific pages contain curated structure information related to a unique DSSTox Identifier (DTXSID), calculated intrinsic properties, structural identifiers, and detailed record and citation information.

Structure Identification Using High Resolution Mass Spectrometry Data and the EPA's CompTox Chemistry Dashboard

Antony Williams^{1,*}, Andrew McEachran², Jon Sobus³

¹U.S.-EPA, Office of Research and Development, National Center for Computational Toxicology (NCCT), Research Triangle Park, NC ²Oak Ridge Institute for Science and Education (ORISE) Participant, Research Triangle Park, NC ³U.S.-EPA, Office of Research and Development, National Exposure Research Laboratory (NERL), Research Triangle Park, NC

Dashboard Chemical Page Features

| Chemical Properties External Links | Synonyms Product Composition ToxCast in | Vitro Data Exposure | PubChem (| Comments | | | | | Chemical Properties External Links Synonyms Product Compo | sition ToxCast in Vitro Data Exposu |
|-------------------------------------|--|---------------------|---------------|--------------|------------------|----------------|----------------------|------------------|---|-------------------------------------|
| Summary | Download as: CSV Excel SDF | | | | | | | | - | Freq |
| Octanol-Water Partition Coefficient | Download as: CSV Excel SDF | | | | | | | | building_material flooring | |
| (LogP) | Property | Average (Exp.) | Median (Exp.) | Range (Exp.) | Average (Pred.) | Median (Pred.) | Range (Pred.) | Result Unit | building_construction | |
| Water Solubility | Octanol-Water Partition Coefficient (LogP) | 3.38 (2) | 3.43 | 3.43 | 3.42 (2) | 3.42 | 3.20 to 3.64 | - | building construction boats | |
| Melting Point | Water Solubility | 5.26e-04 (1) | 5.26e-04 | 5.26e-04 | 2.22e-03 (2) | 2.22e-03 | 7.56e-04 to 3.68e-03 | mol/L | | |
| Boiling Point | Melting Point | 155 (7) | 156 | 153 to 158 | 138 (2) | 138 | 132 to 144 | °C | building_material flooring, building_material walls | |
| Vapor Pressure | Boiling Point | 200 (1) | 200 | 200 | 349 (2) | 349 | 334 to 364 | °C | manufacturing plastics | |
| · | Vapor Pressure | - | - | - | 7.06e-08 (1) | 7.06e-08 | - | mmHg | surface_treatment metals | |
| Soil Adsorption Coefficient | Soil Adsorption Coefficient | - | - | - | 2.92 (2) | 2.92 | 2.74 to 3.10 | - | manufacturing metals | |
| Octanol-Air Partition Coefficient | Octanol-Air Partition Coefficient | - | - | - | 8.39 (1) | 8.39 | - | - | | |
| Atmospheric Hydroxylation Rate | Atmospheric Hydroxylation Rate | - | - | - | -10.4 (1) | -10.4 | - | cm3/molecule*sec | manufacturing machines | |
| Biodegradation Half Life | Biodegradation Half Life | - | - | - | 15.1 (1) | 15.1 | - | days | paint | |
| Bioaccumulation Factor | Bioaccumulation Factor | - | - | - | 173 (1) | 173 | - | - | adhesive | |
| | Bioconcentration Factor | 1.64 (1) | 1.64 | 1.64 | 82.0 (3) | 82.0 | 1.38 to 173 | - | F | |
| Bioconcentration Factor | | | | | | | | | | |

Data under the Chemical Properties tabs include experimental and predicted data (LogP, water solubility, melting points, etc. Predicted properties were modeled using a suite of QSAR models

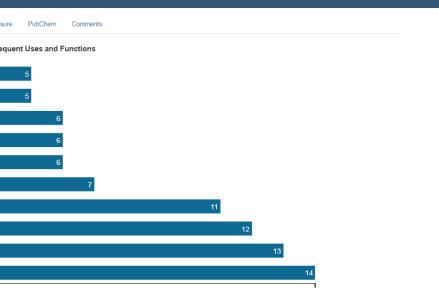
Product composition, frequent use and function data are compiled from US EPA's ToxCast program uses high-throughput screening to the US EPA's Consumer Product Category Database (²CPCat, Dionisio evaluate, predict, and prioritize toxicity of chemicals. Bioassay et al 2015) for over 43,000 chemicals. information. data for all chemicals in ToxCast can be viewed

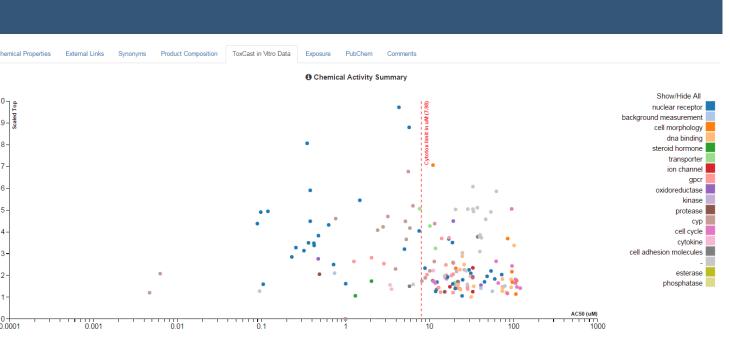
Advanced Searching for Chemical Identification Using Mass Spectrometry Data

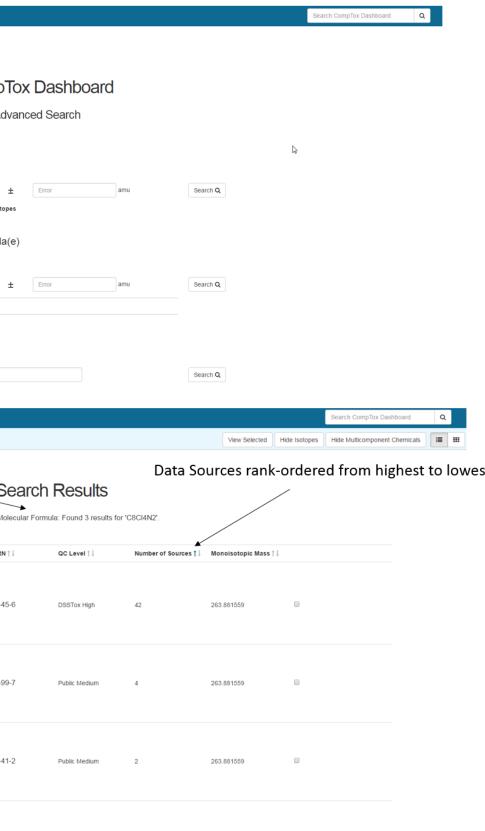
| | ed States commental Protection Home Advanced Search cy | h | | | | | Search CompTox | Dashboard C | ٩ | | itates mental Protection Home Advanced Search | | |
|-----------|--|---|------------------------|-------------------------|--------------------------|------------------------|---------------------|--------------------------------------|-----------------|------------------|---|--|------------------------------------|
| R | | | | | | | | | | | | | |
| - | | | mpTox Da Advanced S | | | | | | | | | C | omp |
| Enter mor | noisotopic mass here | Mass Search ا ف الس MinMax 263.881559 amu | ± .005 ore isotopes | amu | Search | lf mo using mono | re appro a Min/N | priate, ca ⁄lax appro mass wir | oach for | Enter for | mula here | Mass Search Mass amu Single component Generate Molecular F Mass amu Mass amu Options ~ | ^{gnore isotop} Formula |
| | d States onmental Protection Home Advanced Sc | Molecular Formula Sea Molecular Formula Ignore isotopes | arch | | Search | ٩ | 2 | Search CompTox Das | shboard Q | 3 ⊛€₽₩ | ed States connential Protection Home Advanced Sear | Molecular Formula Se CBCHN2 Innore isotopes | earch |
| Numbe | er of results found in | n the search | Search | n Results | Data S | Sources rank | | d from h | ighest to lowes | | er of results found in | the search | S |
| | Download as: CSV + Excel + SDF + | Searched by Mass and s | single component che | micals: Found 22 result | s for '263.881559 ± .005 | 5 amu'. | | | | | Download as: CSV + Excel + SDF + | Searce | hed by Mo |
| | Structure | Preferred Name † 1 | CAS-RN †↓ | QC Level †↓ | Number of Sources † | ↓ Monoisotopic Mass †↓ | Mass Differen †↓ | nce | | | Structure | Preferred Name $\uparrow\downarrow$ | CAS-RN |
| | | Chlorothalonil | 1897-45-6 | DSSTox High | 42 | 263.881559 | -0.0000 | | | | | Chlorothalonil | 1897-45 |
| | | 1,2-Benzenedicarbonitrile, 3,4,5,6-t | 1953-99-7 | Public Medium | 4 | 263.881559 | -0.0000 | | | | | 1,2-Benzenedicarbonitrile, 3,4,5,6-t | 1953-99 |
| 2 | Br CH3 Br | Phenol, 2,4-dibromo-6-methyl- | 609-22-3 | Public High | 3 | 263.878541 | -0.0030 | | | 4 | | p-Phthalodinitrile, tetrachloro- | 1897-41 |

Unknown chemicals can be brought to the top of a list by rank-ordering by the number of associated data sources within Comparison of results of searching and rank-ordering a reference database. Monoisotopic masses or chemical formulae associated with spectral features are searched within by data sources between the Dashboard (top) and the Dashboard and the results rank-ordered. Figs. 1 &2 demonstrate searching an unidentified monoisotopic mass ChemSpider (bottom) split up by compound classes. observed via NTA and rank-ordering the results to bring the most likely chemicals to the top. Figs. 3 & 4 demonstrates the process when a user has already generated a molecular formula.

| | Mass-based | d Searching | Formula-based Searching | | | |
|---------------|------------|-------------|-------------------------|------------|--|--|
| | Dashboard | ChemSpider | Dashboard | ChemSpider | | |
| Average Rank | 1.3 | 2.2 | 1.2 | 1.4 | | |
| Position | | | | | | |
| Percent in #1 | 85% | 70% | 88% | 80% | | |
| Position | | | | | | |



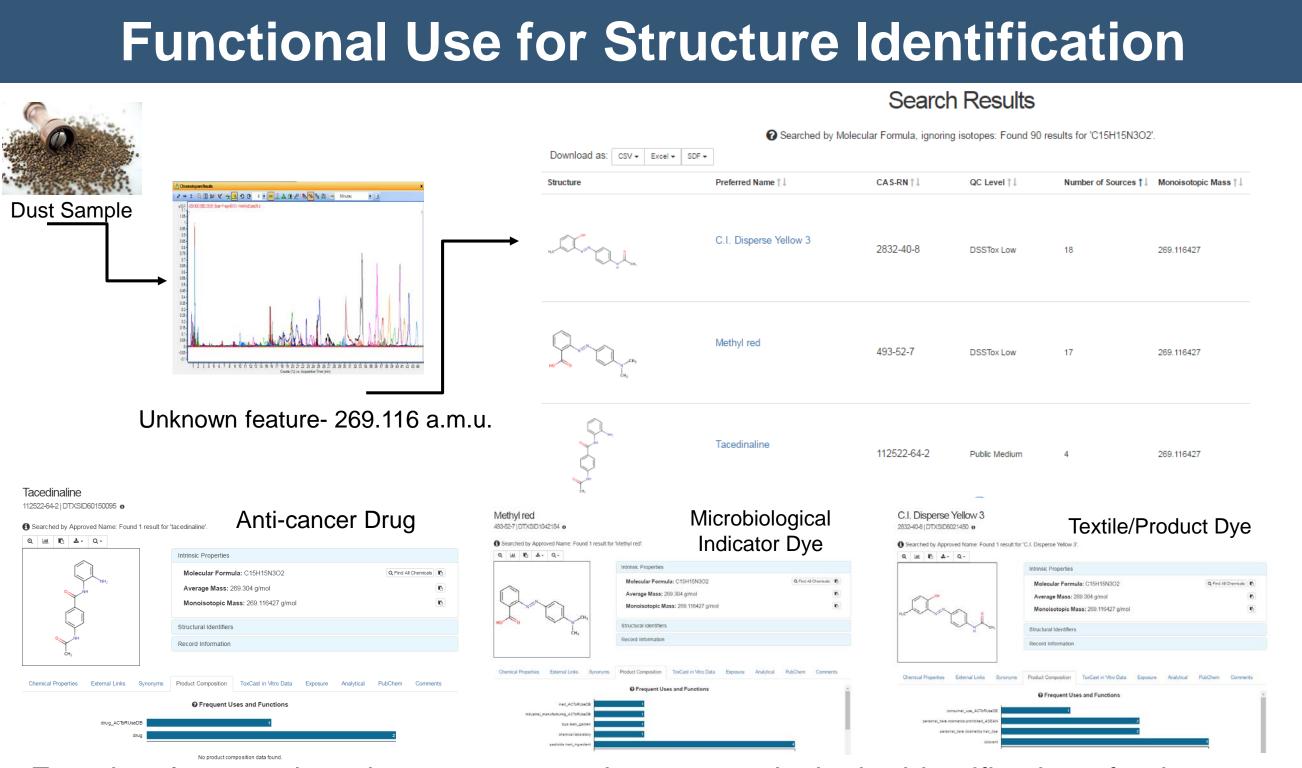




| Compound class | Number in class | Average Rank | Number of compounds in each position rank-ordered | | | | | | |
|--------------------------------|-----------------|-----------------|---|----|----|----|-----|--|--|
| | | | #1 | #2 | #3 | #4 | #5+ | | |
| Pharmaceutical Drug | 72 | 1.3 | 59 | 8 | 3 | 2 | | | |
| Manufacturing Chemicals | 42 | 1.2 | 38 | 1 | 1 | 2 | | | |
| Personal Care Products | 8 | 2.6 | 6 | | | | 2 | | |
| Steroid Hormones | 7 | 1.0 | 7 | | | | | | |
| Perfluorochemicals | 6 | 1.3 | 5 | 1 | | | | | |
| Pesticides | 12 | 1.3 | 10 | 1 | 1 | | | | |
| Veterinary Drugs | 3 | 1.0 | 3 | | | | | | |
| Dyes | 2 | 1.0 | 2 | | | | | | |
| Food product/natural compounds | 4 | 1.5 | 3 | | 1 | | | | |
| Illicit Drugs | 2 | 1.5 | 1 | 1 | | | | | |
| Misc. Molecules | 3 | 1.0 | 3 | | | | | | |

| Compound class | Number in class | Average Rank | Number of compounds in each position rank-ordered | | | | | | |
|--------------------------------|-----------------|-----------------|---|----|----|----|-----|--|--|
| | | | #1 | #2 | #3 | #4 | #5+ | | |
| Pharmaceutical Drug | 72 | 1.4 | 55 | 9 | 6 | 2 | | | |
| Manufacturing Chemicals | 42 | 5.5 | 28 | 6 | 3 | | 5 | | |
| Personal Care Products | 8 | 6.1 | 3 | 1 | | | 4 | | |
| Steroid Hormones | 7 | 1.0 | 7 | | | | | | |
| Perfluorochemicals | 6 | 1.2 | 5 | 1 | | | | | |
| Pesticides | 12 | 2.3 | 6 | 2 | 3 | | 1 | | |
| Veterinary Drugs | 3 | 1.3 | 2 | 1 | | | | | |
| Dyes | 2 | 1.0 | 2 | | | | | | |
| Food product/natural compounds | 4 | 3.8 | 2 | | | 1 | 1 | | |
| Illicit Drugs | 2 | 2.0 | 1 | | 1 | | | | |
| Misc. Molecules | 3 | 1.3 | 2 | 1 | | | | | |

Advanced searching using rank-ordering methods to identify unknowns was conducted n both the Dashboard and in ChemSpider to compare the new application (Dashboard) with the community standard (ChemSpider). 162 unique chemicals were searched by ooth monoisotopic mass and formula and results rank-ordered by data sources. The Dashboard significantly outperformed ChemSpider in terms of the average rank position within each search method. Additionally, the percentage of chemicals in the #1 position was higher for the Dashboard in both cases.



- Functional use and product occurrence data can assist in the identification of unknown chemical features when sample information is known

- When data source count alone cannot definitively distinguish an unknown, the next step is to look at functional use information

- In this example, an unknown feature from a house dust sample led to several potential chemicals but was most likely to be a textile/product dye over an anti-cancer drug or microbiological dye

Future Work

• "MS-Ready Structures":

Mass spectrometry users observe data in spectra as the desalted form of a chemical with no stereochemistry information. These "MS-ready forms" of chemical substances are being incorporated in the Dashboard so that searching for substances is performed on the detectable form of the chemical - the desalted, non-stereo form of the chemical

Non-Targeted Analysis Collaboration Research Trial:

The US EPA is conducting a collaborative research trial with more than 20 labs and institutions from around the world in an effort to develop, improve, and standardize analytical and data processing methods in NTA and suspect screening analysis

References

Distributed structure-searchable toxicity (DSSTox) public database network: a proposal, A.M. Richard, C.R. Williams, Mutat Res. 2002 Jan 29;499(1):27-52. Dionisio, Kathie L., et al. "Exploring consumer exposure pathways and patterns of use for chemicals in the environment." Toxicology Reports 2 (2015): 228-237. McEachran et al. Identifying "Known Unknowns" Using the US EPA's iCSS CompTox Dashboard." In review.