Targeted Lists of Chemicals to Support Non-Targeted Analysis via the US-EPA CompTox Chemicals Dashboard

Antony J. Williams* and Charles Lowe

1. Center for Computational Toxicology & Exposure, Office of Research & Development, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA

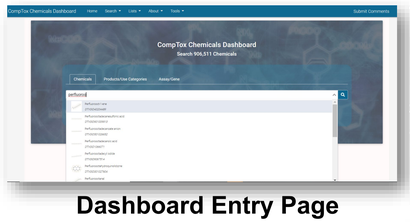
ORCID: 0000-0002-2668-4821
Antony Williams I williams.antony@epa.gov I 919-541-1033

SETAC NTA Meeting
Durham, NC
May 22-25, 2022

Overview

At the US-EPA non-targeted analysis (NTA) uses high-resolution mass spectrometry to better understand the identity of a wide variety of chemicals present in environmental samples. Data processing remains challenging and there is inherent uncertainty in confidently identifying chemicals from candidate lists. We have developed functionality within the CompTox Chemicals Dashboard (available at https://comptox.epa.gov/dashboard) to support this analysis. These tools include the generation of "MS-Ready" structures (1) to optimize database searching, consensus ranking using chemical metadata (2), and *in silico* MS/MS fragmentation prediction for spectral matching (3). A number of chemical lists have also been developed to support specific applications including the identification of chemicals in human breath (i.e., the volatilome), chemicals in human media (e.g., blood, saliva), and in water. Other lists include PFAS (per- and polyfluoroalkyl substances) chemicals and toxins (e.g., microcystins and cyanotoxins). Combining the synergies of a database containing ~900,000 chemicals with over 320 segregated chemical lists with dashboard search functionality provides a comprehensive workflow to support NTA. The Dashboard provides a freely available web-based application to support structure identification and NTA. Expansion of the types of data hosted in the database can support different targeted and non-targeted projects. *This abstract does not reflect EPA policy.*

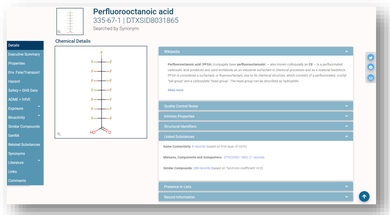
The CompTox Chemicals Dashboard



Dashboard Entry Page

The landing page of the Dashboard is a simple text entry box allowing a type-ahead search for systematic, trade and trivial names, CAS Registry Numbers and InChI chemical identifiers.

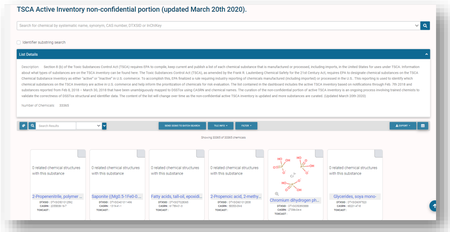
Where possible, links are provided to related Wikipedia articles. Structure file formats are available for download to the desktop (SMILES and molfile) and an executive summary report regarding chemical toxicity is provided.



Chemical Record Page: PFOA

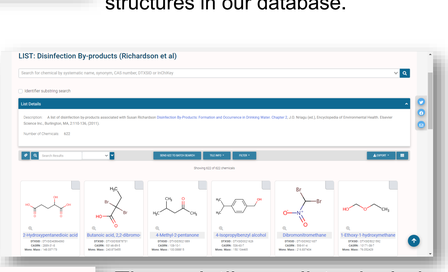
Accessing Chemical Lists

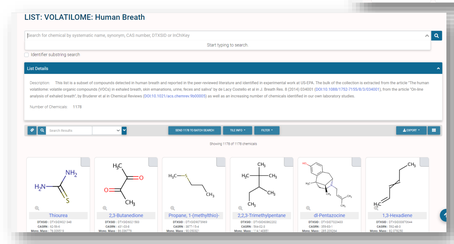
The dashboard provides access to >320 individual chemical lists. The lists are aggregated from publications, public domain databases, assembled from regulatory lists and assembled to support specific research activities in the agency. Example lists are shown below. Some remain under constant expansion (e.g., the list of ToxCast chemicals) or under versioned release (i.e., the PFAS structure list, presently in its fourth iteration).



The TSCA (Toxics Substances Control Act) inventory list is refreshed with every release, often twice a year. The inventory contains a high proportion of UVCB (unknown variable composition or biological substance) chemicals (without structures) but ~~commonly~~ many are mapped to representative or component structures in our database.

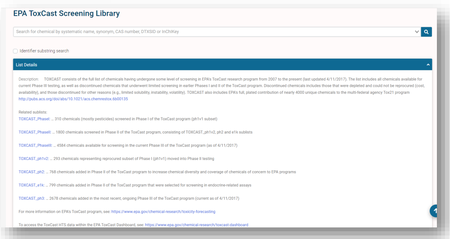
The list of Disinfection by-products is under constant curation and is expanded based on the availability of an updated review article from Susan Richardson at the University of South Carolina.





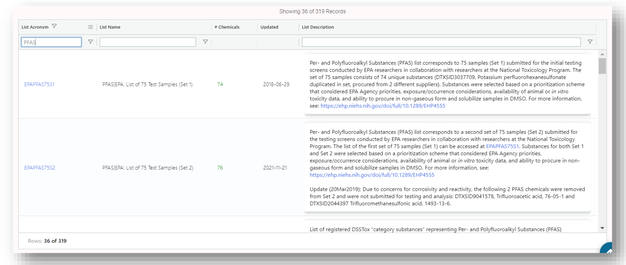
The volatilome lists include chemicals detected in human breath and in saliva. These lists are under constant curation based on new reports and scientific publications published in the literature.

Our Center measures the bioactivity of chemicals in hundreds of assays and adds new chemicals and assays every year. These lists are updated with new chemicals for every release of the Dashboard and link to bioactivity data.



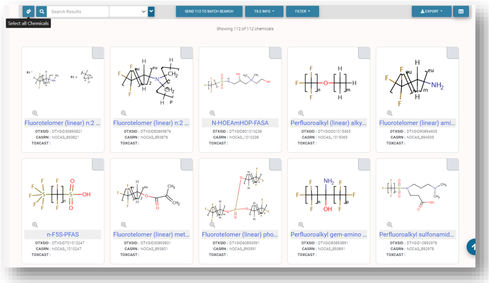
PFAS Chemical Lists

There are >30 individual PFAS lists on the Dashboard. These include a list based on structural filters, UVCB chemicals, PFAS categories represented as Markush structures, and regulatory lists (4).



List of chemical lists of PFAS chemicals: 36 lists and growing

A growing list of Markush structure representations is available. These structures can be enumerated and mapped to distinct chemical structures as members of the categories. We use ChemAxon enumeration tools.



References

1) McEachran et al. (2018) "MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies. *J Cheminform* 10:45.
2) McEachran et al. (2020) Revisiting Five Years of CASMI Contests with EPA Identification Tools. *Metabolites.* 10(6): 260.
3) McEachran *et al.* (2019) Linking *in silico* MS/MS spectra with chemistry data to improve identification of unknowns, *Scientific Data* 6(141)
4) Assembly and curation of lists of per- and polyfluoroalkyl substances (PFAS) to support environmental science research, Williams et al. *Front. Environ. Sci.* 10:850019

Acknowledgements

The authors thank the CCTE chemical curation team for their rigorous work and the Dashboard software development team for their ongoing development of the dashboard. We acknowledge contributions from the community especially to the NORMAN Network