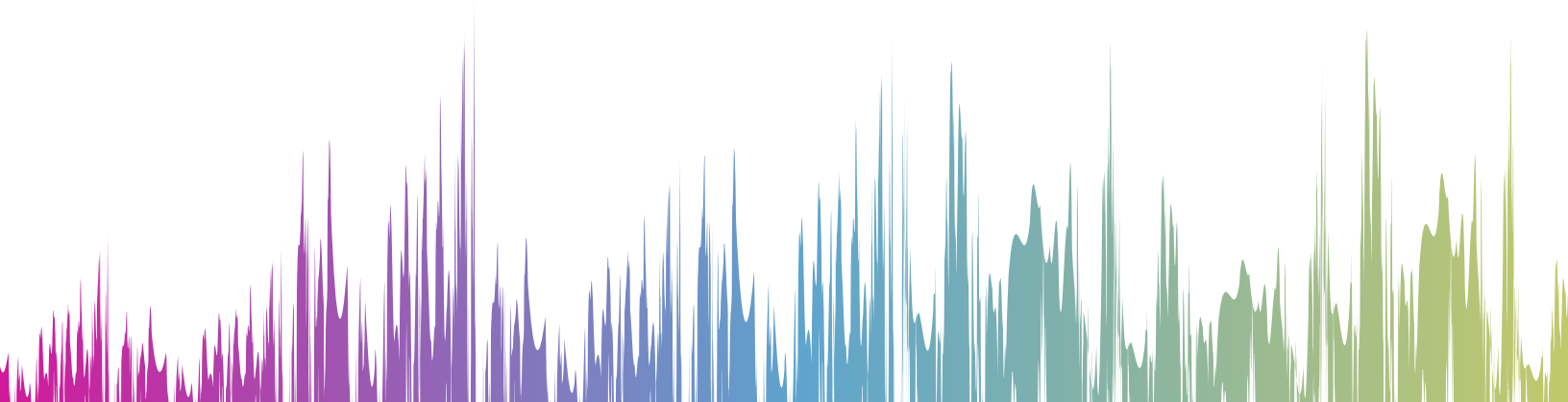# Nontarget Analysis of Organic Chemicals in the Environment

Synthetic organic chemicals—man-made carbon-containing molecules—are ubiquitous in modern life. They provide numerous societal benefits and are found in industrial chemicals, pesticides, pharmaceuticals, plastics and their additives, and personal care product ingredients. As such, thousands of organic chemicals are produced and sold worldwide, in large quantities (more than 500 tons per year), while thousands of new chemicals are continuously discovered and catalogued.

The lifetime of some organic chemicals extends beyond their intended use, eventually leading to their dispersal in the environment, where they are commonly designated "organic contaminants." Chemicals from industrial or consumer products are purposefully released into the environment—pesticides, for example—or may be discharged to aquatic environments through wastewater, vented to the atmosphere through exhausts, or released to terrestrial environments through biosolids or landfills.

Understanding potential exposure and health effects associated with organic contaminants, pinpointing their sources, and measuring the efficacy of strategies to contain, treat, or reduce them requires analytical tools capable of measuring and identifying low-level contaminants in environmental media.

Mass spectrometry (MS)—an analytical technique that measures ionized molecules' mass–to–charge ratio

(m/z)—is a preferred method for detecting organic contaminants. MS can determine the number and type of atoms in a molecule but not their full connectivity or arrangement. Reference standards—samples with known composition and purity—are then used to identify the measured analytes. Reference standards enable calibration of MS response to chemical concentration, an essential aspect of quantitative chemical risk assessment. When reference standards are used in MS assays, methods are termed "targeted analysis."

Targeted Analysis

Targeted analysis has been the preferred technique for measuring organic contaminants for more than 60 years. It has provided essentially all the available data on the environmental fate of organic contaminants. In targeted analysis, an analyte is identified based on retention time and m/z, and its concentration is determined by intensity as compared with a reference standard. However, obtaining reference materials needed for targeted analysis can be challenging in cases where chemicals are difficult to synthesize or purify in the laboratory or are not commercially available. Such is

Targeted analysis can only confirm what is searched for in a sample.

	Samples	Signal Comparison With	Outcome
Targeted Analysis	Environmental or biological sample AND reference standard	Reference standard	Chemical identity and concentration
Nontargeted Analysis	Environmental or biological sample	Chemical databases and spectral libraries	Proposed structures of chemicals

Figure 1: Targeted vs. Nontargeted Analysis.

the case when contaminants of interest are substances with unknown or variable composition, by-products of industrial synthesis, or transformation products formed from chemical or biological reaction during use or disposal. Even if all reference materials of interest could be obtained, the number of analytes currently measured in a single targeted analysis would remain on the order of a few thousand. In contrast, the number of chemicals with physicochemical properties and use-patterns suggestive of potential environmental dispersal is on the order of tens of thousands. As such, targeted analysis may produce a blind spot in the range of measurable analytes, precludes measuring previously unreported or unexpected contaminants, and subsequently inhibits comprehensive risk characterization for organic chemicals.

Nontarget Analysis

Analytical chemists have developed nontargeted screening approaches that leverage the unique capabilities of high-resolution, accurate-mass (HRAM) instrumentation to expand capabilities and address limitations in targeted analysis methods. HRAM can record mass spectral data on all analytes entering the MS and their associated fragment ions. This approach provides information on the potential chemical structure of unknowns, greatly expands the number of detectable molecules in a single analysis compared with targeted approaches, and enables reuse of data to detect newly discovered compounds. HRAM is by no means a new technology and has long played a role in the identification of unknown organic compounds. However, recent advancements in instrument sensitivity and scanning speed, coupled with commercial availability and lower operating costs, has made HRAM a potential option for routine environmental analysis. With the increased availability of HRAM instrumentation, environmental chemists have gained access to the necessary data to perform high-throughput, comprehensive, nontargeted screening of environmental samples. While HRAM instrumentation has driven progress in nontargeted analysis, concurrent and continuous advancements in allied technologies—such as chromatography and ionization techniques—influence the chemical universe detected by HRAM analysis and can greatly impact data quality.

Whereas HRAM experiments of the past may have sought to acquire data to support the identification of a few, often new, contaminants, primarily via expert interpretation and annotation, most contemporary HRAM research seeks to identify thousands of substances in a single analysis. Given these ambitious goals, analysts must forgo intensive data interpretation in favor of automation.

Nontarget analysis can identify far substances in a sample, suspected or otherwise.

Tools that Support Nontarget Analysis

Data analysis processes for HRAM data include various commercial and open-source software tools, including:

» Spectral libraries, used for assigning plausible chemical formulas or structures based on matches to experimental or computational mass spectra, and

» Chemical structure and information databases, used for storing and searching computer representations of molecules, chemical fate and effects properties, literature citations, patent information, and production volumes.

Spectral libraries minimally consist of chemical identifiers and structure representations, and mass spectra collected from reference standards or generated by computer predictions. The dependence on extant reference spectra when conducting spectral library searching for structure annotation may suffer from the same limitation as targeted analysis. Namely, compounds that cannot be purchased or made and are not present in the library may never be matched. Computational mass spectrometry tools alleviate reference standard dependency by using models or rules to predict chemical behavior within a mass spectrometer. Automated identification of organic small molecules by mass spectrometry remains an evolving field of research. As such, researchers and practitioners of HRAM nontarget analysis are grappling with challenging questions regarding quality control and assurance, and performance evaluations.

Chemical structure databases and their associated metadata play a role in current implementations of HRAM nontarget analysis. Namely, these databases serve as sources of plausible contaminant structures. Where no structure can be assigned from a chemical database, researchers may implement computationally expensive structural generation approaches or revert to laborious expert judgement and experience for proposing structures. In addition to the computational MS tools described above, various techniques can be used to rank possible candidate structures based on chemical and nonchemical metadata. For instance, where chemical structure references are found in scientific literature, they are typically for highly studied compounds whose production and use might lead to environmental

dispersal. Well-curated chemical databases are more reliable for nontargeted analysis because the structural representations and identifiers have been cleaned and harmonized. While open databases designed specifically for environmental applications, such as the U.S. EPA's CompTox Chemicals Dashboard, are improving accessibility, commercial databases, like the Chemical Abstracts Service Database or the Wiley Registry of Mass Spectral Data, still remain some of the largest, best annotated, and well-curated sources of chemical information for HRMS nontarget analysis.

State of the Science

A critical and ongoing debate in the nontarget analysis community centers around the question, "What constitutes definitive chemical structure identification?" While some researchers favor annotation schemes employing identification levels based on various lines of experimental evidence, others advocate for the traditional approach of only reporting identifications confirmed with authentic reference standards. Adherents to the latter approach argue that without reference standards, identifications could be rendered suspect or meaningless due to unacceptably high false-positive rates. Alternatively, proponents of reference standard-free identification might argue that methods requiring confirmation with authentic standards have unacceptably high false-negative rates since they provide chemical information on only a tiny fraction of the molecules present in a sample. Ensuring the continued utility of nontargeted analysis—especially considering recent interest in using nontarget analysis within regulatory frameworks—requires consensus from the community regarding consistent and proper reporting practices for contaminant identification.

There is a need for consistent reporting practices for contaminant identification.

Application of Nontarget Analysis Environmental Assessment

At present, most HRAM nontarget analyses only address the question of "Which contaminants are present?" and does not attempt to answer, "How much con-

taminant is present?" For application of nontarget data in quantitative chemical risk assessment, it is feasible to take a tiered approach employing nontarget screening to identify contaminants relevant to a system, followed by targeted analysis to quantify contaminants in that system. A more time- and cost-efficient solution would be reached if nontarget analysis predicted contaminant concentrations with uncertainty bounds. Therefore, researchers are working to develop techniques for generating instrument response factors based on structure–property relationships and machine learning.

However, in the absence of definitive identification and concentrations, researchers are already using the data provided by nontarget analysis to holistically evaluate chemical occurrence. Such approaches couple occurrence data for unidentified contaminants with other monitoring techniques, such as targeted analysis, water quality data, treatment performance assessment, or toxicity testing, and can supplement environmental assessment. Further, these techniques can be used to advance knowledge of organic contaminant fate and potential effects. These approaches are poised to benefit from big data techniques and will undoubtedly revolutionize how organic contaminants are studied in environmental systems.

Outlook

HRAM nontarget analysis has already proven valuable for characterizing organic contaminants in environmental samples. In particular, nontarget analysis has been employed for identifying previously unreported or unexpected organic pollutants in various aquatic environments, including surface waters, wastewater effluents, storm water runoff, and agricultural runoff. Additionally, nontarget analysis has helped to identify organic contaminants in indoor environments such as chemicals in house dust and building materials. Furthermore, nontarget analysis has provided a robust framework for the high-throughput identification of chemical transformation products, such as those formed during wastewater treatment. Finally, efforts to link nontarget analysis with bioassays and epidemiological endpoints promise to provide unique insights into links between chemical exposure and human and ecological health. These early successes in the application of HRAM nontarget analysis to environmental systems, and the wealth of vigorous research in this area, provide optimism for the future utility of these techniques in helping environmental chemists address the need for a comprehensive understanding of organic chemical environmental fate.

Acknowledgments

Copyright