Non-Targeted and Suspect Identification of Per- and Polyfluoroalkyl Substances through Databases and Other Tools

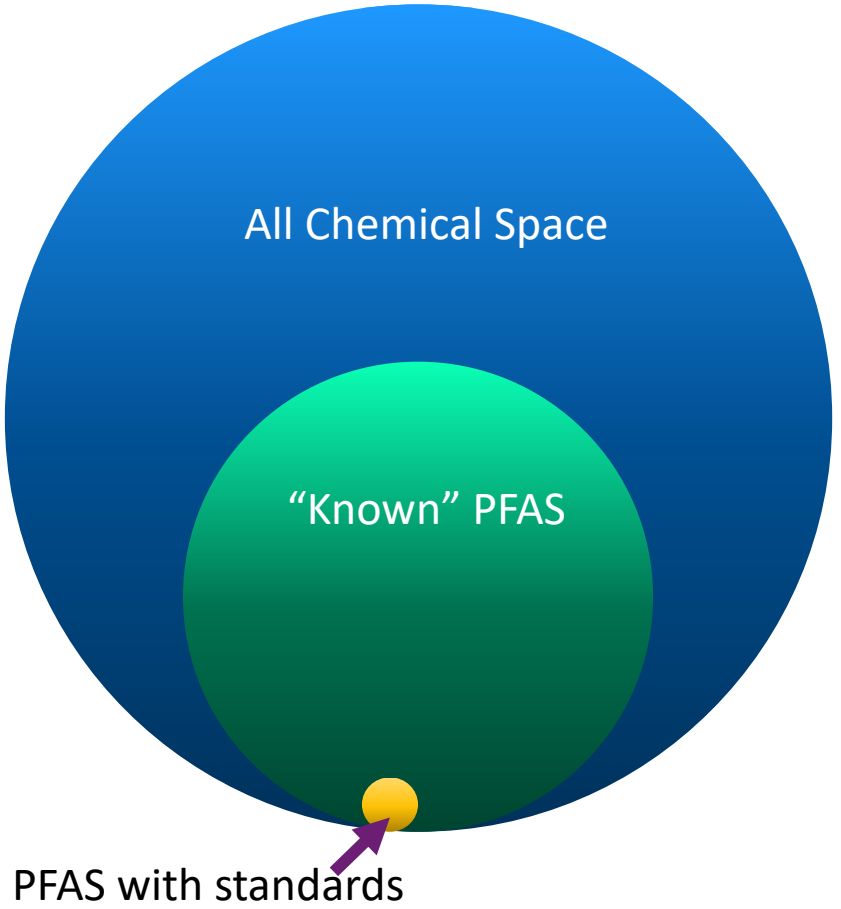
Benjamin Place, Research Chemist

National Institute of Standards & Technology

SERDP • ESTCP
SYMPOSIUM
2021 • Enhancing DoD's Mission Effectiveness

Per- and Polyfluoroalkyl Substances

- Estimates for the total number of different PFAS in the environment and/or used in varying industries can be over 4,000.
- There are a limited number of analytical standards available for these compounds.
  - Estimated ~100 standards commercially available.
  - This makes traditional targeted approaches (LC-MS/MS) more difficult
- **Non-Targeted Analysis** and **Suspect Screening Analysis** are two approaches towards the detection of known and unknown PFAS without the availability of authentic analytical standards.

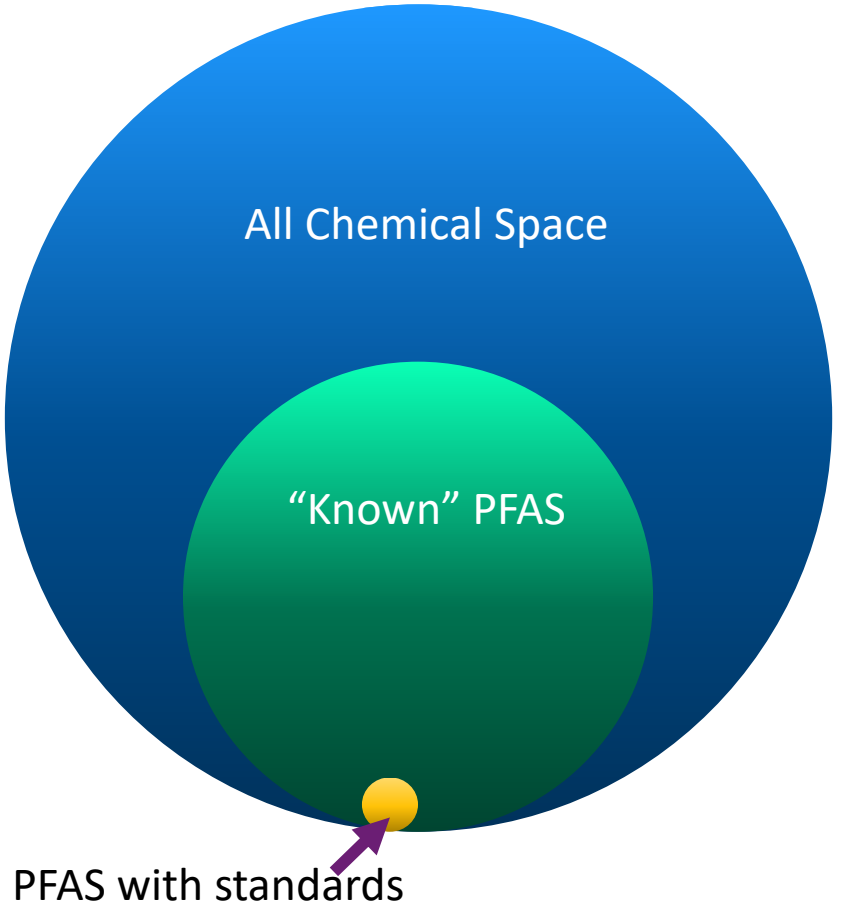


NIST

Non-Targeted Analysis – An Overview

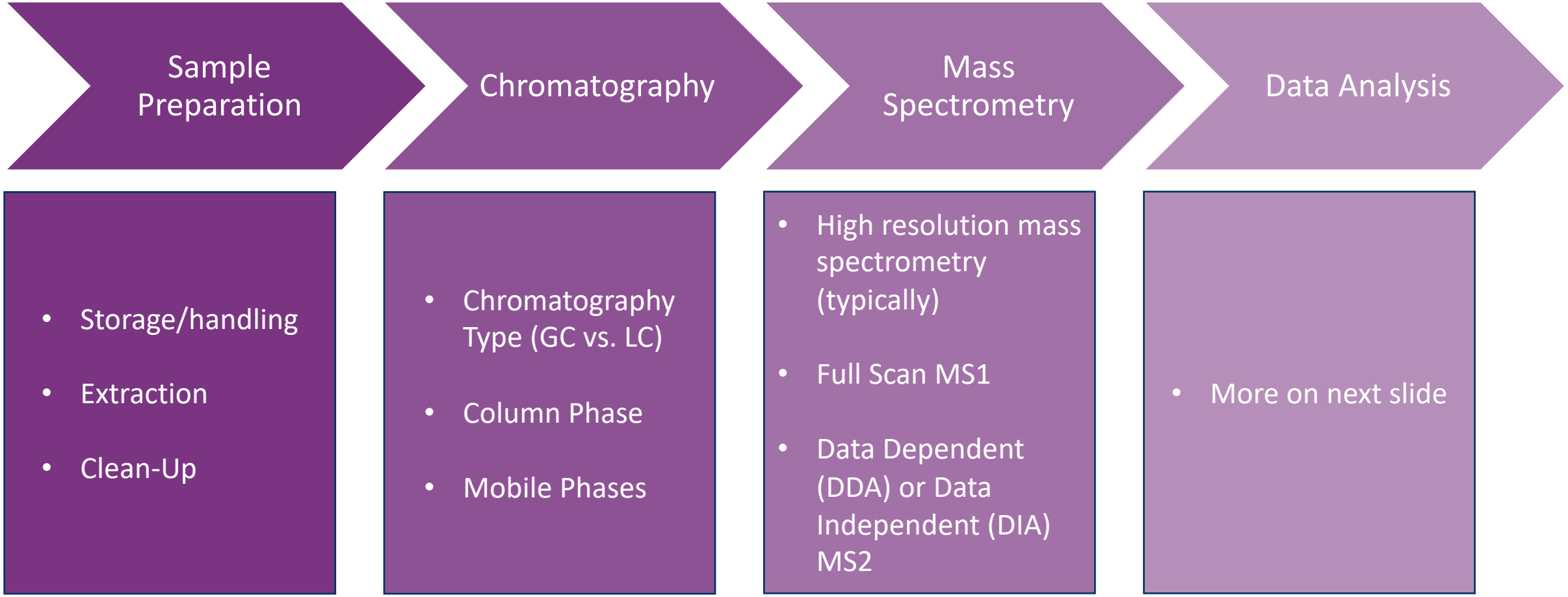
Non-targeted analysis (NTA) is a theoretical concept that can be broadly defined as the characterization of the chemical composition of any given sample without the use of *a priori* knowledge regarding the sample's chemical content.

Suspect screening analysis (SSA) is the identification of chemicals and/or chemical classes detected by an instrument, typically a mass spectrometer, by comparison to a predefined user list or library containing known chemicals of interest.



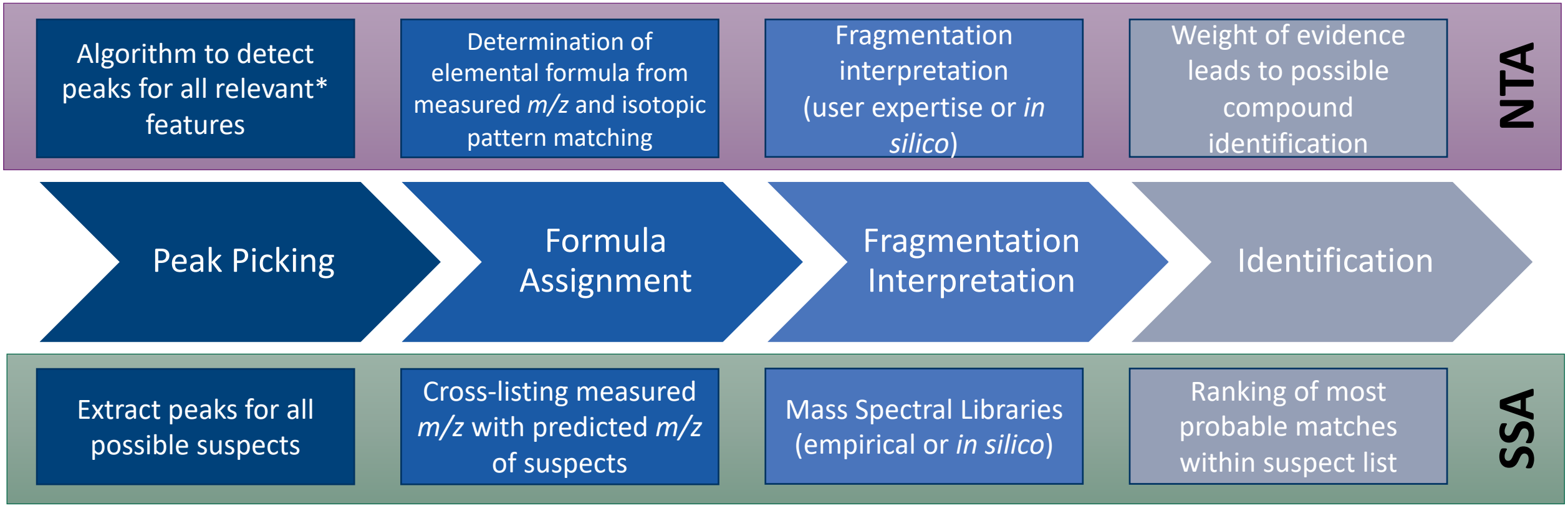
Benchmarking and Publications for Non-Targeted Analysis Working Group. https://nontargetedanalysis.org/

Non-Targeted Analysis – Analytical Workflow



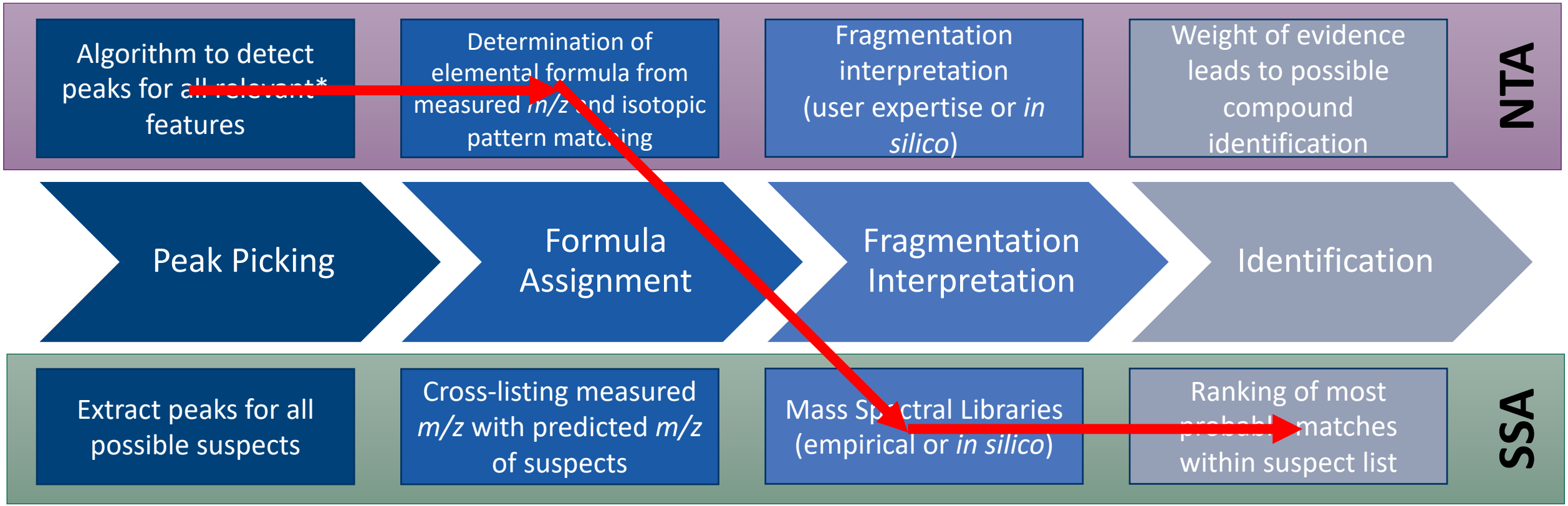
Non-Targeted Analysis – Data Workflow

Non-Targeted Analysis – Data Workflow



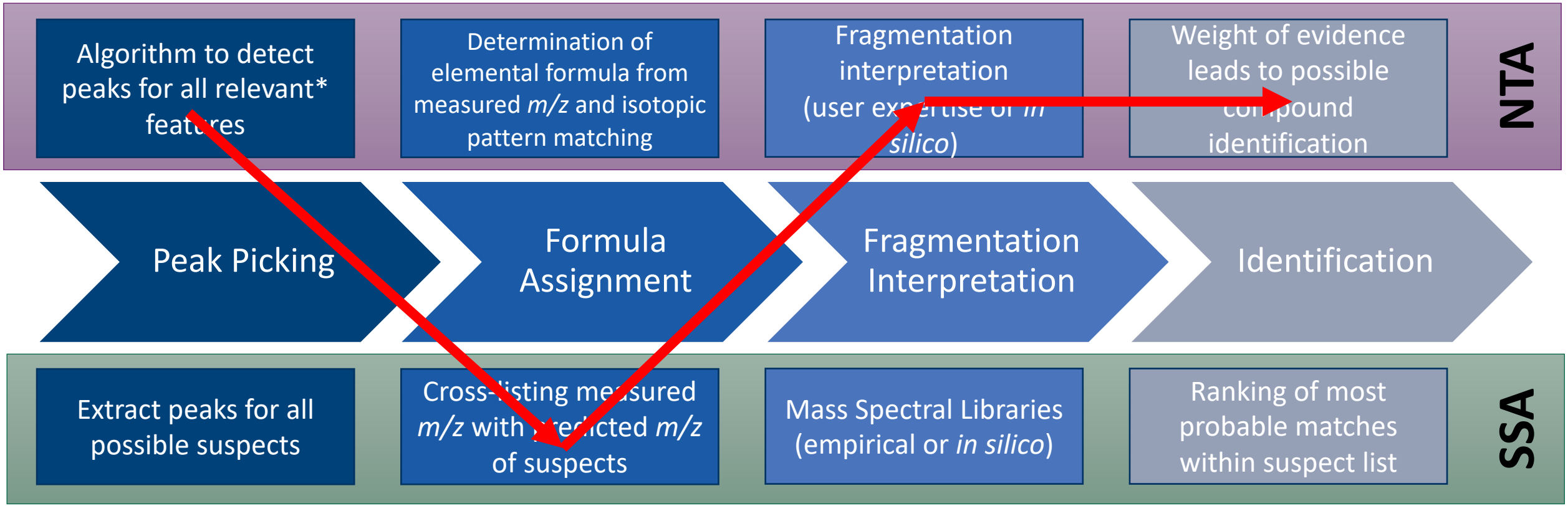
* Using predetermined parameters to define a peak as **relevant**

Non-Targeted Analysis – Data Workflow



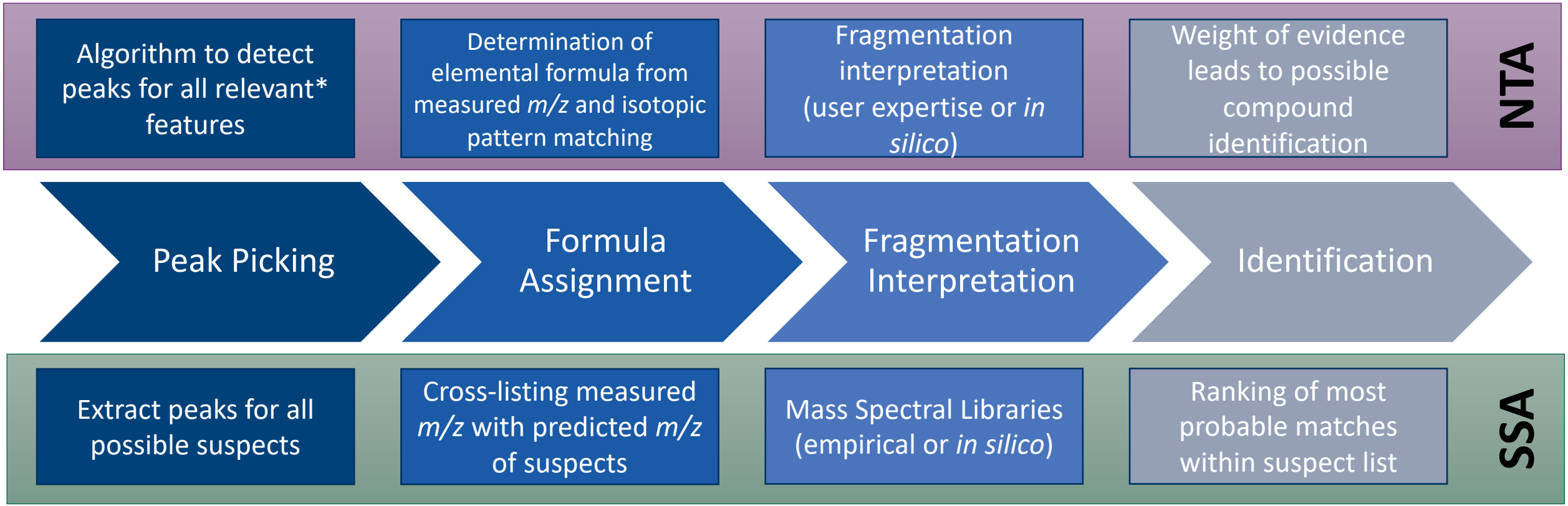
* Using predetermined parameters to define a peak as **relevant**

Non-Targeted Analysis – Data Workflow



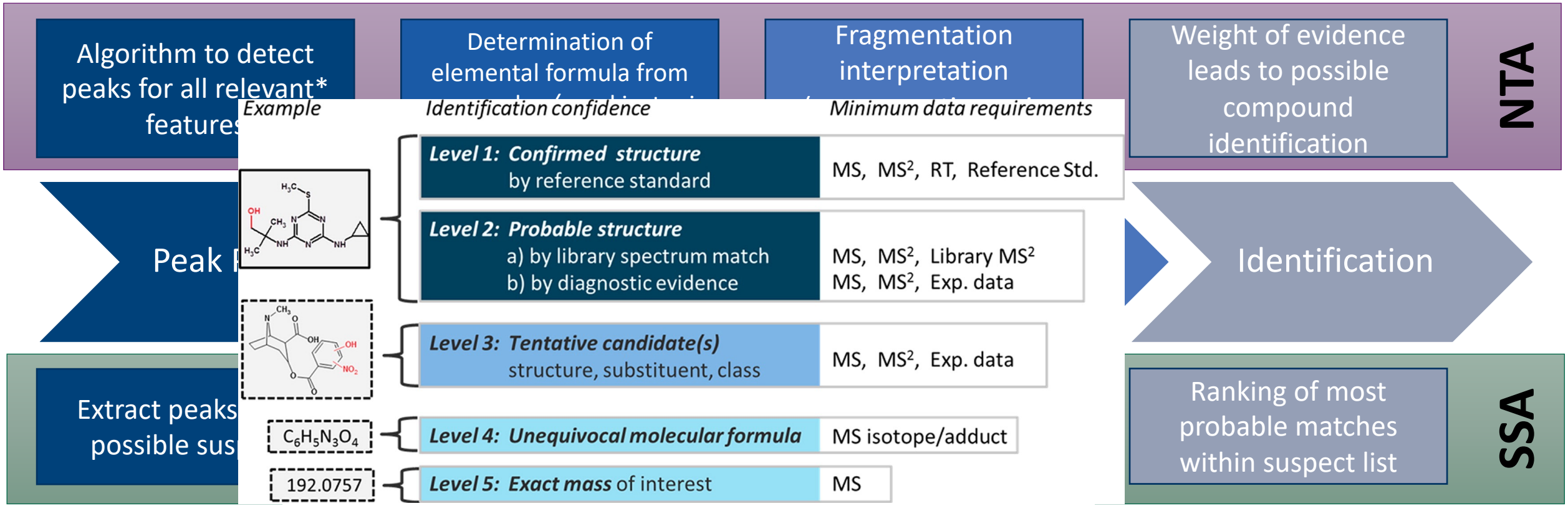
* Using predetermined parameters to define a peak as **relevant**

Non-Targeted Analysis – Data Workflow



* Using predetermined parameters to define a peak as **relevant**

Non-Targeted Analysis – Data Workflow



Schymanski, et al. **2014** DOI: 10.1021/es5002105

* Using predetermined parameters to define a peak as **relevant**

Non-Targeted Analysis – PFAS

There are specific properties of PFAS that can be used to selectively screen the data.

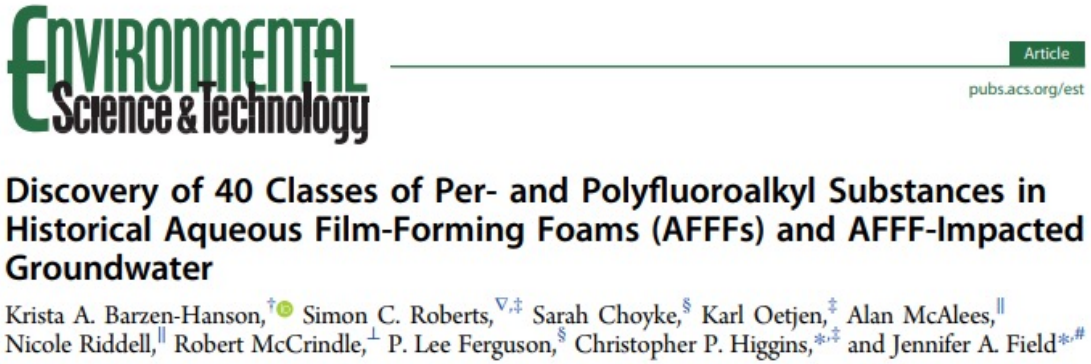
Non-Targeted Analysis – PFAS

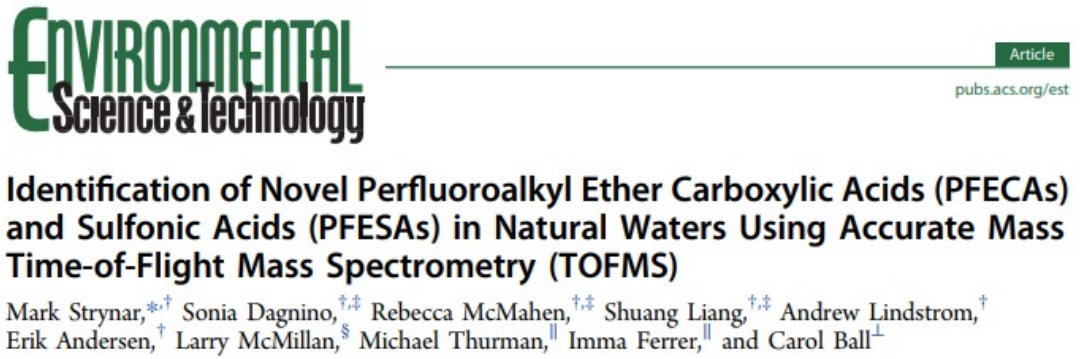
There are specific properties of PFAS that can be used to selectively screen the data.

- Low to negative mass defect

^{1}H = 1.0078 amu
^{19}F = 18.9984 amu

$C_8H_{15}O_2$ = 143.1072 amu
(mass defect = 0.1072 amu)
$C_8F_{15}O_2$ = 412.9658 amu
(mass defect = -0.0342 amu)

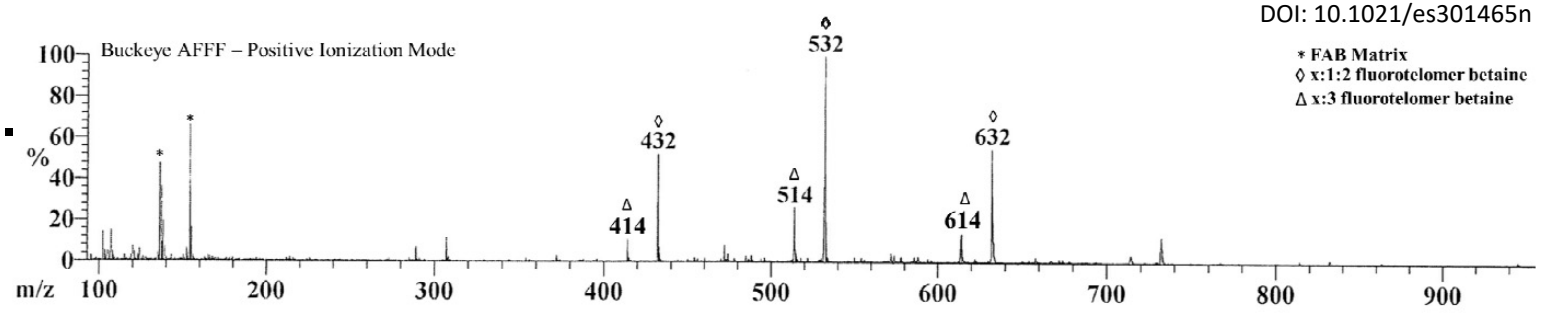




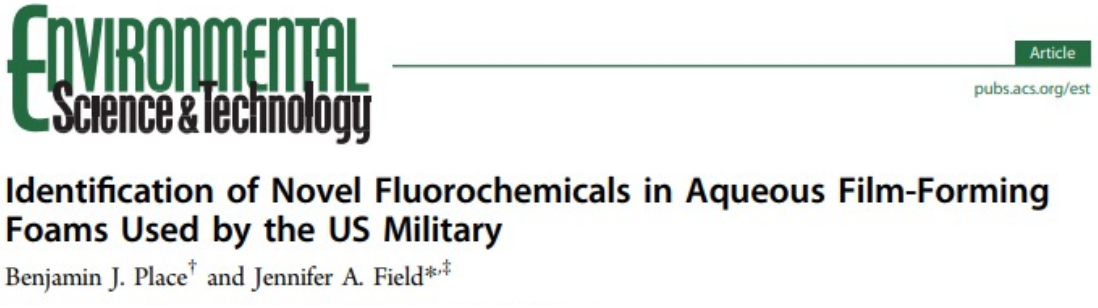
Non-Targeted Analysis – PFAS

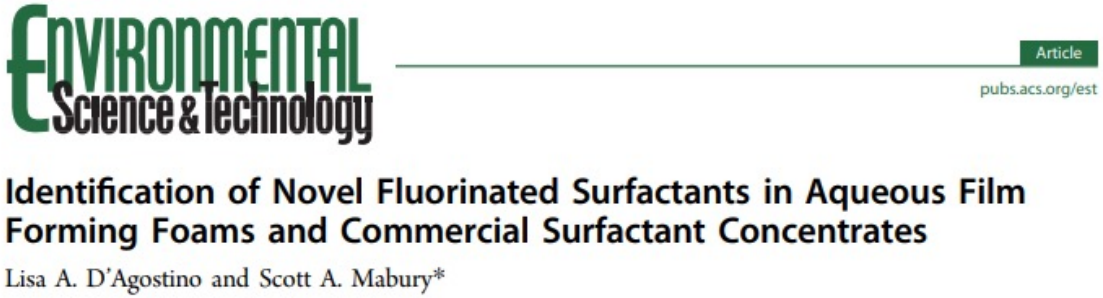
There are specific properties of PFAS that can be used to selectively screen the data.

- Low to negative mass defect
- Co-occurrence of homologs



Place and Field (2012)
DOI: 10.1021/es301465n

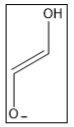




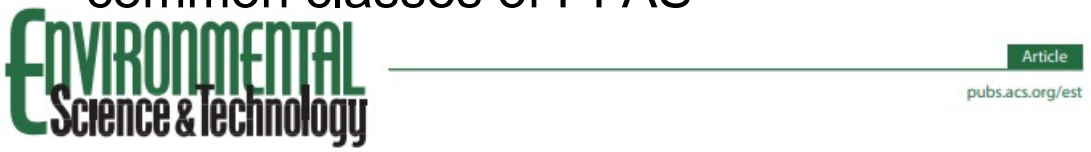
Non-Targeted Analysis – PFAS

There are specific properties of PFAS that can be used to selectively screen the data.

- Low to negative mass defect
- Co-occurrence of homologs
- Specific fragmentation patterns related to common classes of PFAS

Formula	Exact Mass	Structure
$C_2H_3O_2^-$	59.0139	
CF_3^-	68.9958	
NO_2S^-	77.9655	
O_3S^-	79.9574	

Barzen-Hanson, et al. (2017)
DOI: 10.1021/acs.est.6b05843



Article

pubs.acs.org/est

Discovery of 40 Classes of Per- and Polyfluoroalkyl Substances in Historical Aqueous Film-Forming Foams (AFFFs) and AFFF-Impacted Groundwater

Krista A. Barzen-Hanson,[†] Simon C. Roberts,[▽,‡] Sarah Choyke,[§] Karl Oetjen,[‡] Alan McAlees,[∥] Nicole Riddell,[∥] Robert McCrindle,[⊥] P. Lee Ferguson,[§] Christopher P. Higgins,[*,‡] and Jennifer A. Field[*,#]

Article

Cite This: *Environ. Sci. Technol.* 2019, 53, 4717−4727

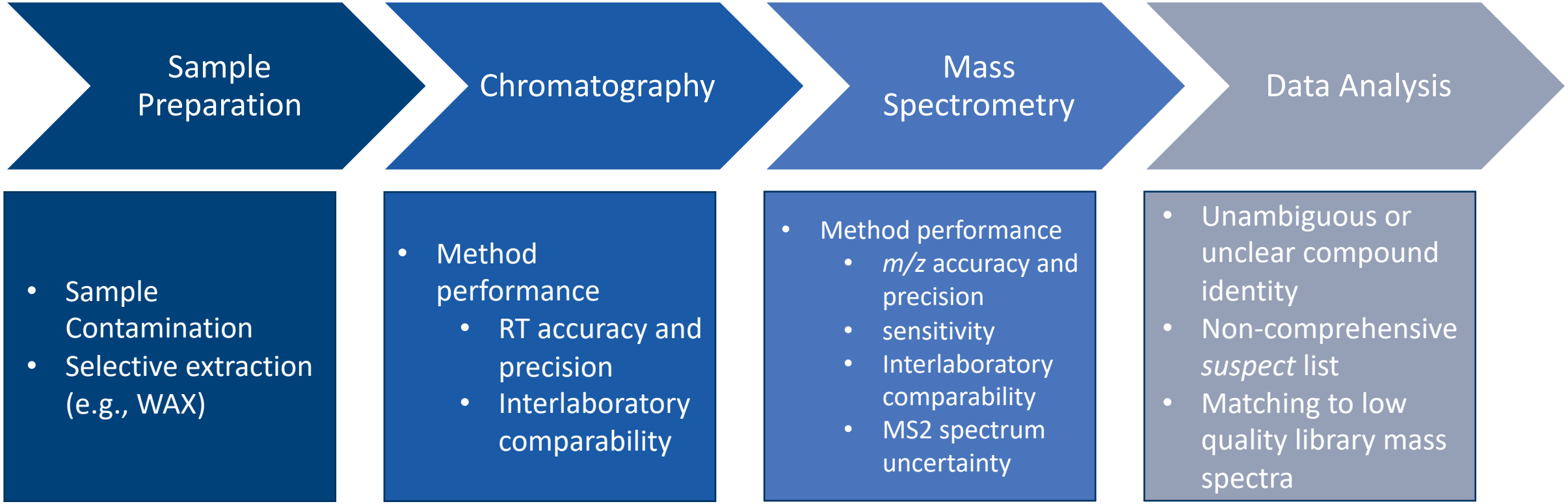
pubs.acs.org/est

Identification of Per- and Polyfluoroalkyl Substances in the Cape Fear River by High Resolution Mass Spectrometry and Nontargeted Screening

James McCord[†] and Mark Strynar[*,‡]

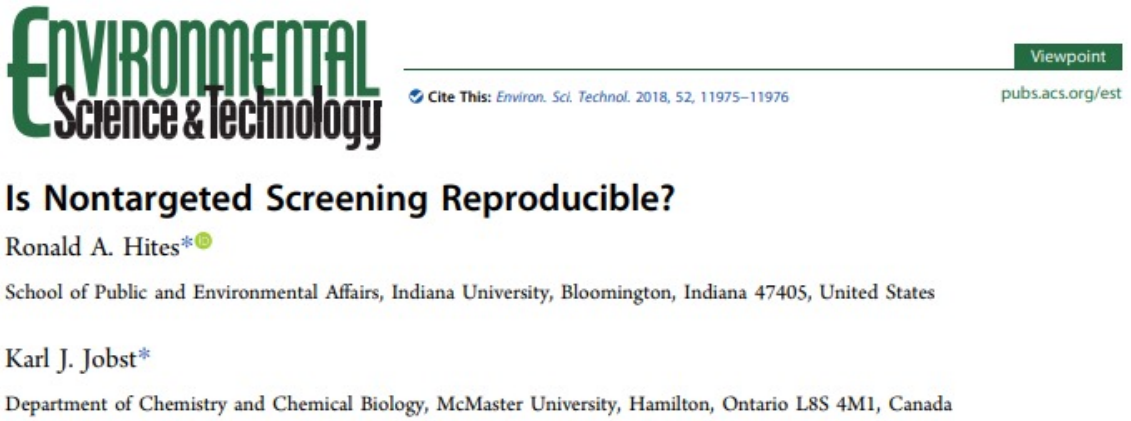
SERDP • ESTCP SYMPOSIUM
#SerdpEstcp2021

Sources of Uncertainty or Bias in the NTA Workflow



NTA Study Design and Reporting

- Within the design of the study itself, there are uncertainties/biases that must be considered.
  - Does your experimental design provide enough evidence to identify a novel PFAS?
  - How has the sample collection been designed to ensure no false positives?
  - What is the performance of the method?
  - How can your results be compared to the results of other NTA studies?
  - What materials, data, and/or data tools will you use to validate your method?



Benchmarking and Performance for Non-Targeted Analysis (BP4NTA)

- In 2018, NTA researchers formed the BP4NTA Working Group, based on the common interests and values:
  - Shared terms and definitions can lead to better comparison of methods and results.
  - We can determine benchmark performance once methods and results reporting are more harmonized.
  - Once benchmarks are set, we can then tackle concepts like “proficiency testing” for NTA measurements.

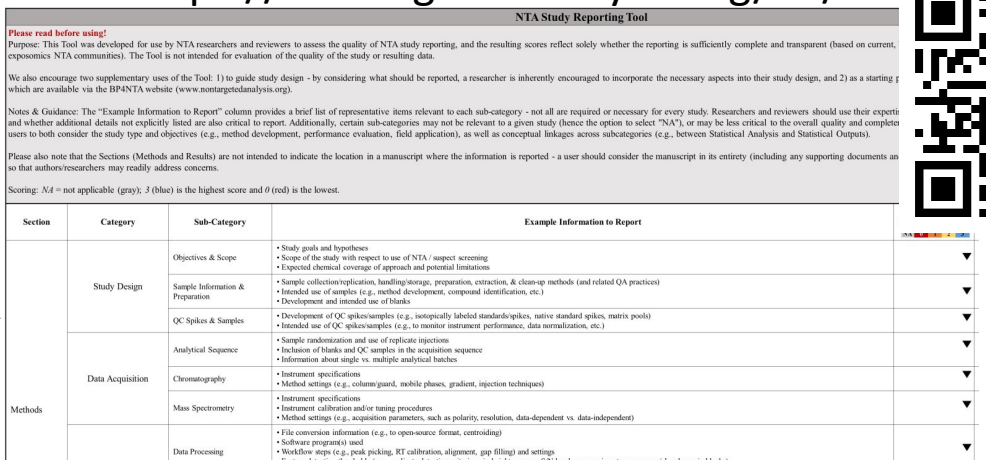
BP4NTA Introduction Manuscript: Place, et al. (2021) *Anal. Chem. In Revision.*

https://nontargetedanalysis.org

Method Reporting Tools

- BP4NTA Members developed the NTA Study Reporting Tool (SRT)
  - Designed to enable study designers, manuscript reviewers, and journal editors to better understand what type of information should be considered and provided in an NTA manuscript.
    Peter, et al., *Anal. Chem.* 2021. DOI: 10.1021/acs.analchem.1c02621

https://nontargetedanalysis.org/srt/



NTA Study Reporting Tool

Please read before using!

Purpose: This Tool was developed for use by NTA researchers and reviewers to assess the quality of NTA study reporting, and the resulting scores reflect solely whether the reporting is sufficiently complete and transparent (based on current, exposomics NTA communities). The Tool is not intended for evaluation of the quality of the study or resulting data.

We also encourage two supplementary uses of the Tool: 1) to guide study design - by considering what should be reported, a researcher is inherently encouraged to incorporate the necessary aspects into their study design, and 2) as a starting p which are available via the BP4NTA website (www.nontargetedanalysis.org).

Notes & Guidance: The "Example Information to Report" column provides a brief list of representative items relevant to each sub-category - not all are required or necessary for every study. Researchers and reviewers should use their experti and whether additional details not explicitly listed are also critical to report. Additionally, certain sub-categories may not be relevant to a given study (hence the option to select "NA"), or may be less critical to the overall quality and completen users to both consider the study type and objectives (e.g., method development, performance evaluation, field application), as well as conceptual linkages across subcategories (e.g., between Statistical Analysis and Statistical Outputs).

Please also note that the Sections (Methods and Results) are not intended to indicate the location in a manuscript where the information is reported - a user should consider the manuscript in its entirety (including any supporting documents an so that authors/researchers may readily address concerns.

Scoring: *NA* = not applicable (gray); *3* (blue) is the highest score and *0* (red) is the lowest.

Section	Category	Sub-Category	Example Information to Report
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
		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
		QC Spikes & Samples	• Development of QC spikes/samples (e.g., isotopically labeled standards/spikes, native standard spikes, matrix pools) • Intended use of QC spikes/samples (e.g., to monitor instrument performance, data normalization, etc.)
	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
		Chromatography	• Instrument specifications • Method settings (e.g., column/guard, mobile phases, gradient, injection techniques)
		Mass Spectrometry	• Instrument specifications • Instrument calibration and/or tuning procedures • Method settings (e.g., acquisition parameters, such as polarity, resolution, data-dependent vs. data-independent)
		Data Processing	• File conversion information (e.g., to open-source format, centroiding) • Software program(s) used • Workflow steps (e.g., peak picking, RT calibration, alignment, gap filling) and settings

- To enable systematic collection of NTA method metadata for the NIST database, a macro-enable Microsoft Excel workbook has been developed.
  - Vocabulary is controlled
  - Exports data to open JSON format

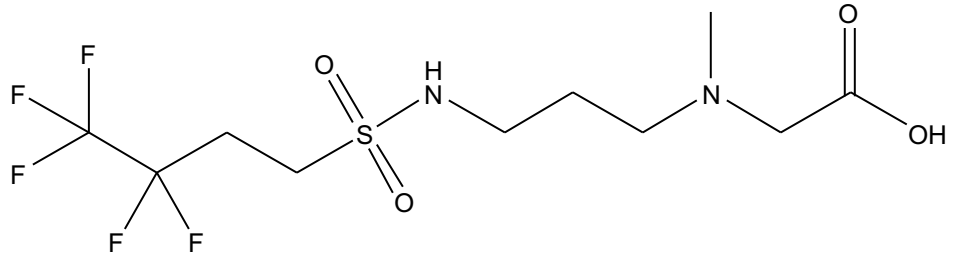
Mass Spectrometry	
User Input	
Description	Value
Mass Spectrometer Vendor	ThermoFisher Scientific
Mass Spectrometer Model	Q-Exactive
Ionization Mode	electrospray ionization
Polarity	negative
Ionization Voltage/Current	2500
Ionization Voltage/Current Units	V
First Mass Analyzer	quadrupole
Second Mass Analyzer	orbitrap
Fragmentation Type	HCD
Fragmentation Energy	30
Fragmentation Energy Type	fixed
Fragmentation Energy Units	normalized
MS2 Experiment	DDA
Isolation Width/Window (Da)	0.7
Instrument Mass Accuracy (ppm)	5
MS1 Resolution	70000
MS2 Resolution	17500


```
},
"massspectrometry": {
    "msvendor": "ThermoFisher Scientific",
    "msmodel": "Q-Exactive",
    "ionization": "electrospray ionization",
    "polarity": "negative",
    "voltage": "2500",
    "vunits": "V",
    "massanalyzer1": "quadrupole",
    "massanalyzer2": "orbitrap",
    "fragmode": "HCD",
    "ce_value": "30",
    "ce_desc": "fixed",
    "ce_units": "normalized",
    "ms2exp": "DDA",
    "isowidth": "0.7",
    "msaccuracy": "5",
    "ms1resolution": "70000",
    "ms2resolution": "17500",
    "source": ""
},
```

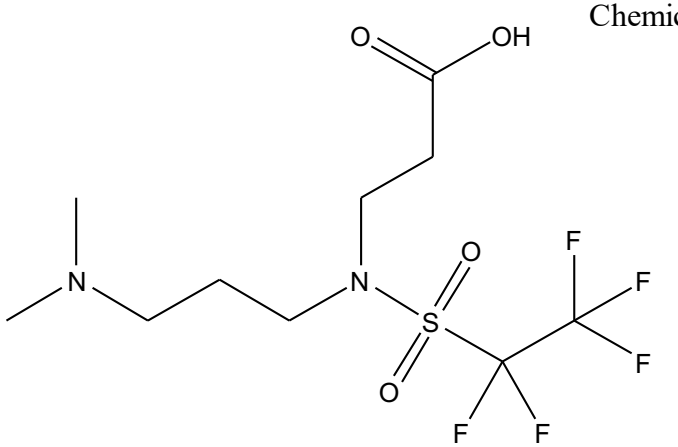

https://github.com/USNISTGOV/NISTPFAS

Reporting PFAS Identities

- Proving that an identified PFAS is **new** requires a significant effort to check prior lists (and study results) for all previously-discovered PFAS
  - Not all are easily searchable.
  - Naming conventions vary between researchers.
  - Compounds with the same molecular formula can have vastly different chemistries.



N-methyl-*N*-(3-((3,3,4,4,4-pentafluorobutyl)sulfonamido)propyl)glycine
Chemical Formula: $C_{10}H_{17}F_5N_2O_4S$
Exact Mass: 356.0829



3-((*N*-(3-(dimethylamino)propyl)-1,1,2,2,2-pentafluoroethyl)sulfonamido)propanoic acid
Chemical Formula: $C_{10}H_{17}F_5N_2O_4S$
Exact Mass: 356.0829

NIST Suspect List of Possible PFAS

- Purpose: To create a central location for PFAS identified by peer-reviewed manuscripts, other documentations, or computational tools.
  - Identities are attributed to source.
  - Structures (as InChI) are the key value (not name, acronym, elemental formula, etc.).
  - Index values are constant
  - Additional information (acronym, chemical class, etc.) available.
- Build upon efforts by Colorado School of Mines, Oregon State University, and others.
- Public contributions are allowed, does not require chemical registration or other official entry protocol.

NIST Suspect List of Possible Per- and Polyfluoroalkyl Substances

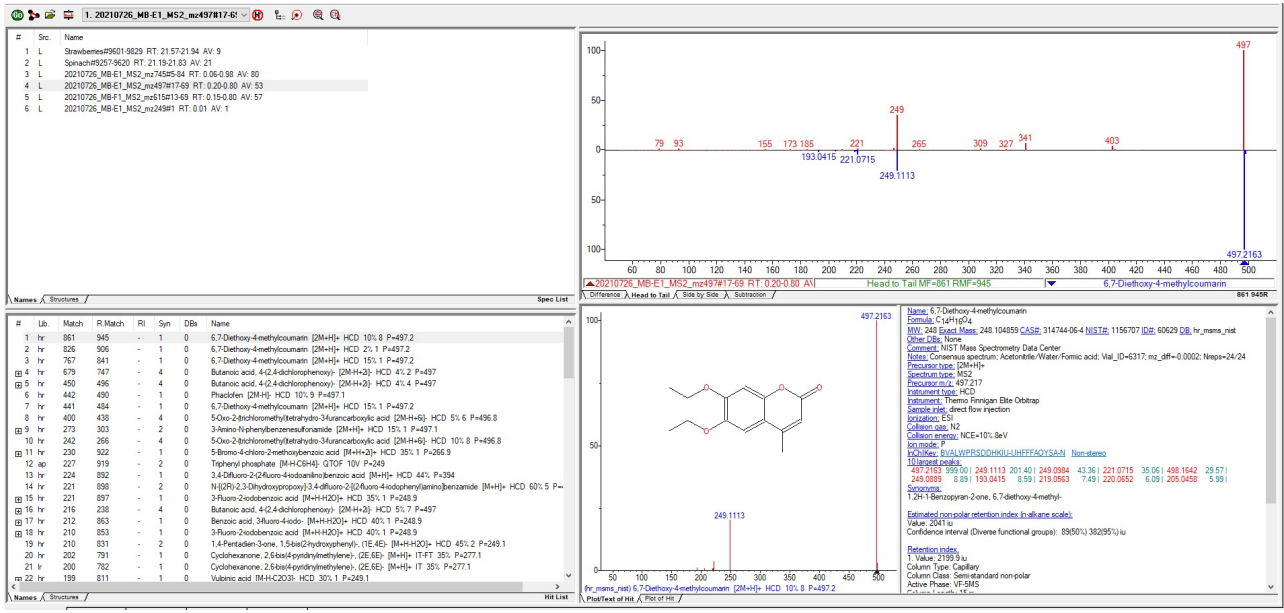
ID	CHEMICAL_NAME	INCHI	SMILES	INCHIKE'	FORMUI	FIXEDMASS
1	1H-Perfluoro-3,3-bis(trifluorom	InChI=1S/C8HF13/c1-2-3(6(13,14)15,7(16,17)18)4(9,10)5(11,12)8(19,20)21/h1H	C#CC(C(C((	UEYXJPFQ.	C8HF13	343.9870669
2	Bis(heneicosafluorodecyl)phosp	InChI=1S/C20H9F34O2P/c21-5(22,7(25,26)9(29,30)11(33,34)13(37,38)15(41,42)17	C(CP(=O)((	DEENCVDX	C20H9F34	957.9797256
3	Tris(2-(perfluorododecyl)ethyl)p	InChI=1S/C42H12F75O4P/c43-7(44,10(49,50)13(55,56)16(61,62)19(67,68)22(73,74	C(COP(=O)	JKDZBIHM	C42H12F7	2035.927562
4	2-Chloro-2-propenoic acid 3,3,4	InChI=1S/C9H6ClF9O2/c1-4(10)5(20)21-3-2-6(11,12)7(13,14)8(15,16)9(17,18)19/h	C=C(C(=O)	CVMPVWS	C9H6ClF9(	351.9912611
5	Bicyclo[2.2.1]hept-2-ene, 5,5,6-	InChI=1S/C10H6F10O/c11-6(12)4-1-2-5(3-4)7(6,13)21-10(19,20)8(14,15)9(16,17)1	C1=CC2CC	JZEKVGIW	C10H6F10	332.025897
6	3,3,4-Trifluoro-4-(heptafluoropı	InChI=1S/C12H8F10O/c13-8(14)6-4-1-2-5(3-4)7(6)9(8,15)23-12(21,22)10(16,17)11	C1=CC2CC	NGEYGRCI	C12H8F10	358.0415471
7	5-(Nonafluorobutyl)bicyclo[2.2.	InChI=1S/C11H9F9/c12-8(13,7-4-5-1-2-6(7)3-5)9(14,15)10(16,17)11(18,19)20/h1-2	C1=CC2CC	QAIOQFZL	C11H9F9	312.0560543
8	3-(Heptafluorobutyryl)camphor	InChI=1S/C14H15F7O2/c1-10(2)6-4-5-11(10,3)8(22)7(6)9(23)12(15,16)13(17,18)14	CC1(C)C2C	PEWOESYI	C14H15F7	348.0960273
9	Perfluorobutylsulfonamide	InChI=1S/C4H2F9NO2S/c5-1(6,3(9,10)11)2(7,8)4(12,13)17(14,15)16/h(H2,14,15,16	C(C(C(F)(F)	FUVKFLJW	C4H2F9NC	298.9662533
10	N-(3,4-Dichlorophenyl)-2,2,3,3,4	InChI=1S/C10H4Cl2F7NO/c11-5-2-1-4(3-6(5)12)20-7(21)8(13,14)9(15,16)10(17,18)	c1cc(c(cc1	INMDAGXI	C10H4Cl2F	356.9558166
11	N-(3,5-Dichlorophenyl)-2,2,3,3,4	InChI=1S/C10H4Cl2F7NO/c11-4-1-5(12)3-6(2-4)20-7(21)8(13,14)9(15,16)10(17,18)	c1c(cc(cc1	WGNKYUV	C10H4Cl2F	356.9558166
12	2-Propenoic acid, 2-methyl-, 2-[	InChI=1S/C28H28F17N3O8S/c1-6-48(9-10-54-19(50)46-16-8-7-14(4)17(11-16)47-2	CCN(CCOC	WUGSPVX	C28H28F1	889.1325656

https://github.com/USNISTGOV/NISTPFAS

Mass Spectral Libraries

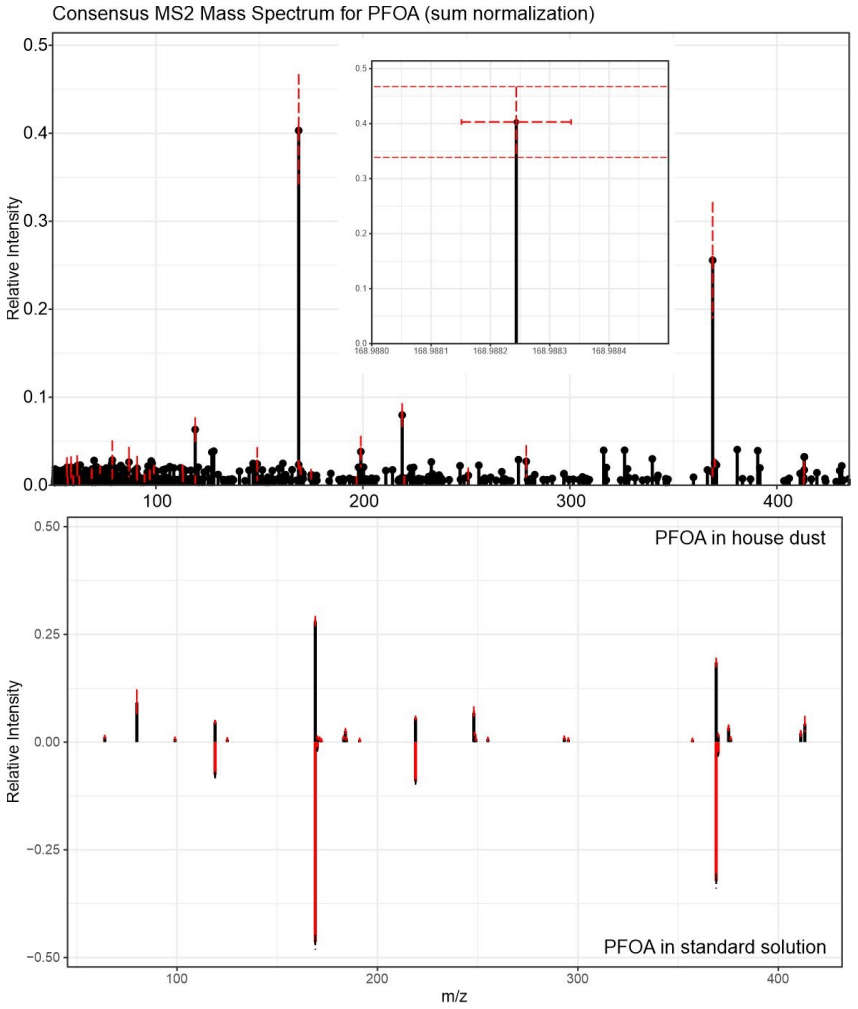
- For suspect screening analysis, libraries containing mass spectra produced from chemical standards can be extremely powerful.
- Mass spectral matches provide some of the strongest evidence for identification of compounds in lieu of analytical standards.
- Conventionally, most mass spectral libraries contain spectra for compounds that have commercially-available standards

NIST MS Search Program
https://chemdata.nist.gov



Uncertainty of Mass Spectra

- 2021 published study:
  - There is quantifiable variability about the measured *m/z* and intensity for fragmentation ions.
  - Sample matrix may affect the presence and/or variability of fragment ions for the same compound.
  - By calculating the variability of fragmentation mass spectra, the confidence of mass spectral match can be assessed.

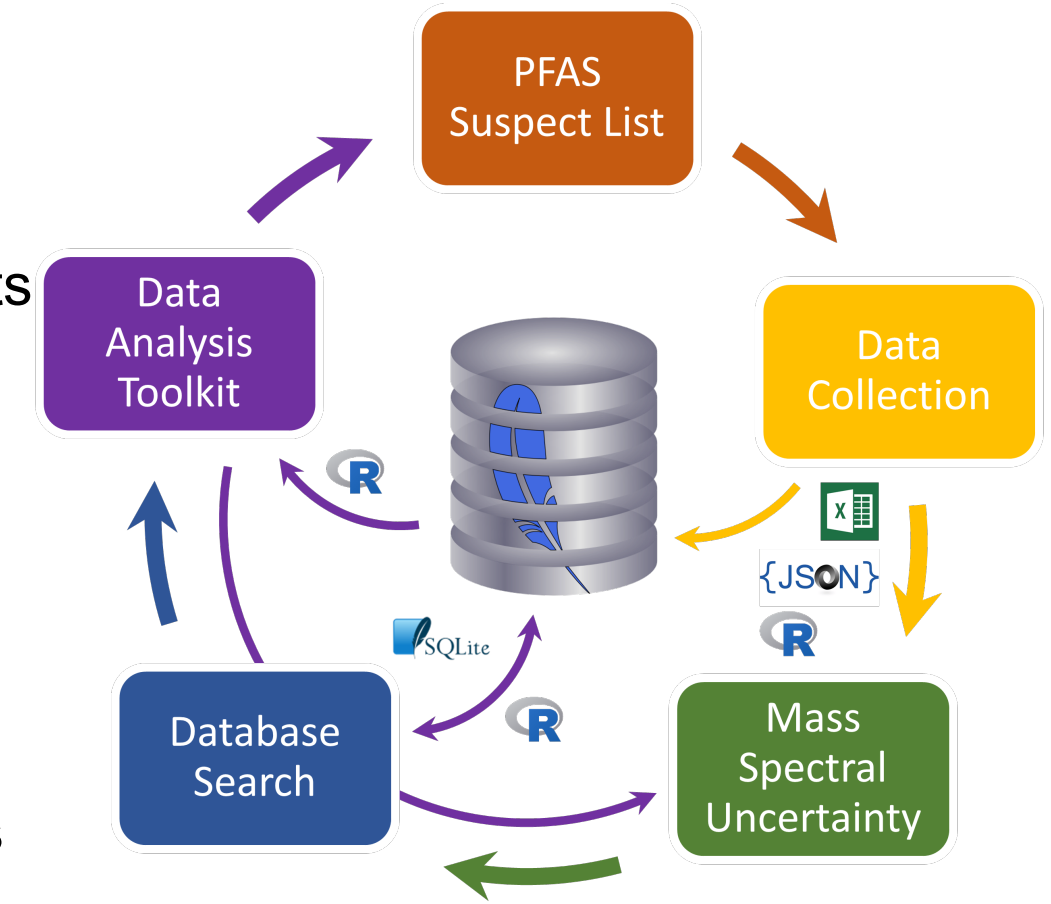


Place, BJ, *JASMS* (2021)
DOI: 10.1021/jasms.0c00423

NIST PFAS MS Database

ER20-1056 Project:

- Creation of a database infrastructure to include:
  - Structure-specific PFAS suspects
  - Reference mass spectra for identified PFAS
  - Mass spectra of PFAS in standards and matrix samples with the variability assessed
  - Open schema (SQLite) that enables researchers to build advanced tools for data analysis

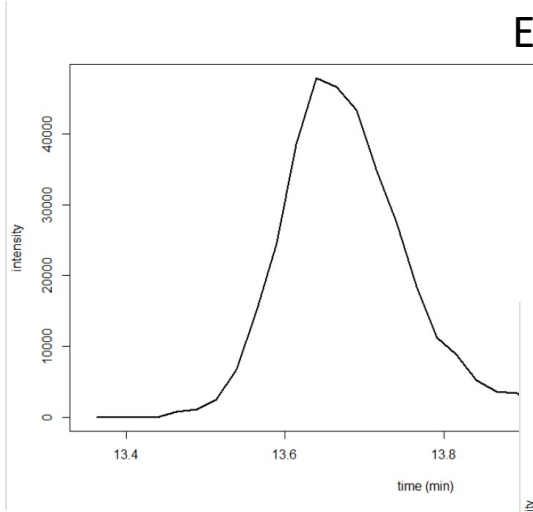


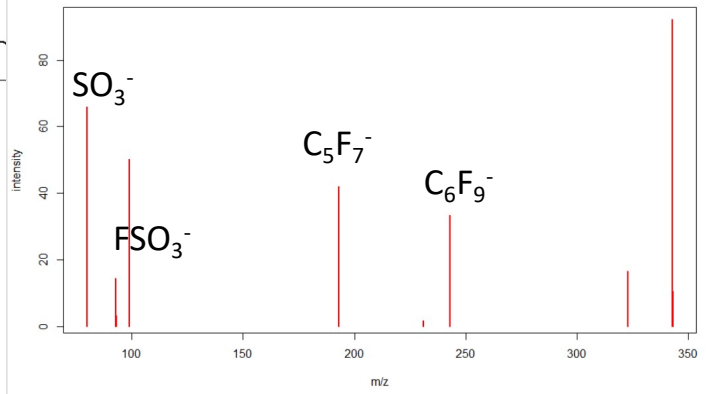
NIST PFAS MS Database

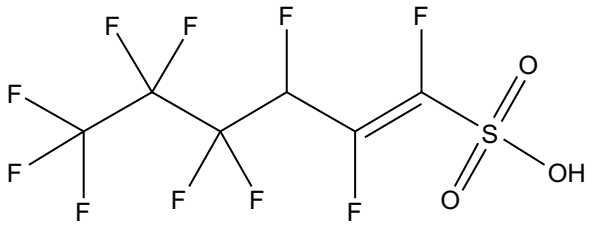
ER20-1056 Project:

- NIST PFAS Suspect List
  - 4,884 individual structures
- Current database population
  - Around 291 mass spectra from standards and commercial mixtures.
  - Representing 99 individual PFAS
  - Multiple energy levels, different instrumentation.
- Open to submission!

Example Entry Data







(*Z*)-1,2,3,4,4,5,5,6,6,6-decafluorohex-1-ene-1-sulfonic acid
Chemical Formula: $C_6H_2F_{10}O_3S$
Exact Mass: 343.9565
Source: Barzen-Hanson et al. 2017

Additional Knowledge Gaps

- In lieu of reference mass spectra, *in silico* fragmentation prediction can be a powerful tool, but there is a lack of understanding of the accuracy of these predictions for PFAS.
- Benchmark performance for any NTA workflow needs to be established.
  - Requires common communication of performances in study results
  - “Good” performance must be empirically determined.
- There are no reference materials or reference data for evaluating and/or comparing NTA performance between laboratories.

Conclusions

- Non-targeted analysis and suspect screening analysis are powerful tools for the identification of unknown PFAS.
  - But there are sources of uncertainty and bias that must be addressed within any NTA workflow.
- Through the SERDP project, NIST is developing reference data and data tools to address current NTA data gaps.
- Future webinars and training programs are being developed.
- An interlaboratory study (and first glance at using the new database!) is planned for the end of 2022.
- If you would be interested in a training and/or the interlaboratory study please contact us at pfas@nist.gov

Acknowledgements

- This project has been supported by SERDP, project ER20-1056

- NIST PFAS Researchers:
  - Jared Ragland, Nik Blonder, Alix Rodowa, Kathy Peter, Jessica Reiner
- Collaborators:
  - Field Group (OSU), Higgins Group (CSM), Xiao Group (UND), Helbling Group (Cornell), Ferguson Group (Duke)