Mapping Chemical Space in Non-Targeted Analysis

Gabrielle P. Black[1], Charles Lowe[2], Tarun Anumol[3], Jessica Bade[4], Kristin Favela[5], Christine Fisher[6], Yong-Lai Feng[7], Alan Hood[8], Ann M. Knolhoff[6], Andrew D. McEachran[3], Jamie Nunez[4], Katherine T. Peter[9], Natalia Quinete[10], Jon Sobus[2], Eric M. Sussman[11], William Watson[5], Antony J. Williams[2], Samanthi Wickramesekara[8], and Thomas M. Young[1]

[1]Department of Civil & Environmental Engineering, University of California Davis, Davis, CA, USA , [2]U.S. EPA, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, NC, USA , [3]Agilent Technologies, Inc., Santa Clara, CA, USA , [4]Pacific Northwest National Laboratory, Richland, WA, USA , [5]Southwest Research Institute, San Antonio, TX, USA , [6]U.S. Food and Drug Administration, Center for Food Safety and Applied Nutrition, College Park, MD, USA , [7]Health Canada, Ottawa, ON, Canada , [8]U.S. Food and Drug Administration, Center for Devices and Radiological Health, Silver Spring, MD, USA , [9]University of Washington Tacoma, Center for Urban Waters, Tacoma, WA 98421, [10]Department of Chemistry and Biochemistry, Institute of Environment, Florida International University, North Miami, FL, USA, [11]MCRA LLC, Washington, DC, USA

INTRODUCTION

The methods and software tools implemented in non-targeted analysis (NTA) are numerous, allowing for highly customizable workflows with varying chemical space coverage. Predicting or defining this applicability domain for each workflow remains a challenge. An additional need therefore exists for approaches that can define the region of chemical space detectable by a selected (or planned) NTA method. Such approaches could theoretically enhance performance of NTA methods by minimizing false positives (i.e., instances when an undetectable compound is reported as being present) and increasing confidence in putative positive identifications that fall within the defined method applicability domain. The development of explicit chemical space mapping tools could also give researchers the ability to reduce the vast known chemical universe into lists of plausibly detectable and identifiable compounds. These Amenable Compound Lists (ACLs) could then be used as identification libraries and in annotation efforts as part of data processing workflows. Ultimately, understanding method boundaries will allow researchers to communicate and compare methods and results more easily, and better assess method needs on a project-by-project basis.

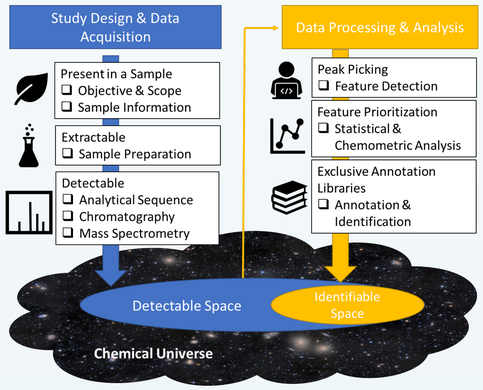


Figure 1: Filtering points in NTA workflows that can be used to define the detectable and identifiable space are consistent with the points outlined in the Study Reporting Tool (SRT, Peter et al 2021). Each point signifies a decision-making step with a high probability of influencing the chemical space coverage of a given method. The detectable space is defined by compounds' presence in a sample, the ability for the applied sample preparation method to extract such a compound, and the analytical platform configuration's ability to detect it. The identifiable space is dependent upon the post-acquisition data processing, beginning at peak picking (abundance thresholds, binning parameters, etc.), feature prioritization (fold change analyses, effects-directed, etc.), and the use of exclusive annotation libraries (only investigating features with mass matches, in silico hits, etc.).

ANTICIPATED APPLICATIONS

Prospective Use

Selection of Optimized Method Procedures
- Align experimental conditions with study goals

Preparation of Amenable and Unamenable Chemical Screening Lists
- Increase computation efficiency, improve downstream performance metrics, inform annotation and matching thresholds

Selection of Quality Control Mixtures
- Create physical mixtures of project-appropriate chemicals to be used for validation of chemical space boundaries

Retrospective Use

Annotation Prioritization
- Prioritize feature annotations present in Amenable Compound Lists

Improve NTA Performance Metric Accuracy
- Improve accuracy of compounds considered *true* versus *false* detections

THE CHEMSPACE TOOL

The ChemSpace Tool would incorporate multiple filtering steps based on method and instrumentation parameters to reduce input libraries to Amenable Compound Lists (ACLs) of compounds that are plausibly detectable and identifiable in analyzed samples. These filtering steps should be based on an ensemble of chemometric tools to determine which compounds likely are within the boundaries of the chemical space of a given method. The tool would partition chemical space into two parts, (1) the detectable space and (2) the identifiable space (Figure 1).

Eight steps have been identified as commonly used in NTA and/or are suspected to be highly influential to the resultant detectable space (Figure 2). Each step would produce an ACL that occupies a defined region of chemical space (e.g., water-soluble compounds, HLB extractable, LC-MS amenable, etc.). Ultimately, the eight ACLs resulting from these filtering steps would be compared, where the overlapping compounds define the detectable space (white overlapping region, Figure 2).

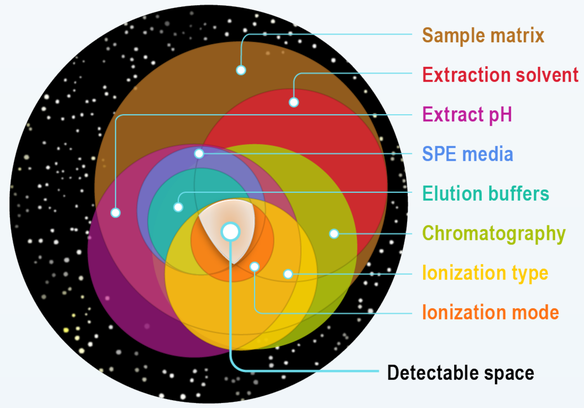
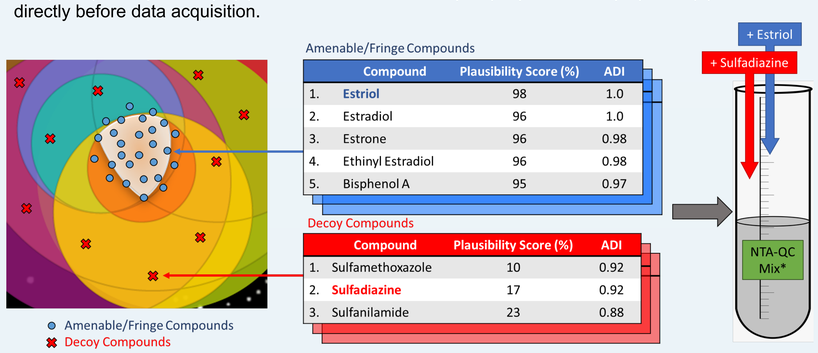


Figure 2: An example of the refinement of the detectable space based eight filtering steps (1) sample matrix (e.g., water), (2) extraction solvent (e.g., water), (3) extract pH, (4) solid phase extraction media (e.g., HLB), (5) elution buffer (e.g., methanol), (6) chromatography (e.g., Liquid Chromatography), (7) ionization type (e.g., electrospray), and (8) ionization mode (e.g., positive). Each filtering step would produce step-wise ACL's where the overlap zone represents the detectable space.

CHEMSPACE QC MIX

Based on the minimum, median, and maximum values of the descriptors used to create the detectable space ACL, a set of validation steps with quality control (QC) compounds will be suggested. A standardized NTA-QC mix will serve as the foundation for these steps, similar to the mixes proposed by Knolhoff et al (2021) and expanded upon based on ChemSpace filtering parameters. Each suggested compound would be accompanied by a Plausibility Score, or the likelihood in which the compound is included in the mapped chemical space, and an Applicability Domain Index (ADI), informing the reliability of the prediction. Compounds that fall within the mapped chemical space in addition to those outside the mapped space, or decoys, will be included in the ChemSpace QC mix. The mix would be spiked into (1) matrix and processed alongside samples (when possible), (2) in the extract of a sample (or pooled sample) and (3) in solvent directly before data acquisition.



REFERENCES

Knolhoff, A. M., Premo, J. H., & Fisher, C. M. (2021). A proposed quality control standard mixture and its uses for evaluating nontargeted and suspect screening LC/HR-MS method performance. *Analytical Chemistry*, *93*(3), 1596–1603. https://doi.org/10.1021/acs.analchem.0c04036

Peter, K. T., Phillips, A. L., Knolhoff, A. M., Gardinali, P. R., Manzano, C. A., Miller, K. E., Pristner, M., Sabourin, L., Sumarah, M. W., Warth, B., & Sobus, J. R. (2021). Nontargeted Analysis Study Reporting Tool: A Framework to Improve Research Transparency and Reproducibility. *Analytical Chemistry*, *93*, 13870–13879. https://doi.org/10.1021/acs.analchem.1c02621

CONTACT

Gabby Black
gnpecora@ucdavis.edu

Charles Lowe
lowe.charles@epa.gov

PROSPECTIVE USE CASE

The ChemSpace tool may aid researchers in understanding the influence of their study design, data acquisition, data processing and analysis methods on the chemical space covered. Deconstructing each of these steps can play a key role in methods development and allow researchers to modify their methods that expand upon or refine their chemical space coverage prior to analyzing samples .

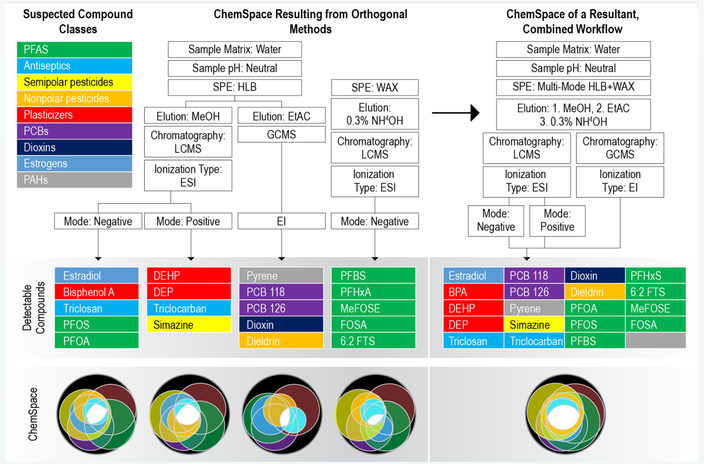


Figure 4: A researcher has a list of compound classes they suspect to be present in drinking water. Prior to choosing a sample preparation workflow, using predictive models they can evaluate the chemical space coverage of four different methods they are considering. Based on ChemSpace coverage of each, they find that a method combining of various aspects of each method provides the ChemSpace coverage needed to encompass compounds in each of their suspected compound classes.

RETROSPECTIVE USE CASES

The ChemSpace tool may prove to be equally as beneficial in a retrospective sense as it pertains to annotation and performance evaluation. Often in NTA, the number of plausible annotations far exceeds the number of detected features seeking to be identified. Frequently, researchers use filtering tools to eliminate unlikely or implausible structures based on retention time, platform amenability, etc. before attempting to annotate a feature. The ACL will offer the opportunity to further eliminate implausible structures in tandem with other filtering steps, increasing the confidence in annotation (Figure 5a).

Implementing the ChemSpace tool in a *post hoc* example can allow for third party evaluation of vastly different methods by adjusting the evaluation based on chemical space coverage of the individual methods. For example, normalizing sensitivity based on the different chemical spaces covered by each lab's methods results in more comparable sensitivity results. (Figure 5b).

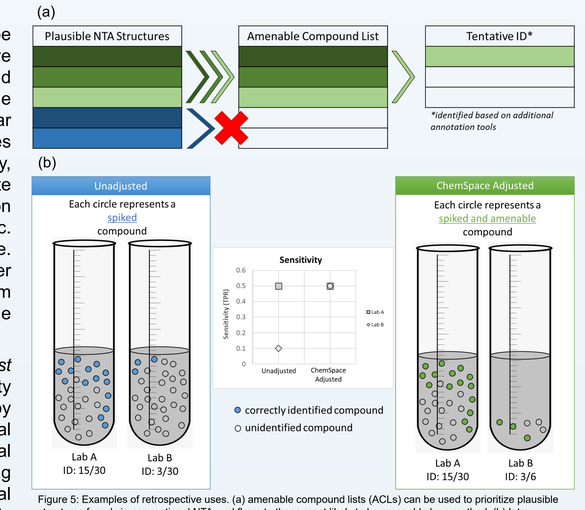


Figure 5: Examples of retrospective uses. (a) amenable compound lists (ACLs) can be used to prioritize plausible structures found via conventional NTA workflows to those most likely to be amenable by a method. (b) Intra-laboratory comparison of sensitivity (and other performance metrics) can be "normalized" by evaluating reported results in the context of chemical space. In this example, Lab B's chemical space covers only 6 of 30 spiked compounds. The true positive rate of their detections is 10% when left unadjusted for chemical space but increases to 50% when chemical space coverage is considered. When considering amenable chemical space, the sensitivity of the two labs are comparable.

OUTLOOK

The benefits of the ChemSpace tool spans from cutting down on method development time, improving annotation prioritization and overall accuracy, enhancing method transferability, and providing context for methods and results. In addition to transparent and detailed reporting of all workflow steps, chemical space delineation would allow researchers not only to compare results on an intra-laboratory or intra-project basis but would also allow more streamlined adoption of existing methods to new projects. Most importantly, understanding chemical space provides important context for results, thus allowing researchers and readers to differentiate whether un-detected compounds are truly absent from the sample or were not amenable to the method. While the framework for *what* this tool could encompass has been developed, many details of *how* this tool will perform are still needed. As we start to build and test this tool, we ask that chemometrics experts and chemists alike step forward to help fill modeling gaps, and for researchers to begin using some of the already available tools to discuss chemical space in NTA reporting.