#27

Leveraging the US EPA ToxCast chemical library to benchmark suspect screening and non-targeted analysis methods

Elin M. Ulrich[1,*], Jon Sobus[1], Ann Richard[2], Chris Grulke[2], Antony Williams[2], and Annette Guiseppi-Elie[1]

1. National Exposure Research Laboratory, US EPA, Research Triangle Park, NC USA
2. National Center for Computational Toxicology, US EPA, Research Triangle Park, NC USA

*Ulrich.Elin@epa.gov | 919-541-3717

Abstract

Suspect screening (SSA) and non-targeted analysis (NTA) methods offer practical means to characterize xenobiotic chemicals in a variety of environmental and biological media more efficiently and with broader scope than is possible with targeted methods. These approaches use a variety of analytical instrumentation, data processing methods, acceptance criteria, and reporting standards.

We are conducting a round-robin collaborative trial to evaluate a range of approaches currently used in SSA and NTA. Experiments will evaluate method performance as a function of increasing experimental complexity based on the number of compounds in the mixtures as well as the components in the underlying matrix.

The results will be compared to the actual chemical list to assess the best approaches based on correct identifications, identification certainty, false negatives, and false positives.

GOAL: Produce benchmark method(s) for analytical, reporting, and data analysis to facilitate further analyses and identify areas for improvement.

Background

Exposome

- Totality of enviro exposures throughout lifetime; includes diet, lifestyle, indirect exp. (Wild, 2005)
- 70-90% of disease risk probably environmental (Rappaport and Smith, 2010)
- >84,000 chemicals registered for U.S. use, but little is known about exposure from them

Toxicity ForeCaster (ToxCast™) and Tox21

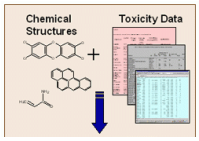
https://www.epa.gov/chemical-research/toxicity-forecaster-toxcasttm-data

- Launched in 2007
- High throughput toxicity screening for hazard prioritization
- Over 3,800 chemicals in EPA's libraries
- Over 1,100 assays on portions of library
- Well curated library of chemicals tested
- Chemical purity QA information available

Distributed Structure-Searchable Toxicity Database (DSSTox)

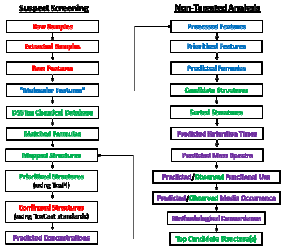
https://www.epa.gov/chemical-research/distributed-structure-searchable-toxicity-dsstox-database

- Includes all ToxCast chemicals (and more!)
- Highly curated with matching:
  - CASRN
  - Structure (SMILES, InChI)
  - Name
- 154,000+ substances with ≥ QC level 4 will be shared as suspect screening list
- DB contains molec. formula for test substance Can calculate monoisotopic mass after desalting
- Access >700K chemicals, physchem properties, advanced searches for structure identification at iCSS Chem Dashboard https://comptox.epa.gov/



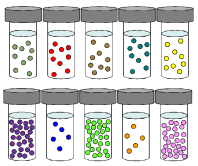
QC Lvl	Description	Chems
1	Curated, validated	4,535
2	Curated, confirmed by multiple sources	16K
3	EPA SRS, no conflicts in ChemID or PubChem	33K
4	ChemID, no conflicts in PubChem	101K
5	ACToR or PubChem	584K
6	Conflicts in public sources	310K

Suspect screening and non-targeted analysis at EPA



- Entails complex workflows combining:
  - Analytical chemistry
  - Data Processing & Analysis
  - Mathematical and QSPR modeling
  - Informatics and Web Services
- Test/evaluate each step, whole process
- Evaluate performance characteristics- % correct, false positive/negative rates

Experimental



Three categories of experiments will be used:

Chemical Standards
- Ten mixtures with high structural diversity
- Known chemicals from ToxCast
- Focus on environ chems with exposure potential
- ~100-400 per mixture, some replicates

Environmental Matrices Unspiked and Spiked
- NIST SRM 2585- Organic contaminants in house dust
- NIST SRM 1957- Organic contaminants in non-fortified human serum
- Silicone passive sampler, environmentally exposed

- Each laboratory will/may use their own:
  - in-house instrumentation
  - methods
  - suspect screening lists
  - data processing

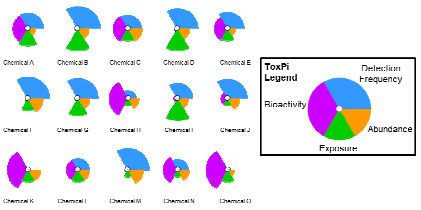
- ~20 groups expressed interest
- **Extracts** of standardized environmental matrices provided to reduce variability
- **Liquid and gas** chromatography used to assess coverage of chemical space
- List of **known chemicals will be disclosed** after initial analyses and reports
- **Individual chemicals** comprising each mixture will be available upon request

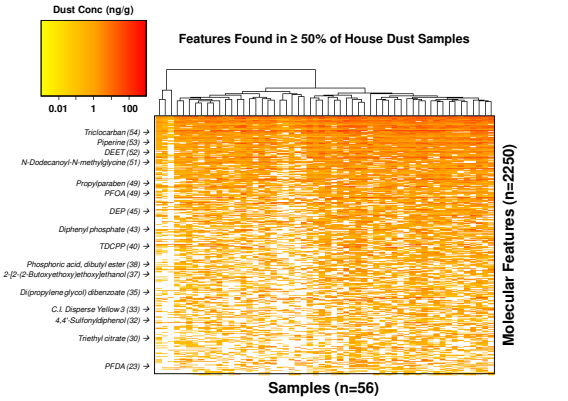
Research Questions

- What percentage of standard mixture chemicals are correctly identified?
- Which methods perform better overall? For specific chemical classes?
- Does the complexity of the mixture/matrix impact performance?
- What types of method/analysis parameters improve performance?
- What chemical space is being covered by each method? Overlap? Can we model these behaviors?
- What can be done to expand coverage?
  - Physicochemical parameters
  - Suspect list
  - Sensitivity
  - Matrix effects
- What unintended components or by-products are in standard mixtures?
  - Impurities
  - Reaction products
  - Degradation products
- In environmental samples, what chemicals do methods agree are present? Does this agree with SRM reported data? Is this predictable?

Example Outputs

- ToxPi prioritizations
- Heat maps
  - molecular features
  - known chemicals
- Concentration estimates
- Hierarchical clustering
  - analysis parameters
  - chemical exposure
- **Benchmark method(s)**





Short Presentation, tentatively in the Auditorium

Monday ~17:15 in workshop #3 "Suspect screening in the environment"