United States Environmental Protection Agency

ExpoCast
exposure forecasting

NAMs for Exposure: Non-Targeted Analysis

Jon Sobus, Ph.D.

US EPA CSS-HERA Board of Scientific Counselors
Chemical Safety Subcommittee Meeting

February 2-5, 2021

Stakeholder Needs
(OCSPP; EPA Regions)

Chemical safety evaluations require an improved understanding of chemical exposure scenarios and pathways

High-priority exposure data needs → consumer products, indoor environments, occupational settings, ambient environments, ecological pathways

Challenges

- High-quality exposure data are unavailable for many chemicals
- Measurement data traditionally generated using “targeted” methods
- Targeted analytical methods:
  - Require *a priori* knowledge of chemicals of interest
  - Produce data for few selected analytes (10s-100s)
  - Require standards for method development & compound quantitation
  - Are blind to emerging contaminants
  - Can’t keep pace with the needs of 21st century chemical safety evaluations

Research Objective

Rapid Exposure Modeling and Dosimetry Output 2.7:

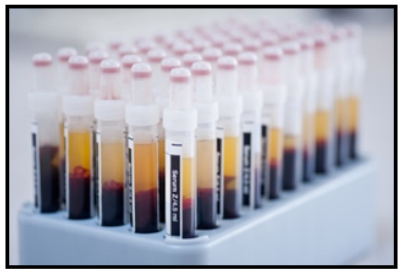
Develop, evaluate, and apply ***non-targeted analysis (NTA)*** methods, alongside targeted monitoring methods, to identify critical sources and pathways of human and ecological exposures

Key Question:

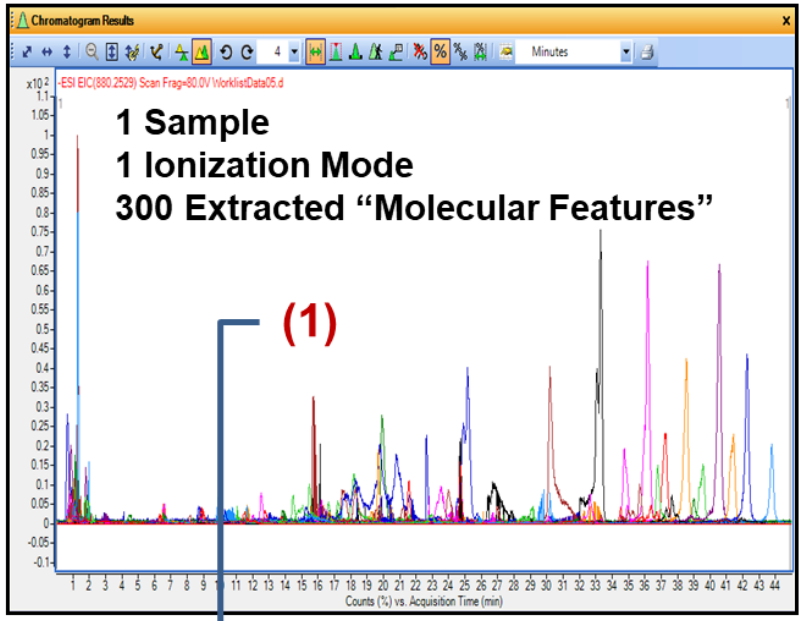
Are NTA methods suitable to meet the needs of 21st century chemical safety evaluations?

General NTA Workflow Steps

Samples



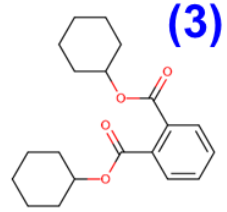
High-Resolution MS



1) Prioritize "molecular features"
2) Correctly assign formulas
3) Correctly assign structures
4) Predict chemical concentrations
5) Determine chemical sources

(2) $C_{20}H_{26}O_4$

(3)



(4) 12 µg/mL

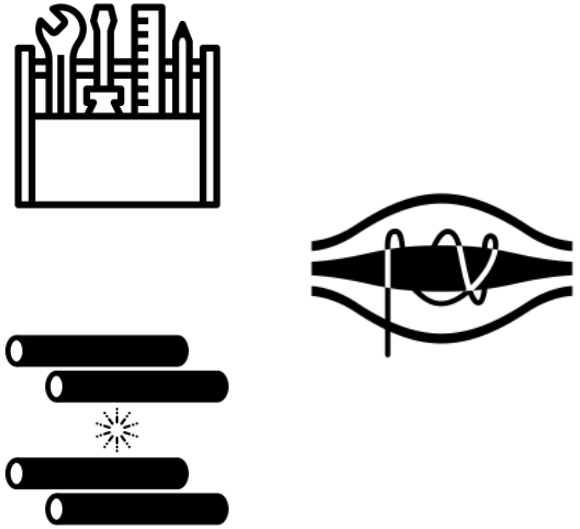
(5)

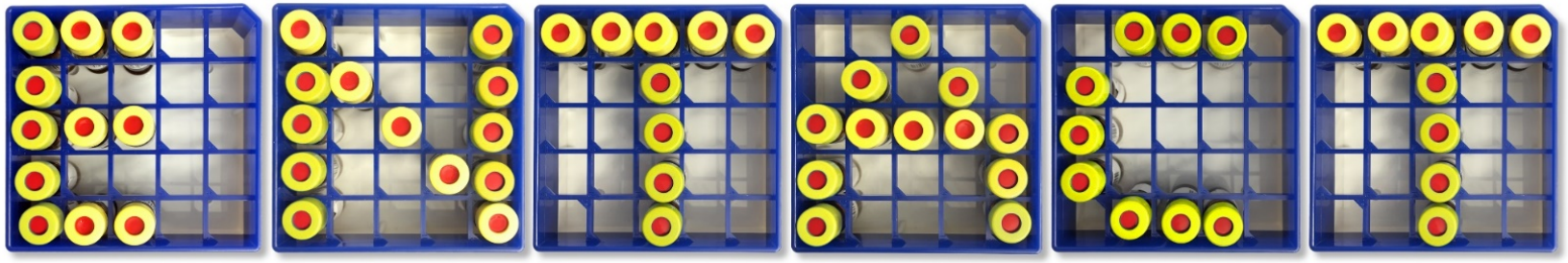
Ongoing Research Activities

- **Evaluate NTA State-of-the-Science**
  - EPA's Non-Targeted Analysis Collaborative Trial (ENTACT)
- **Develop and Disseminate Guidance Materials**
  - Benchmarking and Publications for NTA (BP4NTA)
- **Build Tools to Ensure Transparency & Reproducibility**
  - NTA Study Reporting Tool (NTA SRT)
  - EPA NTA Web Application (NTA WebApp)
- **Address Priority Data Needs with Proof-of-Concept Applications**

Evaluating NTA Science-of-the-Science

- How variable are tools and results from lab to lab?
- Are some methods/workflows better than others?
- How does sample complexity affect performance?
- What chemical space does a given method cover?
- How sensitive are specific instruments/methods?

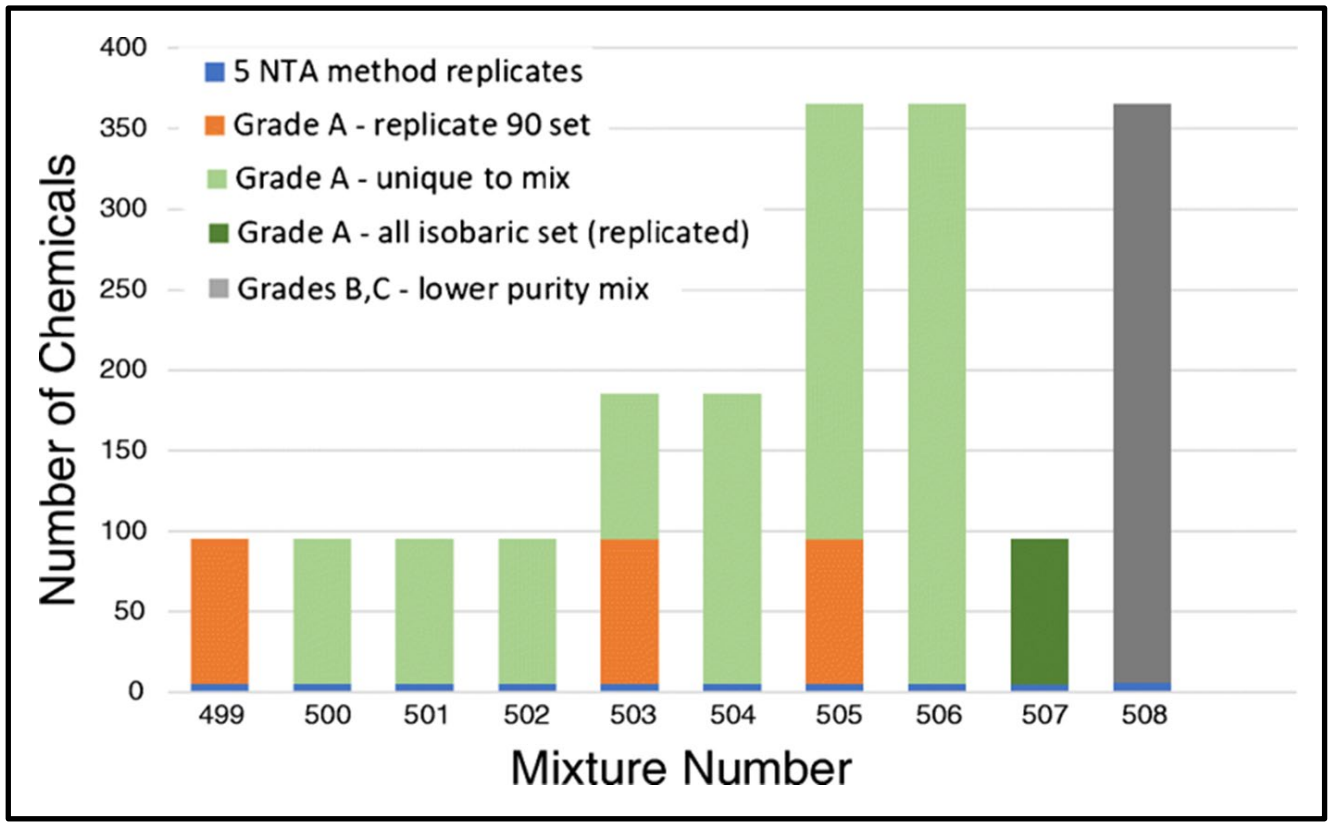




EPA's Non-Targeted Analysis Collaborative Trial

ENTACT Study Design (Part I)

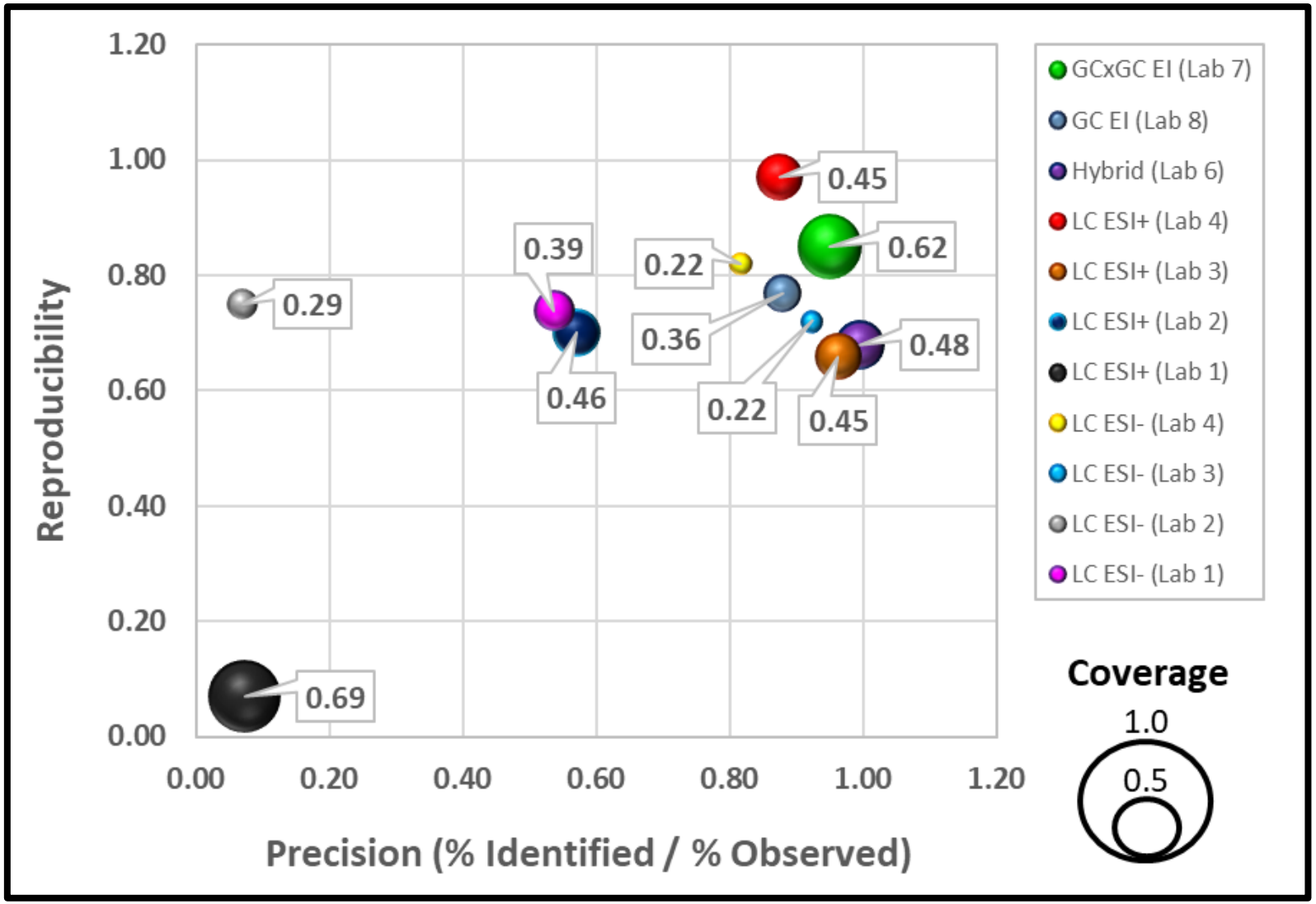
- ~30 global participants, 19 results submitted to date
- 10 synthetic mixtures of ToxCast substances (n=1269)



Replication in substance spikes offers a unique means to assess NTA method reproducibility!

Ulrich et al. 2019. doi: 10.1007/s00216-018-1435-6

Performance Comparison Across Methods



Metrics:

Bubble Size →
What % observed?
(of those spiked)

X-Axis →
What % correct?
(of those observed)

Y-Axis →
What % consistent?
(of those correct)

Take-Away Messages from ENTACT (to date…)

- Lack of transparency in methods/results reporting
- Method procedures change over short time increments
- Biased self-reporting → highlight strengths, mask weaknesses
- Blinded ToxCast mixtures allow for NTA performance assessment
- Standard performance measures highly variable across labs/methods
- Standard performance assessment methods/benchmarks must be adopted
- Benchmarks require input/consensus from NTA community
- Community focus must be on QA/QC and guidance (and innovation)

Developing and Disseminating Guidance Materials

- BP4NTA → Borne out of 2018 ENTACT workshop
- ~100 U.S. and international members
  - Government, academia, and industry
- Working Group Objectives:
  - Short term → define common NTA terms, concepts, and performance metrics
  - Short term → provide recommendations on research & reporting best practices
  - Long term → establish proficiency testing levels (ASTM/ISO)
- Products (including 3 manuscripts):
  - Website with key resources and links: https://nontargetedanalysis.org/
  - Guidance documents with definitions & supporting info
  - "NTA Study Reporting Tool" to standardize reporting (proposals & manuscripts)

Building Tools to Ensure Transparency & Reproducibility

The "NTA Study Reporting Tool" (NTA SRT):

- Standardized framework for reviewing quality of NTA reporting
- Aids NTA study design and review (proposals & manuscripts)
- Follows chronology of typical NTA studies with detailed examples
- Scale-based scoring (numeric & colorimetric) for individual study attributes
- HTML interactive version via BP4NTA website (hyperlinks → supporting docs.)
- Fillable PDF version available for download (via website)
- Comment box for periodic updates/revisions (via website)
- Working with journal editors for initial testing and deployment

NTA Study Reporting Tool (draft version)

NTA Study Chronology

Study Sections & Categories			Example Information to Report	Numeric & Colorimetric Scoring	Rationale/Notes
Methods	Study Design	Objectives & Scope	• Study goals and hypotheses • Scope of the study with respect to use of NTA / suspect screening • Expected chemical coverage of approach and potential limitations	1	
		Sample Information & Preparation	• Sample collection/replication, handling/storage, preparation, extraction, & clean-up methods (and related QA practices) • Intended use of samples (e.g., method development, compound identification, etc.) • Development and intended use of blanks	2	
		QC Spikes & Controls	• Development of spikes/controls (e.g., isotopically labeled standards/spikes, native standard spikes, matrix pools) • Intended use of QC or other spikes/controls (e.g., to monitor instrument performance, data normalization, etc.)	2	
	Data Acquisition	Analytical Sequence	• Sample randomization and use of replicate injections • Inclusion of blanks and QC samples in the acquisition sequence • Information about single vs. multiple analytical batches	3	
		Chromatography	• Instrument specifications • Method settings (e.g., column/guard, mobile phases, gradient, injection techniques)	3	
		Mass Spectrometry	• Instrument specifications • Instrument calibration and/or tuning procedures • Method settings (e.g.,	3	
			• File conversion inform • Software program(s) u • Workflow steps (e.g., p • Feature detection thres • Data correction or nor	2	
			• Software programs(s) p • Basic statistical analys settings/thresholds • Chemometric analysis	NA	
			• Software program(s) u • Libraries and databases • Workflow steps (e.g., p • Workflow methods &	2	
Results	Data Outputs	Statistical & Chemometric Outputs	• Basic statistical outputs (e.g., adj. p-values, standard deviations, test statistics) • Results of chemometric analyses (e.g., reported classifications/groupings of features or samples, observed trends in the data) • Visuals/plots (e.g., Venn diagrams, heatmaps, clustering dendrograms, volcano plots, network diagrams, PCA and loading plots) • New statistical metrics, algorithms, packages, and/or scripts	NA	
		Identification & Confidence Levels	• Reported identifications and associated confidence levels (e.g., levels described by Schymanski et al.) • Supporting data for annotation/identification (e.g., formula match scores, fine isotope pattern, retention time match, MS/MS match scores, source of MS/MS spectra) • For features with lower confidence IDs, (i.e., not standard-confirmed), proposed tentative structures and other annotated data • Semi-quantification or quantification data • Exported MS/MS spectra (e.g., as a library, database, or deposition into online repository)	3	
	QA/QC Metrics	Data Acquisition QA/QC	• Quality: Adherence to QA/QC protocols for sample preparation and data acquisition • Boundary: Description of the potential impacts of methods (sample prep, chromatographic, MS) on observable chemical space • Accuracy: Reported chromatographic and mass accuracy • Precision: Variability of observed retention time, precursor mass error, and abundance	1	
		Data Processing & Analysis QA/QC	• Quality: Outcomes of QC checks along the data processing & analysis workflow • Boundary: Impact of data processing & analysis method(s) on observed chemical space, observed limits of detection/ID • Accuracy: Performance measures (True Positive Rate, False Positive Rate, etc.) for known compounds or samples with known classification • Precision: Reproducibility/repeatability of performance measures for known compounds or samples with known classification; Calculations such as False Discovery Rate, F1 score, etc.	0	

Hyperlinked (HTML version) to supporting information

3-4 bullet point examples for each of the 13 sub-categories

Not exhaustive – intended to guide reviewers; relies on reviewer expertise/discretion.

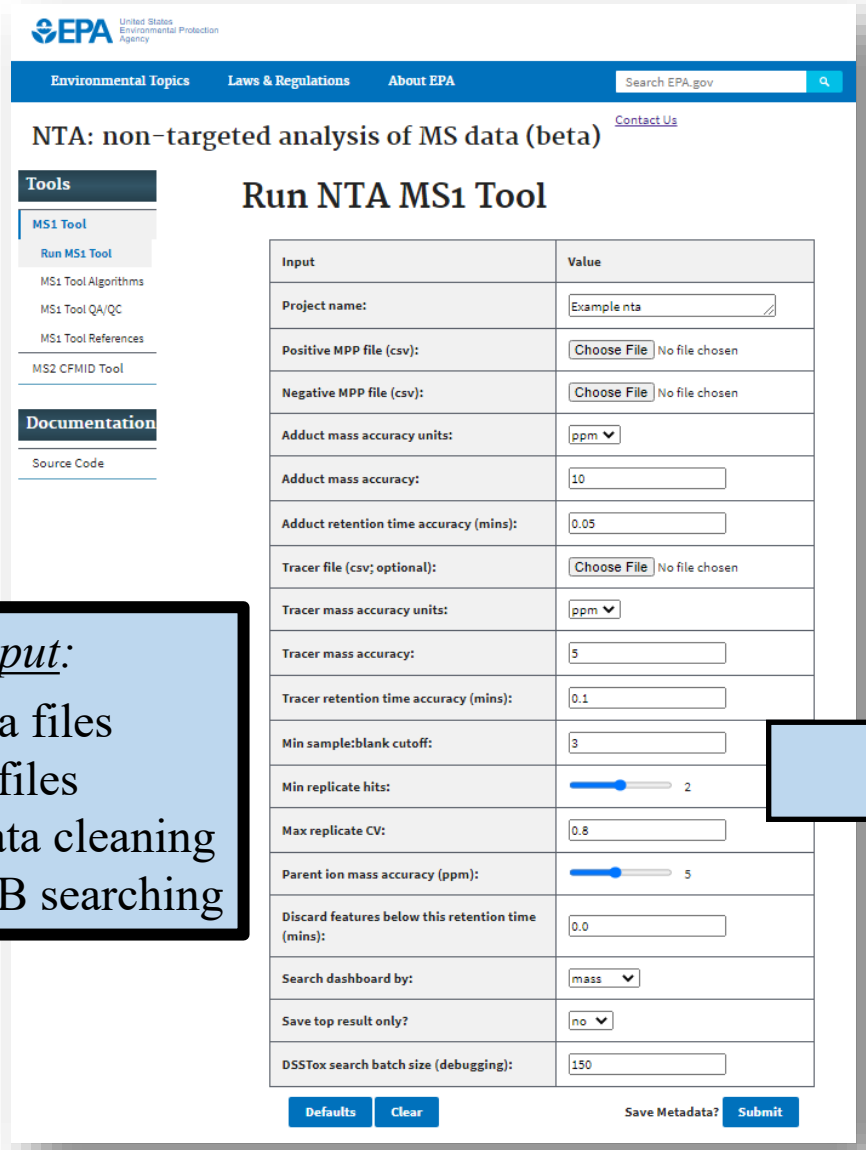
Space for reviewer to explain assigned score

Building Tools to Ensure Transparency & Reproducibility

The EPA NTA WebApp:

- Queries NTA data against DSSTox DB (~900K substances)
- Aggregates metadata to aid candidate prioritization
- Calculates match metrics to aid candidate filtering
- Provides interactive visualization of chemical candidates
- Processes data for advanced statistical analyses
- Standardizes and documents procedures for NTA data analysis
- Adheres to recommendations from BP4NTA workgroup
- Produces publication-ready output in accordance with NTA SRT

EPA's NTA WebApp

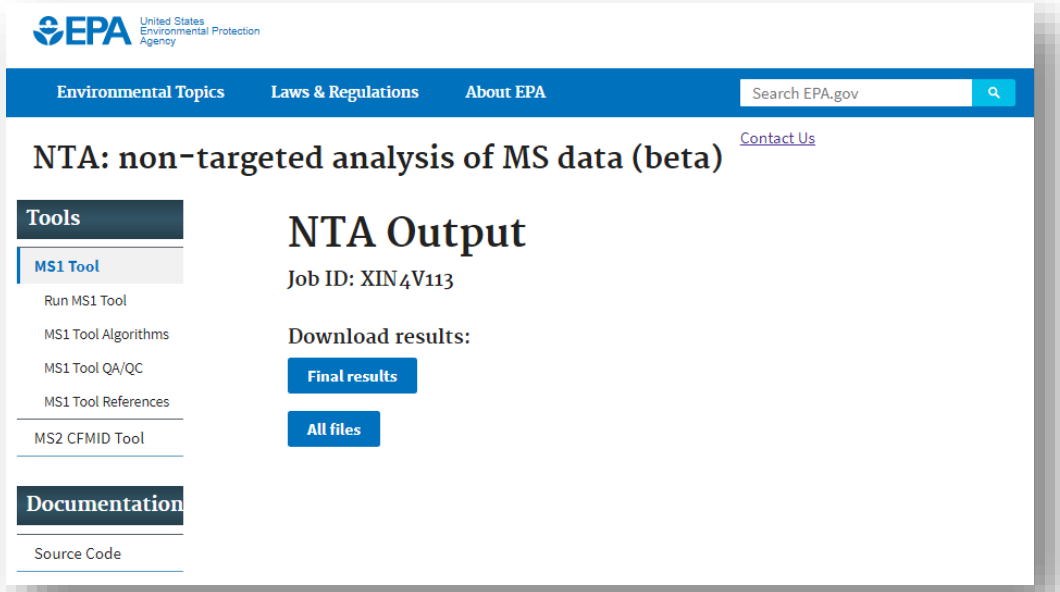


WebApp Input:

- Experimental data files
- Tracer (QA/QC) files
- Parameters for data cleaning
- Parameters for DB searching

WebApp Output:

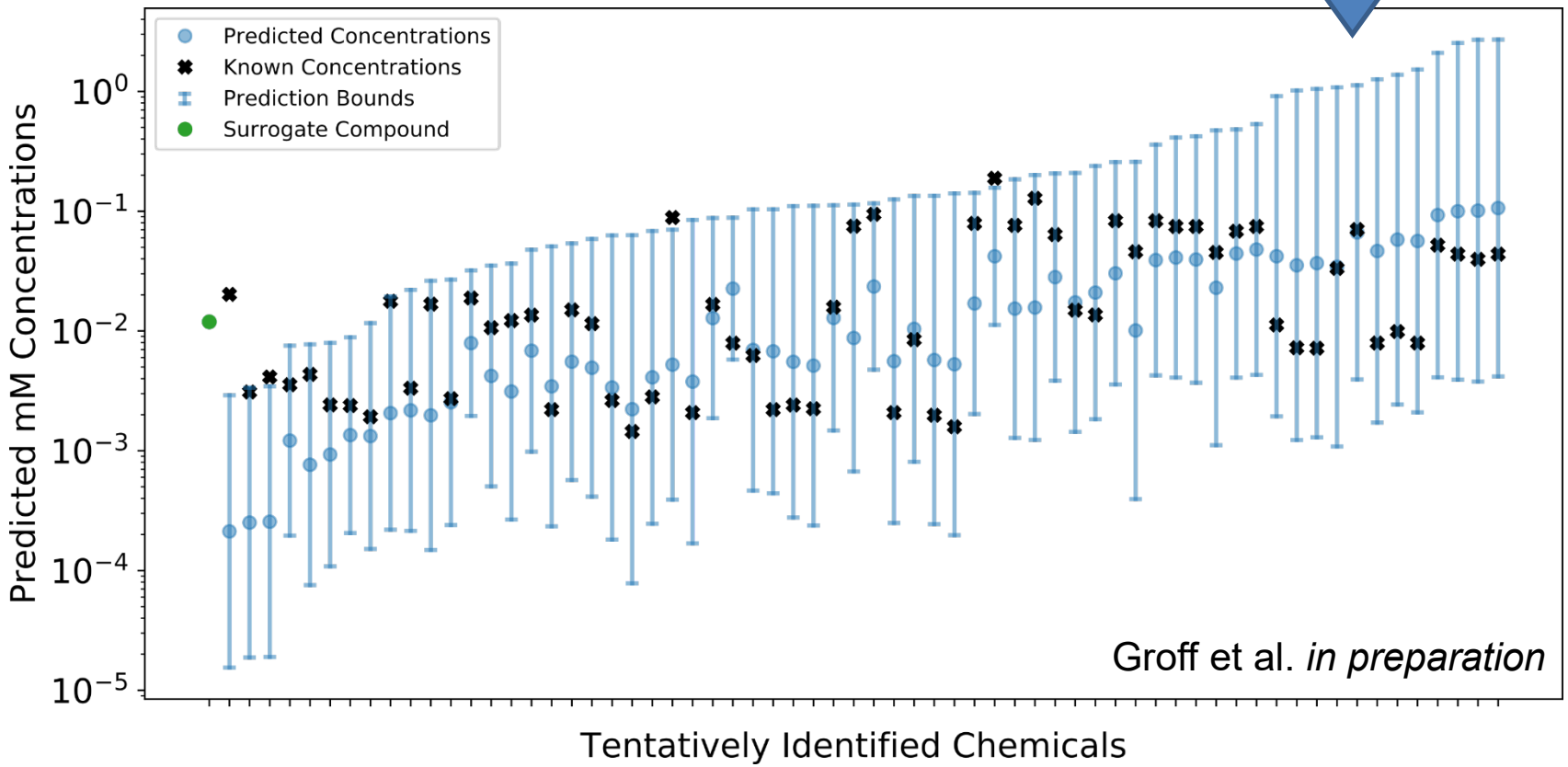
- QA/QC tracer results
- Cleaned, unannotated file for stats analysis
- Cleaned, annotated file with DSSTox chemicals
- Complete file with all chemicals & metadata



Addressing High Priority Data Needs with Proof-of-Concept Applications

- Characterizing chemical contents of products (including UVCBs)
- Characterizing data-poor xenobiotics in biological tissues & fluids
- Identifying xenobiotic metabolites produced from *in vitro* assays
- Developing semi-quantitative (SQ) methods for risk-based interpretation
- Characterizing emerging contaminants in Brita filters (SQ proof-of-concept)
- Developing a framework for rapid response NTA

SQ NTA Proof-of-Concept



- Analysis of Brita filter extracts via GC-HRMS.
- Concentration estimates can be above or below true value.
- Prediction intervals used to bound SQ concentration estimates.
- 95% prediction intervals shown; Can use 99%, 99.9%, etc.
- Tentatively identified compounds ranked by upper bound estimates.
- Upper bound estimates compared to level-of-interest to set priorities.
- Priority compounds further examined using targeted methods.

Contributing Researchers
(EPA Affiliation Unless Otherwise Noted)

- **ENTACT**:
  - ***Co-leads:*** E. Ulrich and J. Sobus
  - ***Research Team:*** A. Williams, A. Chao, S. Newton, C. Lowe, C. Grulke, A. Richard, J. Grossman (ORISE)
- **BP4NTA**:
  - ***Overall Co-leads:*** E. Ulrich and B. Place (NIST)
  - ***Website Co-leads:*** S. Newton and S. Nason (CAES)
- **NTA SRT**:
  - ***Co-leads:*** K. Peter (NIST) and A. Phillips
  - ***Research Team:*** P. Gardinali (FIU), A. Knolhoff (FDA), C. Manzano (SDSU), K. Miller, M. Pristner & B. Warth (U. of Vienna), L. Sabourin & M. Sumarah (Agri-Food Canada), J. Sobus
- **NTA WebApp**:
  - ***Research Team:*** J. Minucci, A. Chao, T. Purucker, A. Williams, J. McCord, H. Al-Ghoul (ORISE), M. Russell, C. Lowe, L. Groff (ORISE), J. Sobus
- **SQ NTA**:
  - ***Research Team:*** L. Groff (ORISE), H. Liberatore, J. McCord, S. Newton, E. Ulrich, J. Sobus

Additional EPA Contributors

This work was supported, in part, by ORD's Pathfinder Innovation Program (PIP) and an ORD EMVL award

EPA ORD

Hussein Al-Ghoul*
Alex Chao
Jacqueline Bangma*
Matthew Boyce*
Kathie Dionisio
Louis Groff*
Jarod Grossman*
Chris Grulke
Kristin Isaacs
Sarah Laughlin*
Hannah Liberatore
Charles Lowe
Kamel Mansouri*
Aurelie Marcotte*
James McCord
Andrew McEachran*
Kelsey Miller
Jeff Minucci

EPA ORD (cont.)

Seth Newton
Grace Patlewicz
Allison Phillips
Katherine Phillips
Tom Purucker
Ann Richard
Randolph Singh*
Jon Sobus
Mark Strynar
Elin Ulrich
Ariel Wallace
John Wambaugh
Antony Williams

GDIT

Ilya Balabin
Tom Transue
Tommy Cathey

* = ORISE/ORAU

Relevant EPA NTA Publications

1) Rager JE, Strynar MJ, Liang S, McMahen RL, Richard AM, Grulke CM, Wambaugh JF, Isaacs KK, Judson R, Williams AJ, Sobus JR. Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring. Environ Int. 2016 Mar;88:269-280.
2) McEachran AD, Sobus JR, Williams AJ. Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard. Anal Bioanal Chem. 2017 Mar;409(7):1729-1735.
3) Newton SR, McMahen RL, Sobus JR, Mansouri K, Williams AJ, McEachran AD, Strynar MJ. Suspect screening and non-targeted analysis of drinking water using point-of-use filters. Environ Pollut. 2018 Mar;234:297-306.
4) Sobus JR, Wambaugh JF, Isaacs KK, Williams AJ, McEachran AD, Richard AM, Grulke CM, Ulrich EM, Rager JE, Strynar MJ, Newton SR. Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. J Expo Sci Environ Epidemiol. 2018 Sep;28(5):411-426.
5) Phillips KA, Yau A, Favela KA, Isaacs KK, McEachran A, Grulke C, Richard AM, Williams AJ, Sobus JR, Thomas RS, Wambaugh JF. Suspect screening analysis of chemicals in consumer products. Environ Sci Technol. 2018 Mar 6;52(5):3125-3135.
6) McEachran AD, Mansouri K, Newton SR, Beverly BEJ, Sobus JR, Williams AJ. A comparison of three liquid chromatography (LC) retention time prediction models. Talanta. 2018 May 15;182:371-379.
7) Ulrich EM, Sobus JR, Grulke CM, Richard AM, Newton SR, Strynar MJ, Mansouri K, Williams AJ. EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal Bioanal Chem. 2019 Feb;411(4):853-866.
8) Sobus JR, Grossman JN, Chao A, Singh R, Williams AJ, Grulke CM, Richard AM, Newton SR, McEachran AD, Ulrich EM. Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. Anal Bioanal Chem. 2019 Feb;411(4):835-851.
9) Catron TR, Swank A, Wehmas LC, Phelps D, Keely SP, Brinkman NE, McCord J, Singh R, Sobus J, Wood CE, Strynar M, Wheaton E, Tal T. Microbiota alter metabolism and mediate neurodevelopmental toxicity of 17β-estradiol. Sci Rep. 2019 May 8;9(1):7064.
10) McEachran AD, Balabin I, Cathey T, Transue TR, Al-Ghoul H, Grulke C, Sobus JR, Williams AJ. Linking in silico MS/MS spectra with chemistry data to improve identification of unknowns. Sci Data. 2019 Aug 2;6(1):141.
11) Nuñez JR, Colby SM, Thomas DG, Tfaily MM, Tolic N, Ulrich EM, Sobus JR, Metz TO, Teeguarden JG, Renslow RS. Evaluation of in silico multifeature libraries for providing evidence for the presence of small molecules in synthetic blinded samples. J Chem Inf Model. 2019 Sep 23;59(9):4052-4060.

Publications (cont.)

12) Weitekamp CA, Phelps D, Swank A, McCord J, Sobus JR, Catron T, Keely S, Brinkman N, Zurlinden T, Wheaton E, Strynar M, McQueen C, Wood CE, Tal T. Triclosan-selected host-associated microbiota perform xenobiotic biotransformations in larval zebrafish. Toxicol Sci. 2019 Sep 5;172(1):109–122.
13) Pleil JD, Wallace MAG, McCord J, Madden MC, Sobus J, Ferguson G. How do cancer-sniffing dogs sort biological samples? Exploring case-control samples with non-targeted LC-Orbitrap, GC-MS, and immunochemistry methods. J Breath Res. 2019 Nov 19;14(1):016006.
14) Chao A, Al-Ghoul H, McEachran AD, Balabin I, Transue T, Cathey T, Grossman JN, Singh RR, Ulrich EM, Williams AJ, Sobus JR. In silico MS/MS spectra for identifying unknowns: a critical examination using CFM-ID algorithms and ENTACT mixture samples. Anal Bioanal Chem. 2020 Feb;412(6):1303-1315.
15) Newton SR, Sobus JR, Ulrich EM, Singh RR, Chao A, McCord J, Laughlin-Toth S, Strynar M. Examining NTA performance and potential using fortified and reference house dust as part of EPA's Non-Targeted Analysis Collaborative Trial (ENTACT). Anal Bioanal Chem. 2020 Jul;412(18):4221-4233.
16) Singh RR, Chao A, Phillips KA, Xia XR, Shea D, Sobus JR, Schymanski EL, Ulrich EM. Expanded coverage of non-targeted LC-HRMS using atmospheric pressure chemical ionization: A case study with ENTACT mixtures. Anal Bioanal Chem. 2020 Aug;412(20):4931-4939.
17) McEachran AD, Chao A, Al-Ghoul H, Lowe C, Grulke C, Sobus JR, Williams AJ. Revisiting five years of CASMI contests with EPA identification tools. Metabolites. 2020 Jun 23;10(6):260.
18) Abrahamsson DP, Sobus JR, Ulrich EM, Isaacs K, Moschet C, Young TM, Bennett DH, Tulve NS. A quest to identify suitable organic tracers for estimating children's dust ingestion rates. J Expo Sci Environ Epidemiol. 2020 Jul 13.
19) Washington JW, Rosal CG, Ulrich EM, Jenkins TM. Use of carbon isotopic ratios in nontargeted analysis to screen for anthropogenic compounds in complex environmental matrices. J Chromatogr A. 2019 Jan 4;1583:73-79.
20) McEachran AD, Hedgespeth ML, Newton SR, McMahen R, Strynar M, Shea D, Nichols EG. Comparison of emerging contaminants in receiving waters downstream of a conventional wastewater treatment plant and a forest-water reuse system. Environ Sci Pollut Res Int. 2018 May;25(13):12451-12463.
21) Williams AJ, Sobus JR. Applications of the US EPA CompTox Chemicals Dashboard to support mass spectrometry and breath research. Breathborne Biomarkers and the Human Volatilome. 2019 [in press]; C. Davis, J. Beauchamp, and J. Pleil (ed), Elsevier Inc.
22) McEachran AD, Mansouri K, Grulke C, Schymanski EL, Ruttkies C, Williams AJ. "MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies. J Cheminform. 2018 Aug 30;10(1):45.

Questions?

sobus.jon@epa.gov

The views expressed in this presentation are those of the author and do not necessarily represent the views or policies of the U.S. Environmental Protection Agency.