EPA'S Non-Targeted Analysis Research Program: Expanding Public Data Resources in Support of Exposure Science

Elin Ulrich[1], Jon Sobus[1], Ann Richard[2], Chris Grulke[2], John Wambaugh[2], Seth Newton[1], Andrew McEachran[3], and Antony Williams[2]

[1]National Exposure Research Laboratory, US EPA, RTP, NC; [2]National Center for Computational Toxicology, US EPA, RTP, NC; [3]Oak Ridge Institute for Science & Education (ORISE) Research Participant, RTP, NC

Elin M. Ulrich I ulrich.elin@epa.gov I 919-541-3717

United States Environmental Protection Agency

2420

P777

Abstract

Suspect screening (**SSA**) and non-targeted analysis (**NTA**) methods using high-resolution mass spectrometry (HRMS) offer new approaches to efficiently generate exposure data for chemicals in a variety of environmental and biological media. These techniques aid characterization of the exposome and provide critical information on thousands of chemicals in commerce for which exposure data are lacking.

EPA is advancing such techniques with workflows (feature extraction, formula generation, structure prediction, spectral matching, chemical confirmation), and tools (databases; models for predicting retention time, functional use, media occurrence, and media concentration; and schemes for ranking features and chemicals) to rapidly identify, prioritize, and quantify novel compounds in high-interest environmental and biological samples.

EPA is also leading a Non-Targeted Analysis Collaborative Trial (**ENTACT**) to evaluate a range of SSA and NTA approaches. Four categories of experiments are underway, with analyses focused on:

1) ten standard chemical mixtures from the EPA's ToxCast library;
2) extracts of standardized sample matrices (including house dust, human serum, and environmentally deployed silicone passive samplers);
3) extracts of standardized sample matrices spiked with known chemical mixtures; and
4) approximately 4600 single chemicals from the ToxCast library.

More than 20 laboratories worldwide from academia, government, and private (i.e., vendor) organizations are participating. Each laboratory is using their own SSA/NTA methods, and will submit results to EPA for performance evaluation and public release. A project goal is to produce benchmark methods for sample and data analysis, as well as results reporting, and to identify areas of future research. A further outcome of this work will be to identify which analytical methods are more suitable to detecting specific classes of chemicals in environmental media.

Background & Resources

Exposome

- Totality of environmental exposures throughout lifetime; includes diet, lifestyle, indirect exposure [1]
- 70-90% of disease risk estimated environmental [2]
- >84,000 chemicals registered for U.S. use, with little exposure info

Toxicity ForeCaster (ToxCast™) and Tox21

- Launched in 2007, program details manuscript: [3]
- High throughput toxicity screening for hazard prioritization
- Results for over 1,100 assays on portions of library of >4,000 chemicals
- Well-curated library of chemicals with chemical purity QA information

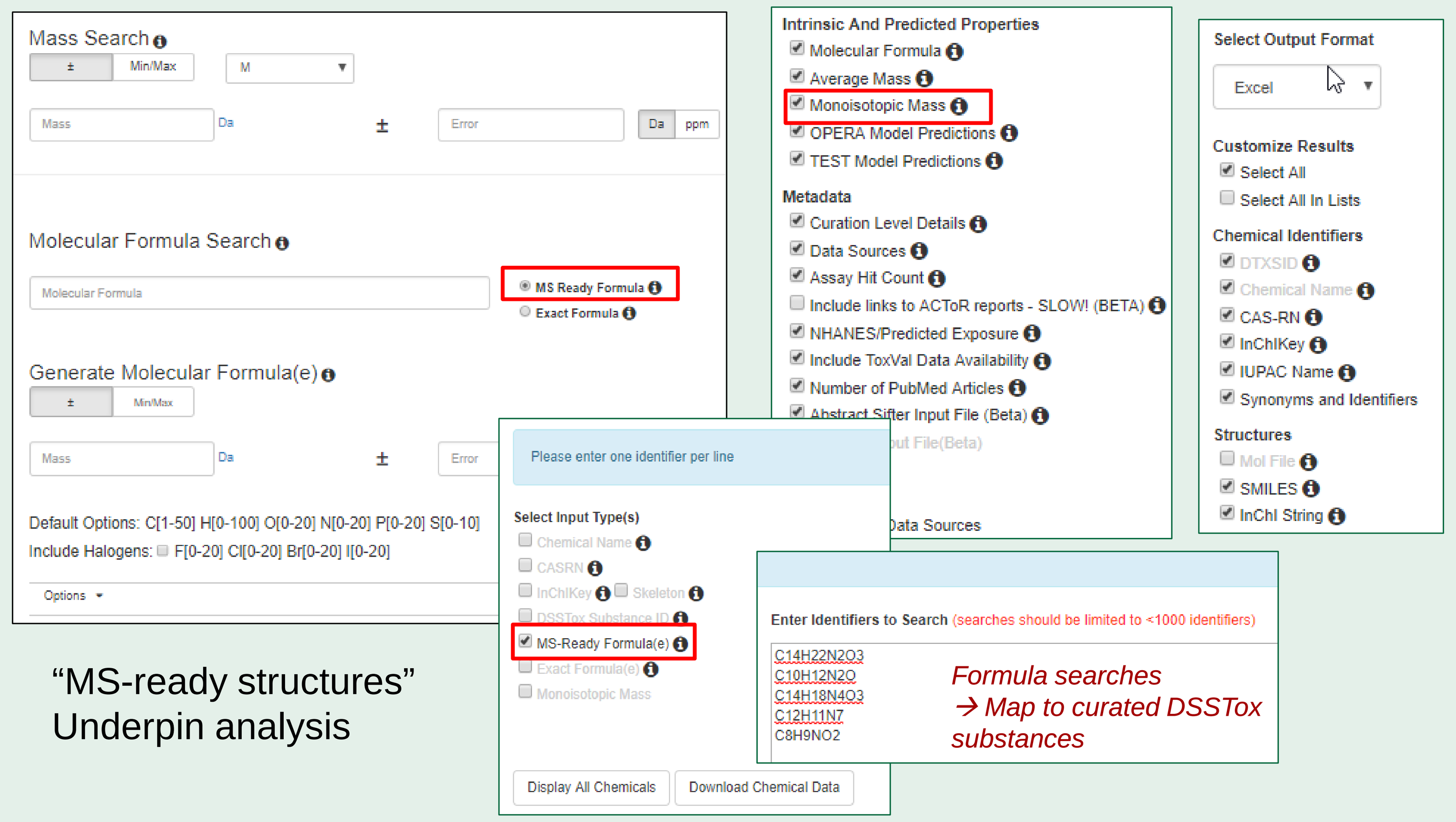
Chemical Structure Resources

Distributed Structure-Searchable Toxicity Database (DSSTox)

- Includes ToxCast chemicals (and more!)
- Highly curated with rule enforcing 1:1:1 mapping of CASRN-Name-Structure (SMILES, InChI)
- 720K+ substances with ≥ Curation QC level 4 shared as SSA list for ENTACT
- DB contains molecular formula for test substance, Monoisotopic mass for desalted formulas

EPA's Chemistry Dashboard https://comptox.epa.gov/dashboard

- Public-portal to DSSTox structure content, lists, downloadable files, predicted & measured phys-chem properties, text-mining, external linkages, and more.
- Advanced searches & tools for structure identification servicing the NTA & MS community:



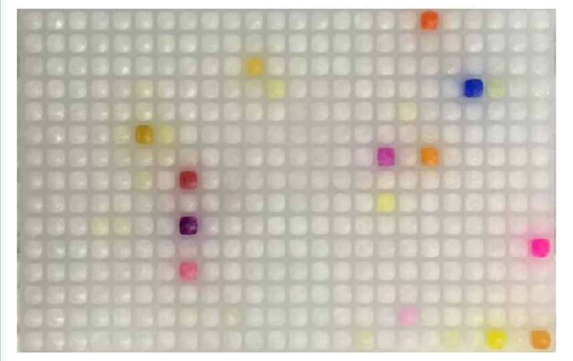
"MS-ready structures" Underpin analysis

Participants

By Sector		By Location	
Academia	15	Canada	1
Government	8	Europe	3
Vendors	5	US	24

Experimental

Three categories of experiments:



Chemical Standards

- Ten mixtures with high structural diversity
- Known chemicals from ToxCast library
- Focus on environmental chemicals with exposure potential
- ~100-400 per mixture, incl. replicates, isobaric cmpds, stereoisomers
- 100 µL aliquots in DMSO, 0.05 mM
- Individual chemicals available upon request by participants

Environmental Matrices: Unspiked & Spiked

- NIST SRM 2585- Organic contaminants in house dust; methanol extract
- NIST SRM 1957- Organic contaminants, non-fortified human serum; acetonitrile extract
- Silicone passive sampler, environmentally exposed; ethyl acetate extract

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

- **Extracts** of standard environmental matrices provided to reduce variability
- **Liquid & gas** chromatography used to assess coverage of chemical space
- List of **known chemicals to be disclosed** after initial analyses & reports

Preliminary Reports of Analysis Methods

Chromatography	Mobile phase	MS type	MS/MS
Dionex LC, Acquity HSS T3	H_2O, ACN, FA	Bruker (Q)TOF ESI +	DDA @ CE 35
Direct infusion	NA	Agilent QTOF nESI/APPI +/-; Thermo FT-ICR ESI +/-	NA, but drift tube ion mobility spectrometry
Agilent LC, Zorbax C8	H_2O, MeOH, AF	Agilent QToF in ESI +/-	DDA @ CEs 10, 20, 40
Dionex LC, XBridge C18	H_2O, MeOH, FA	Thermo Orbitrap QE	DDA @ CE 50
Agilent GC×GC, Rxi-5MS + Rxi-17Silms	Helium	Leco HRT$^+$ in EI and CI for confirmation	NA
Dionex LC, ACE C18-PFP	H_2O, ACN, FA or NH_4OH	Thermo Orbitrap QE in ESI/APCI +/-	DDA @ CEs 15, 30, 45

LC/GC- liq/gas chromatogr; ACN- acetonitrile; MeOH- methanol; FA- formic acid; AF-ammonium formate; QToF- quadrupole time of flight; ESI-electrospray ionization; APPI- atmospheric pressure photoionization; FT-ICR- fourier transform-ion cyclotron resonance; QE- Q Exactive; HRT- high resolution ToF; APCI- atmospheric pressure chemical ionization; DDA- data dependent acquisition; CE- collision energy

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*
  - *Sensitivity*
  - *Suspect list*
  - *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 std ref material reported data? Is this predictable?

Preliminary Results

Mix #	1	2	3	4	5	6	7	8	9	10
Lab 1	132	78	162	142	102	137	105	42	105	81
Lab 2	319	128	384	355	174	392	196	127	537	138
Lab 3	395	197	359	317	221	137	218	90	208	188
Lab 4	39	109	68	137	30	62	124	94	121	107
Lab 5	175	87	197	135	158	156	145	112	186	129
Lab 6	37	34	41	54	41	39	32	29	44	30

Min-75%
76-125%
126-Max

- Shown are total number of "hits" in each of 10 mixtures identified in six laboratories
- Total numbers of actual chemicals in each mixture not yet revealed to participants
- Ranges of under (blue) and over (yellow) reporting totals shown; green are within 25% of actual number
- Potential confounders: sample impurities, reaction products produced within mixture

References

1. CP Wild, "Complementing the Genome with an "Exposome": The Outstanding Challenge of Environmental Exposure Measurement in Molecular Epidemiology" *Cancer Epidemiol. Biomarkers. Prev.* 2005, 14(8), 1847-1850.
2. SM Rappaport and MT Smith, "Environment and Disease Risks" *Science*, 330(6003), 460-461.
3. AM Richard et al. "ToxCast Chemical Landscape: Paving the Road to 21st Century Toxicology" *Chem. Res. Toxicol.*, 2016, 29, 1225-1251.
4. JE Rager et al. "Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring" Environ. Internat., 2016, 88, 269-280.