Scientific Advances in Non-Targeted Chemistry

Elin M. Ulrich

U.S. Environmental Protection Agency
Center for Computational Toxicology and Exposure
Research Triangle Park, NC USA

Outline

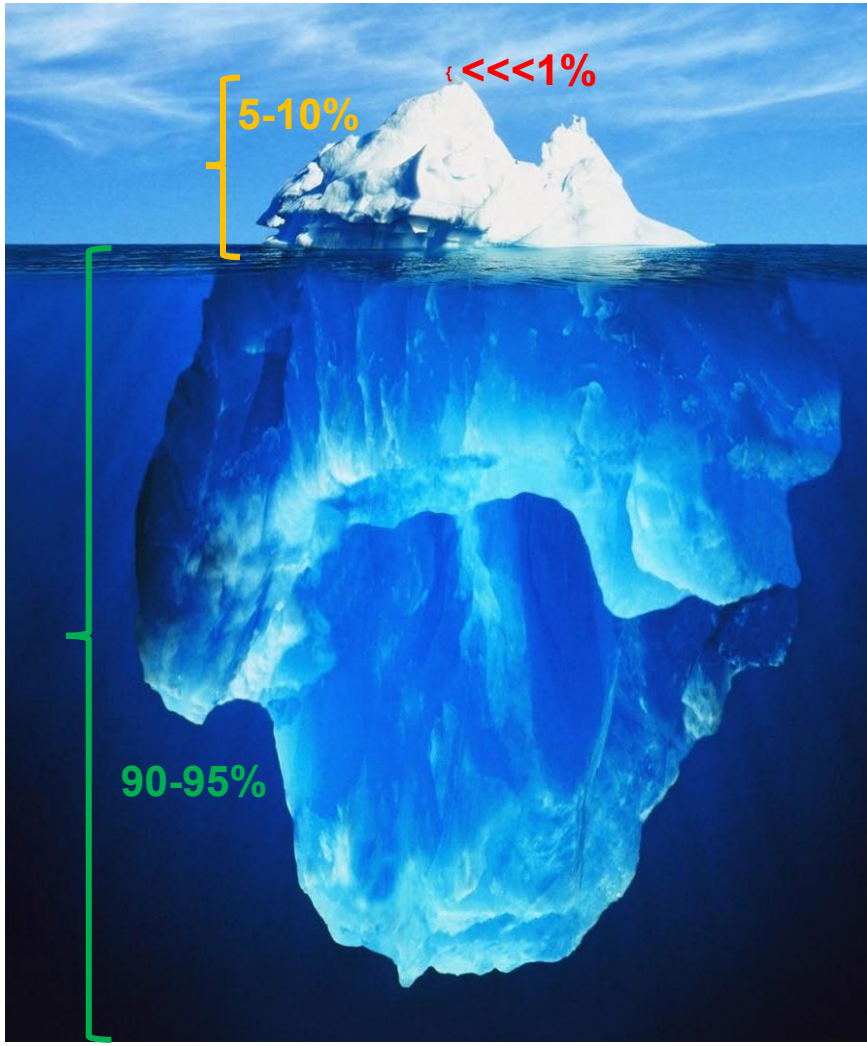
- What is Non-Targeted Analysis?
- Benchmarking and Publications for Non-Targeted Analysis
- Uses of NTA data
- Future of Non-Targeted Analysis

What is Non-Targeted Analysis?

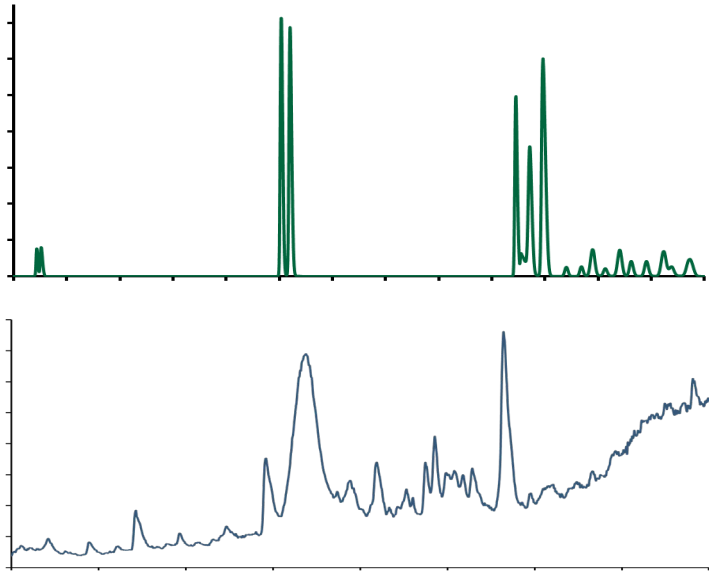
- **Targeted Analysis**
  - “known knowns”
  - Standards, calibration curves

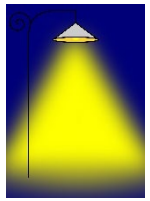
- **Suspect Screening Analysis (SSA)**
  - “known unknowns”
  - Lists of compounds

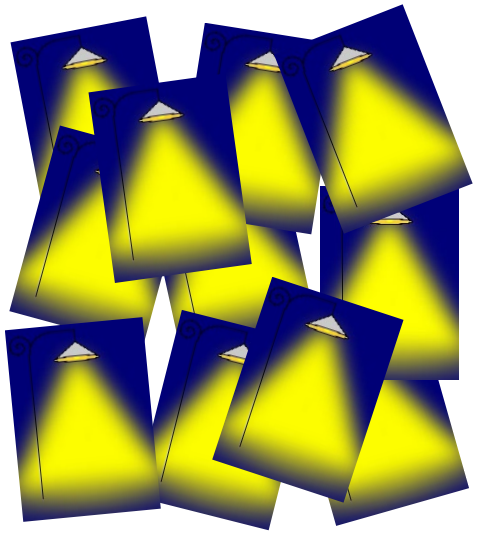
- **Non-Targeted Analysis (NTA)**
  - “unknown unknowns”
  - MS first principles

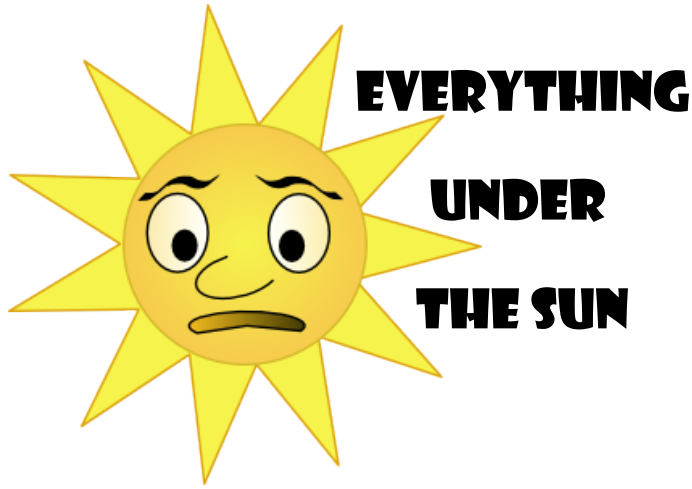


Targeted vs. Non-Targeted Analysis









Targeted **Suspect Screening** **Non-Targeted**

- Difficulty/Time
- Retrospective mining

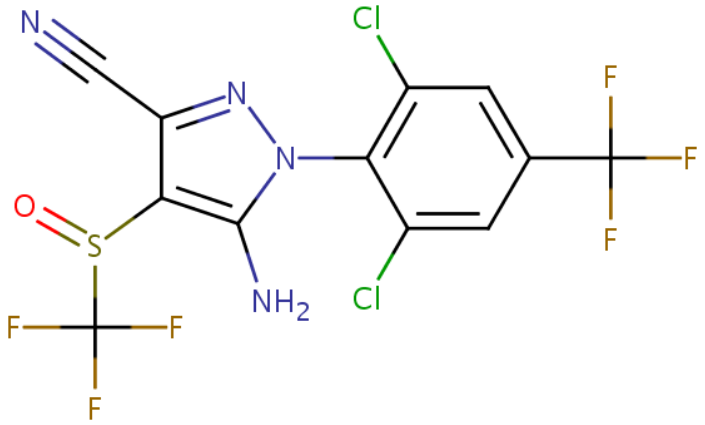
- Quantitative info
- Structure confidence

How does High Resolution MS work?

Atom	Natural Abundance	Exact Mass
^{1}H	99.9885%	1.007825
^{2}H	0.0115%	2.014102
^{12}C	98.93%	12.000000
^{13}C	1.07%	13.003355
^{14}N	99.632%	14.003074
^{15}N	0.368%	15.000109
^{16}O	99.757%	15.994915
^{17}O	0.038%	16.999131
^{18}O	0.205%	17.999159
^{19}F	100%	18.998403
^{32}S	94.93%	31.972072
^{33}S	0.76%	32.971459
^{34}S	4.29%	33.967868
^{36}S	0.02%	35.967079
^{35}Cl	75.78%	34.968853
^{37}Cl	24.22%	36.965903

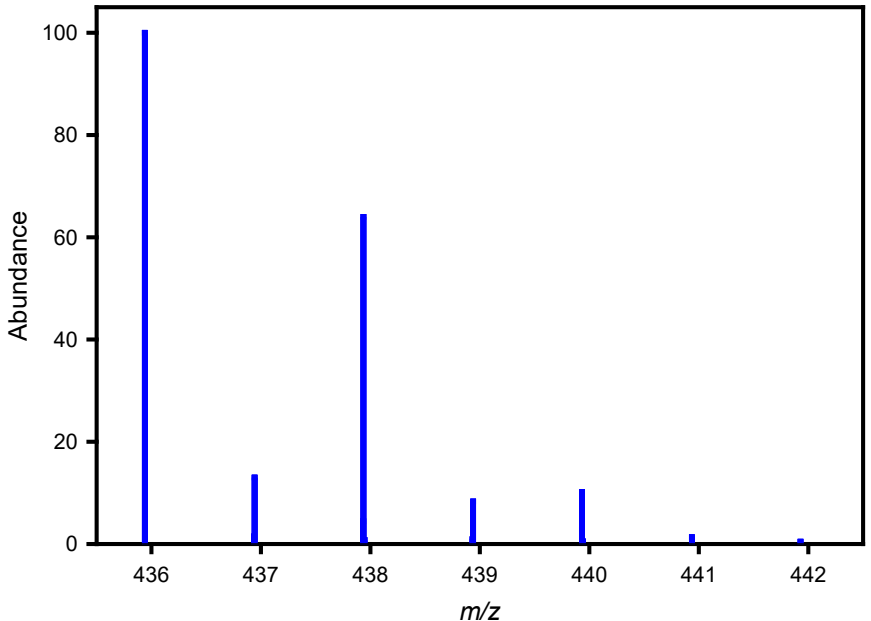
Example: Fipronil

Molecular Formula: $C_{12}H_4Cl_2F_6N_4OS$



Monoisotopic Mass: 435.938706

= (12.0000*12 Carbon) + (1.007825*4 Hydrogen) + (34.968853*2 Chlorine) + (18.998403*6 Fluorine) + (14.003074*4 Nitrogen) + (15.994915*1 Oxygen) + (31.972072*1 Sulfur)

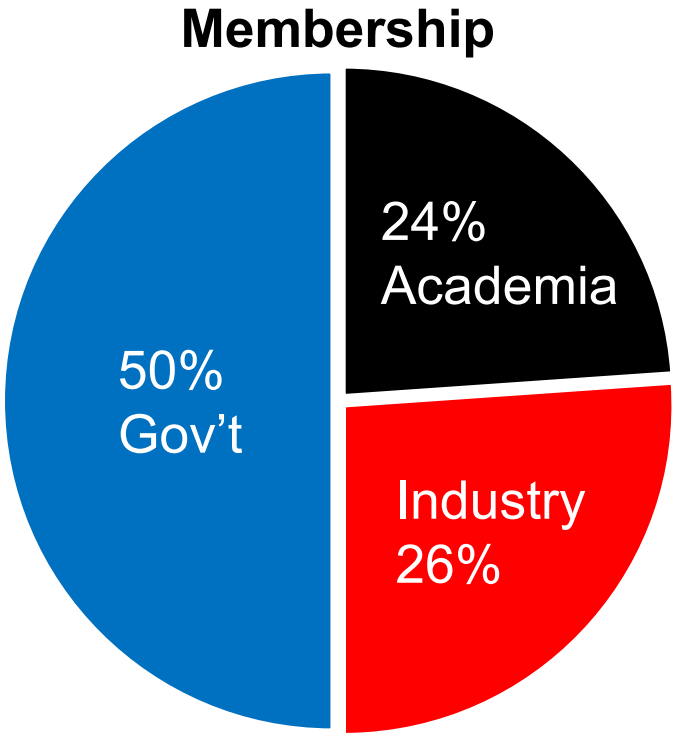


Benefits of Using Non-Targeted Analysis

- Ability to detect many more compounds
  - Includes unknowns, things not in databases (like metabolites)
  - Broad range of chemical space covered (Define!)
- Rapidly screen for knowns
  - Virtually unlimited in number
- Data is collected in a way to allow retrospective analysis
  - When did this compound start showing up?

Benchmarking and Publications for Non-targeted Analysis (BP4NTA)

- ~90 international members
- Membership based on interest in NTA
  - Experience with NTA varies from beginners to experts
  - Wide range of applications: metabolomics, exposure, food, biological, medical devices, environmental
- Leads Ben Place (NIST) and Elin Ulrich (EPA)



Interested? Contact us!

benjamin.place@nist.gov

ulrich.elin@epa.gov

Working Group Objectives

- Create a list of commonly-used NTA terms, concepts, and performance calculation and provide definitions/equations
  - Publish guidance document with terms, use community consensus and feedback
  - Audience includes new researchers, journal reviewers/editors, and experts
- Reporting recommendations and scientific best practices to promote transparency and reproducibility
- Build and maintain coalitions/communications with other groups that have similar interests including metabolomics, NORMAN, mQACC
- Move toward proficiency testing levels for SSA and NTA (ASTM/ISO)
  - Define proficiency expert, competent, etc. (10 years out)

Examples from Working Group

Suspect screening is a methodology that aims to identify chemicals detected via mass spectrometry by comparing to a predefined user list or library containing known chemicals of interest using information such as accurate mass and isotope ratios.

Non-Targeted Analysis (NTA) aims to identify chemicals of interest detected by mass spectrometry without a predefined list of chemicals. Note NTA may include suspect screening, for example, by using a defined list and looking for those features in the data then attempting the identification of any remaining features without a defined list. Due to this complimentary usage of NTA and suspect screening, NTA often encompasses both suspect screening and true NTA.

Results Reporting: Data Processing & Analysis Performance		
Recommended Data to Assess and Report		Proposed Questions for Evaluation
Perf. Metric	Performance Data	
Quality	QC checks along data analysis workflow (e.g., check for QC compounds, check for outlier samples)	Were QC benchmarks established? Did results meet those benchmarks?
Boundary	Quantification or semi-quantification of identified compounds and associated limits of detection/identification Chemical space (e.g., K_{ow}, ionizability, etc.) of the data processing/analysis method (e.g., of the library/database used; constraints introduced by approaches such as mass defect analysis or molecular networking, etc.)	Did sensitivity impact compound detection? Was the chemical space covered by the data analysis method assessed? Did the data processing/analysis methods impact the chemical space of detected/identified compounds?
Accuracy	Performance calculations, such as the True Positive Rate (TPR), False Positive Rate (FPR), True Negative Rate (TNR), and False Negative Rate (FNR) either 1. at the annotation (compound) level, for known compounds in QC spikes/controls, or 2. at the sample level, for samples with known classification/grouping	Annotation level: Do these metrics provide insight about the accuracy/selectivity of the detection and identification workflow with respect to certain compound classes or chemical characteristics? Sample level: Do these metrics provide insight about the accuracy/selectivity of the classification method with respect to certain sample types?
Precision	Reproducibility of detection and identification for QC spikes across sample types or in QC controls (e.g., re-injected pooled samples). Performance calculations, such as the False Discovery Rate (FDR) or Precision.	Was a threshold for repeatable detection and identification listed? Was detection and identification sufficiently reproducible? What factors impacted reproducibility?

Section	Category	Sub-Category	Example Information to Report
Methods	**Experimental Design**	Objectives & Scope	•Study goals, hypotheses, and use/definitions of NTA/SS in the study •Expected chemical coverage of approach (e.g., volatile, nonpolar compounds by GC-EI-MS)
		Sample Information & Preparation	•Sample collection and storage, to include QA •Sample preparation, extraction, and clean-up methods, to include QA •Development and use of blanks
		QC Spikes & QC Controls	•Use of isotopically labeled standards and/or RT reference material •Use of Development and use of positive (spiked) controls, pooled matrix controls, blanks for QC
	Data Acquisition	Run Order Preparation	•Sample replication and randomization •Inclusion of blanks and QC samples in the acquisition sequence
		Chromatography	•Instrument specifications •Method settings (e.g., column, mobiles phase, gradient, injection techniques)
		Mass Spectrometry	•Instrument specifications and Method settings (e.g., resolution, acquisition parameters, DDA vs. DIA vs. AIF) •Instrument calibration and/or tuning procedures
	Data Processing & Analysis	Data Processing	•Software program(s) used, including file conversion •Workflow steps (e.g., centroiding, peak picking, alignment, gap filling), methods, and settings •Peak detection thresholds (e.g., replicate detection criteria; minimum height, area, or S/N levels; comparison to blanks) •Data correction or normalization methods (e.g., RT calibration/indices, peak area/height normalization with IS, blank subtraction)
		Statistical Analyses	•Software program(s) used •Method goals (e.g., feature prioritization, compound class identification, sample classification) •Method type (e.g., clustering, classification, hypothesis testing) and settings
		Annotation & Identification	•Software program(s), libraries, and databases used (including information about in-house databases) •Workflow steps (e.g., formula assignment, suspect screening, MS/MS spectra interpretation, library MS/MS matching) •Workflow methods (e.g., formula prediction method, scoring algorithms) and settings •Thresholds for annotation/identification (e.g., mass error and RT; scores for formula assignment, MS/MS spectral matching)
Results	**Data Outputs**	Identification & Confidence Levels	•Reported identifications and associated confidence levels (e.g., Schymanski et al. levels) •Supporting annotated data (e.g., formula, RT, MS/MS match scores, fine isotope pattern, source of MS/MS spectra) •For unidentified features (i.e., not standard-confirmed), proposed tentative structures and other annotated data •Exported MS/MS spectra (e.g., as a library, database, or deposition into online repository)
		Statistical Outputs	•Visuals/plots (e.g., heatmaps, PCA and loading plots) and statistical output (e.g., adj. p-values) •Reported classifications or groupings of features, identifications, or samples •New statistical packages or code
	QA/QC and Other Performance Metrics	Data Acquisition	•Quality: Deviations from QA practices and results from QC checks for sample preparation and data acquisition •Boundary: Description of the capabilities/chemical space of sample prep, chromatographic, and MS methods •Accuracy: Reported chromatographic and mass accuracy •Precision: Reported variability of retention time, precursor mass error, and abundance
		Data Processing & Analysis	•Quality: Outcomes of QC checks along data analysis workflow •Boundary: Quantification or semi-quantification of identified compounds, limits of detection/identification •Accuracy: Calculations such as TPR, FPR, etc. at annotation level for QC spikes/control samples or at sample level •Precision: Reproducibility of identification for QC spikes across sample types

Kathy Peter (NIST) Lead

Example Uses and Requirements

Decision Context			Example Uses of NTA Data	Example Stakeholders
Sample Classification	*Chemical Annotation*	*Semi-Quantitation*		
Req	Opt	Opt	- Classify locations impacted by point-source emitters - Classify locations impacted by inadvertent environmental releases - Classify exposure status for active or former military personnel - Classify food items not meeting criteria for product certification	- EPA, USGS - FEMA, EPA - DoD, VA - FDA, NIST
Req	Req	Opt	- Identify natural or synthetic chemical nerve agents - Identify chemicals associated with product-related illness - Identify chemicals released in emergency response scenarios - Identify designer drugs used for athletic performance enhancement	- DHS, CDC - CPSC, FDA - FEMA, EPA - DEA, FDA
Req	Req	Req	- Assess occupational health risks from exposure to fire-fighting foams - Assess consumer health risks from exposure to household products - Assess ecological health risks from exposure to urban wastewater - Assess maternal and infant health risk from exposure during pregnancy	- NIOSH, DoD - CPSC, EPA - USGS, EPA - NIEHS, EPA

Content from Jon Sobus, EPA

Uses of NTA Data

- **Exposure surveillance**
  - What chemicals are in food, water, products, dust, blood, etc.?
  - Starting point for generation of targeted methods to allow addition to lists like Unregulated Contaminant Monitoring Rule (UCMR)
- **Chemical prioritization**
  - What are relevant chemicals & mixtures?
  - Lautenberg Act- Risk-based process to determine which chemicals to prioritize for assessment, identifying them as high/low priority substances
- **Exposure forensics**
  - What are chemical signatures of exposure sources?
  - Can assist with enforcement/cleanup efforts
- **Biomarker discovery**
  - What chemicals are associated with health impairment?
  - Provides data for AOPs, new tests for HT toxicity screening efforts

The Future of NTA

- Standardized QA/QC, terminology, review, reporting
  - As possible, standardize methods
- Benchmarking, performance metrics
  - True/False Positives/Negative, chemical space coverage
- Learning from related fields (e.g., metabolomics)
- Semi-quantitative analysis
- Regulatory uses
- “Make non-targeted the new targeted” –Thomas Burke

Acknowledgements and References

Ben Place (NIST)
Jon Sobus (US EPA)
Kathy Peter (NIST)
Sara Nason (CT Agr Exp Stn)
Seth Newton (US EPA)
BP4NTA working group members
EPA NTA researchers

BP4NTA website (Sara and Seth)- https://nontargetedanalysis.org/
CompTox Chemicals Dashboard- https://comptox.epa.gov/dashboard/
SETAC Focused Topic Meeting Nontarget Analysis for Environmental Risk Assessment- https://nta.setac.org/

Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA
https://www.nature.com/articles/s41370-017-0012-y

EPA's non-targeted analysis collaborative trial (ENTACT): Genesis, design, and initial findings
https://link.springer.com/article/10.1007/s00216-018-1435-6

Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance
https://link.springer.com/article/10.1007%2Fs00216-018-1526-4

Examining NTA performance and potential using fortified and reference house dust as part of ENTACT
https://link.springer.com/article/10.1007%2Fs00216-020-02658-w