



# EPA Public Access

Author manuscript

*Environ Sci Technol.* Author manuscript; available in PMC 2024 February 28.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Environ Sci Technol.* 2023 February 28; 57(8): 3075–3084. doi:10.1021/acs.est.2c06804.

## Demonstrating the Use of Non-targeted Analysis for Identification of Unknown Chemicals in Rapid Response Scenarios

**John T. Sloop,**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina 27709, United States; Oak Ridge Institute for Science and Education (ORISE) Participant, Research Triangle Park, North Carolina 27709, United States

**Alex Chao,**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina 27709, United States

**Jennifer Gundersen,**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Environmental Measurement and Modeling, Narragansett, Rhode Island 02882, United States

**Allison L. Phillips,**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessment, Research Triangle Park, North Carolina 27709, United States

**Jon R. Sobus,**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina 27709, United States

**Elin M. Ulrich,**

---

**Corresponding Author Seth R. Newton** – Newton.Seth@epa.gov.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.2c06804>.

Materials; quality control; sample selection and preparation; instrumental analysis; data processing; additional results of mock scenarios 2 and 3; “known unknowns” vs. “unknown unknowns”; hazard comparison module discussion; detailed description of the five steps in the data processing workflow; top 10 candidate features for mock scenario 2 via MPP output; 14 features further inspected during mock scenario 3; tracer compound QC results for each mock scenario; common LC-MS instrumental parameters used in all three LC-MS methods; LC mobile phase gradients used; four situations encountered in any NTA study; chromatogram of matrix dilutions for mock scenario 1; screenshot of MS/MS compound identification results; and hazard report generated for mock scenarios 1 and 2 (PDF)

An example of a full hazard report from the HCM and specific parameters used for various tools in data processing (XLSX)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.est.2c06804>

The authors declare no competing financial interest.

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

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina 27709, United States

**Antony J. Williams,**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina 27709, United States

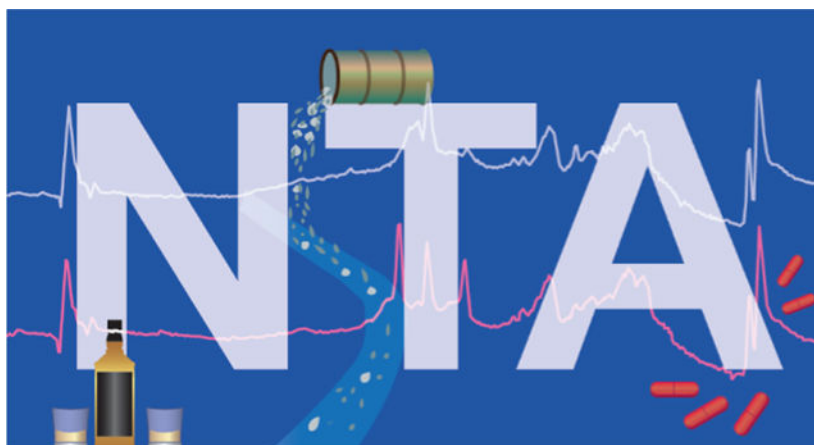
**Seth R. Newton**

U.S. Environmental Protection Agency, Office of Research and Development, Center for Computational Toxicology and Exposure, Research Triangle Park, North Carolina 27709, United States

## Abstract

Several thousand intentional and unintentional chemical releases occur annually in the U.S., with the contents of almost 30% being of unknown composition. When targeted methods are unable to identify the chemicals present, alternative approaches, including non-targeted analysis (NTA) methods, can be used to identify unknown analytes. With new and efficient data processing workflows, it is becoming possible to achieve confident chemical identifications *via* NTA in a timescale useful for rapid response (typically 24–72 h after sample receipt). To demonstrate the potential usefulness of NTA in rapid response situations, we have designed three mock scenarios that mimic real-world events, including a chemical warfare agent attack, the contamination of a home with illicit drugs, and an accidental industrial spill. Using a novel, focused NTA method that utilizes both existing and new data processing/analysis methods, we have identified the most important chemicals of interest in each of these designed mock scenarios in a rapid manner, correctly assigning structures to more than half of the 17 total features investigated. We have also identified four metrics (speed, confidence, hazard information, and transferability) that successful rapid response analytical methods should address and have discussed our performance for each metric. The results reveal the usefulness of NTA in rapid response scenarios, especially when unknown stressors need timely and confident identification.

## Graphical Abstract



## Keywords

non-targeted analysis; LC-MS; high-resolution mass spectrometry; rapid response; hazard comparison

---

## 1. INTRODUCTION

Every year in the U.S., there are thousands of releases of chemicals into the environment, which may threaten public health and/or ecological systems.<sup>1,2</sup> While some events, such as the Deepwater Horizon oil spill and pollutant release caused by the storm surge during Hurricane Harvey, are more well known, small-scale events are extremely common, with over 25,000 calls logged by the National Response Center (NRC), a part of the U.S. Coast Guard, reporting discharges into the environment during 2021.<sup>2-4</sup> Of these 25,000 calls, almost 30% initially reported the discharge to be of an unknown composition.<sup>2</sup> Of those discharges of unknown composition, over 70% were reported to penetrate a body of water near the spill.<sup>2</sup> Various state and federal agencies, including the U.S. Environmental Protection Agency (EPA), are tasked with responding to such incidents and must rapidly identify unknown chemicals.<sup>5-7</sup> These agencies have long relied on targeted analytical methods for identifying and quantifying specific chemicals. However, there is no systematic approach to elucidating the identity of an unknown chemical.

A recent publication by Phillips *et al.* outlined the type of work that the rapid response community performs, the variety of rapid response situations that occur, and the applicability of non-targeted analysis (NTA) as a potential tool for identifying unknown chemicals in such situations.<sup>1</sup> NTA is an emerging field of science with an explicit focus on the characterization of the chemical composition of a given sample without the use of a priori knowledge regarding the sample's chemical content.<sup>8</sup> NTA has applications in a variety of fields for the identification of various compounds, including polar organic pollutants in water samples, pesticide presence in agricultural products, changes in metabolite composition during manufacturing of food products, and PFAS identification in various environmental media.<sup>9-12</sup> While specific details of NTA can be found elsewhere (*e.g.*, [nontargetedanalysis.org](http://nontargetedanalysis.org)), it is worth mentioning that the overall goal of many NTA studies is to identify unknown chemicals with the highest level of confidence that the collected instrumental data and relevant metadata can provide.<sup>13-16</sup> Mining NTA data for confident chemical identifications has been a time-consuming, rigorous research activity in the past; however, advanced informatics tools and integrated workflows are becoming more automated, now making chemical identification *via* NTA a viable procedure to aid rapid response scenarios.

Data collected *via* gas chromatography-mass spectrometry (GC-MS) are, overall, very reproducible. Matching experimental GC-MS spectra to the contents of spectral databases is therefore common practice.<sup>17</sup> Liquid chromatography-ion mobility spectrometry-mass spectrometry (LC-IMS-MS) is also seeing increased use for environmental sample analysis, with recent applications to characterize chemical profiles related to firefighting foams and crude oils.<sup>18-20</sup> Despite these notable uses, most recent NTA workflows and informatics

tools have been specifically developed to support LC-high resolution mass spectrometry (LC-HRMS) applications because of the relative lack of reference spectra in spectral libraries when compared to GC-MS and the ability to easily detect the molecular ion using LC-HRMS. The various tools developed for LC-HRMS studies have made the process of assigning chemical identities to spectral features simpler and have increased identification confidence. Inconsistencies from study to study persist, however, as do the specific data processing approaches used to arrive at a final chemical identity. While no method can truly be “one size fits all”, the introduction of a general NTA method would be beneficial for both existing NTA researchers (as a starting point before performing further novel structure elucidation) and new NTA researchers (as an introduction to how NTA is performed). For these reasons, LC-HRMS was the instrumental approach utilized in this work.

In a rapid response situation where chemical composition is unknown, there are at least four metrics that should be considered when developing a suitable analytical method to aid chemical identification. The first metric is the speed of the analysis. The time needed to deliver results on the identity of chemical(s) present during a release into the environment should be as quick as possible to inform relevant stakeholders about the nature of the chemical(s) and potential danger(s) associated with the release.

The second metric is the confidence in the eventual chemical identification(s). Importantly, it is not possible to truly confirm the identity of a detected chemical without possessing and analyzing a standard of that chemical and comparing it to the sample. There are, however, many individual and orthogonal pieces of information obtained from an LC-HRMS analysis (*e.g.*, the observed mass-to-charge ratio [ $m/z$ ] of the parent ion and isotopologues, retention time [RT], ions associated with the observed compound [adducts or isotopologues], and MS/MS fragmentation patterns) that can be utilized to assign an identity to a chemical at a defined level of confidence. The identification confidence scale proposed by Schymanski *et al.* is routinely used in NTA studies for defining the level of confidence of a chemical identification (levels 1–5), specifics of which can be found elsewhere.<sup>21</sup> While a true level 1 identification is seldom reached during NTA studies, the goal of the work presented here was achieving the highest level of confidence in as short a time as possible and aiming for either a level 2 or 3 identification so a structure can be assigned.

The third metric is the degree of hazard assessment that can be performed for each identified analyte. This is possible when a molecular structure is assigned because there can be some level of toxicity assessment performed. Responders should therefore be informed of which chemicals were released, and the extent to which those chemicals may pose risks to human and/or ecological health. Furthermore, if an elevated health risk is apparent, it is then necessary to know which receptors may be most sensitive to harm and the pathway(s) (*e.g.*, inhalation of outdoor air, dietary consumption of contaminated food products, *etc.*) through which exposures are most likely to occur. Summarizing and disseminating this information in a manner that is easily understandable for responders should be a clear aim of any rapid response situation.

The fourth and final metric for success is the transferability of the designed NTA method/workflow. The end goal for this body of research is to enable federal, regional, state, and

local laboratories to incorporate NTA as a supplement in situations when the chemical of interest is not easily identified by targeted approaches. The work presented here serves as a guide for how this could be done. Specifically, the sample preparation approaches, data collection methods, and data analysis methods convey suitable strategies to rapidly identify stressors in impacted samples and subsequently relay relevant information to responders.

In this work, to address the performance metrics, we designed three different mock scenarios involving chemicals whose identities were unknown to a blinded analyst, intended to mimic situations in which a rapid response would be required. The mock scenarios were chosen to cover various sample media and chemical classes from real scenarios potentially encountered. The three mock scenarios performed in this work were designed to test the ability of an NTA method to characterize: (i) a surrogate of an unidentified nerve agent spiked into a beverage used to poison an individual; (ii) surrogates of novel illicit drugs from a raid on a home; and (iii) an industrial chemical spill into surface water. It was assumed that the samples were received after initial targeted methods failed to identify the unknown chemical(s) known to exist in the samples. Thus, the novel NTA method developed for this work focuses on either one or a small number of chemicals for which there is some a priori knowledge, allowing the analyst to focus on these few chemicals for as confident of an identification in as little amount of time as possible. This differs from typical NTA methods in which the analyst is attempting to identify as many chemicals as possible and there is typically little to no prior knowledge about the contents of the sample. The NTA method employed herein uses two chromatographic approaches: the first to rapidly identify an appropriate sample dilution and ionization polarity, and the second to obtain better separation. It then utilizes both MS and MS/MS data *via* five data processing streams using multiple pieces of data processing software to arrive at a consensus identification, which stands in contrast to most NTA methods that typically use one or two data processing streams. An important aspect of the workflow was the use of a recently developed data processing tool, the NTA WebApp, which markedly decreased the times required for data processing and chemical identification.<sup>22</sup> Finally, we introduce the newly developed Hazard Comparison Module (HCM), a web-based cheminformatics module developed by the EPA that rapidly assembles hazard information from a collection of sources to prepare a hazard profile for all chemicals searched on the module.<sup>23</sup> While the examples shown here are not intended to be inclusive of every possible situation, this is a first step at showcasing the capabilities of HRMS-based NTA approaches as an additional analytical tool for rapid response situations to supplement targeted methods when elucidating chemical identities.

## 2. MATERIALS AND METHODS

The NTA Study Reporting Tool was used in the preparation of this article to guide appropriate reporting of information relevant to NTA studies.<sup>24,25</sup>

### 2.1. Sample Selection and Preparation.

A total of three mock scenarios were planned and conducted, each designed to mimic different situations in which a rapid response for the identification of an unknown chemical(s) would be required. The scenarios were designed to progressively become

more complex to test the boundaries of the NTA method. Spiked chemicals came from (or represent with similar structures) chemical classes that rapid responders are likely to encounter. Sample media cover a range of real-world settings and locations for rapid response situations. Details on specific materials used in this work can be found in Supporting Information Section S1.

Each mock scenario involved two analysts. Analyst 1 was charged with planning and preparing the scenario, and analyst 2 was responsible for selecting/creating an appropriate analytical method (including any sample preparation, data acquisition, data processing, and compound identification steps). One day before analyst 2 was to receive the samples, analyst 1 informed analyst 2 of basic information about the situation (*i.e.*, any observable information that an on-scene coordinator or responder would potentially provide during an actual event) and informed analyst 2 when to expect sample receipt. Analyst 2 was blinded to the identity of the spiked chemicals (and was also blinded to the nature of the scenario until one day prior to beginning laboratory work for that scenario). The one-day warning gave analyst 2 time to research appropriate extraction methods for the suspected chemical class and prepare the laboratory for operations (*e.g.*, cleaning the instrument, preparing mobile phases, *etc.*).

Because the analyte concentration was unknown to analyst 2, the sample and matrix blank were diluted *via* serial dilution into a series of 10-, 50-, 100-, 500-, and 1000-fold dilutions. This was necessary to reduce the risk of contaminating the instrument or saturating the detector with highly concentrated samples. All sample dilutions and blanks were spiked with 20  $\mu\text{L}$  of a mix containing the isotopically labeled tracer compounds listed in Supporting Information Section S1 to monitor instrument performance. This tracers mix was prepared in acetonitrile, with all tracer compounds being at a final concentration of 5  $\mu\text{g}/\text{mL}$ . In each mock scenario, samples and blanks were also further diluted 10-fold to match the LC-MS instrumental starting conditions by mixing 0.1 mL of the prepared sample with 0.9 mL of 1% (v/v) formic acid in DI H<sub>2</sub>O to make the final solutions for instrumental analysis. Specific details on quality control (QC) can be found in Supporting Information Section S2.

The first mock scenario involved identification of a surrogate of a chemical warfare agent that was spiked into an alcoholic beverage intended to poison an individual. The chemical chosen for this scenario was malathion (C<sub>10</sub>H<sub>19</sub>O<sub>6</sub>PS<sub>2</sub>, [DTXSID4020791](#)) spiked at a final concentration of 20  $\mu\text{g}/\text{mL}$  used as a surrogate for Novichok nerve agents, such as Novichok A-234 (C<sub>8</sub>H<sub>18</sub>FN<sub>2</sub>O<sub>2</sub>P, [DTXSID60896946](#)). The second mock scenario involved the identification of a surrogate of alprazolam (common brand name Xanax, C<sub>17</sub>H<sub>13</sub>ClN<sub>4</sub>, [DTXSID4022577](#)) and fentanyl (C<sub>22</sub>H<sub>28</sub>N<sub>2</sub>O, [DTXSID9023049](#)) from a surface wipe sample and a carpet sample spiked with 0.5 mL of 300 and 100  $\mu\text{g}/\text{mL}$  of the alprazolam and fentanyl surrogate, respectively. This scenario was intended to mimic a situation in which a clandestine drug laboratory (any location where illicit drugs are being illegally manufactured or processed, like an individual's home) was discovered. Finasteride (C<sub>23</sub>H<sub>36</sub>N<sub>2</sub>O<sub>2</sub>, [DTXSID3020625](#)) and  $\alpha$ -hydroxy alprazolam (C<sub>17</sub>H<sub>13</sub>ClN<sub>4</sub>O, [DTXSID60190613](#)) were chosen as surrogates for fentanyl and alprazolam, respectively. The structures of the chemicals for the first two mock scenarios can be found in Figure 1. Surrogates were used for the first two mock scenarios to minimize any potential risks

for the analysts associated with these dangerous chemicals and because the chemicals of concern are controlled substances which require a license by the U.S. Drug Enforcement Administration to obtain. The third mock scenario mimicked a situation where identification of various components of an industrial mixture spill (original sample diluted 100-fold) in surface water was required. Specific details of sample selection and preparation for each of the mock scenarios can be found in Supporting Information Section S3.

## 2.2. Instrumental Analysis.

Three LC-MS methods were used during this study, with all data collected using an Agilent 1290 Infinity high pressure (HP) LC (Agilent Technologies, Palo Alto, CA), interfaced with an Agilent 6530B Quadrupole/time-of-flight HRMS and electrospray ionization (ESI). The first was a 9 min, LC-MS “rapid range finding” method, intended to perform quick chromatography for the determination of appropriate sample concentration and ionization polarity (ESI+ and/or ESI-). The second method was a longer, 30 min LC-MS method intended to achieve greater chromatographic separation for selected sample dilution in the chosen ionization mode(s). The third method was a 30 min LC-MS/MS method, operating under the same LC conditions as the 30 min LC-MS method, using data-dependent acquisition (DDA) with the ion(s) of interest added to the preferred ion list. Specific instrumental parameters and details for each of these methods can be found in Supporting Information Section S4.

## 2.3. Data Processing.

Most NTA studies attempt to identify as many chemicals as possible for any given sample set. Because of the enormous amount of data generated from larger sample sets, it is typically feasible to use only one or two data processing approaches (*e.g.*, generating candidate chemical lists for detected MS features or matching collected MS/MS spectra to predicted or reference MS/MS spectra). In a rapid response scenario, however, the analyst is attempting to use NTA to identify a limited number of features (prioritized by sample intensity after blank subtraction) from a very small set of samples. It is therefore possible to use multiple data processing approaches to arrive at a consensus identification without substantially increasing the time required to arrive at a conclusion. Subtle differences exist in seemingly similar chemical identification approaches (*e.g.*, formula matching vs formula prediction), and multiple approaches can reduce the probability that the correct identification was overlooked and provide weight of evidence for a given identification. Therefore, multiple data processing approaches were explored, and a final set of five unique approaches were applied to each scenario, in which three utilized MS data and two utilized MS/MS data. These approaches, in tandem with some a priori information provided to analyst 2 about the nature of the scenarios, were the basis of the novel, focused NTA method applied in this work. Figure 2 shows a schematic of the overall data processing workflow used, and a more comprehensive explanation and specifics of each approach can be found in Supporting Information Section S5, and Table S1 walks through each step with the specific results of mock scenario 1 to serve as an example of the rationale behind each step in the process.

The results from all five data processing approaches were considered when assigning a chemical identification to any given feature in each of the mock scenarios, and the conclusion of chemical identity for a feature required the analyst's judgement when weighing the evidence from each approach. Chemical identifications were assigned a level of confidence based on the identification confidence scale by Schymanski *et al.*, ranked from levels 1 to 5.<sup>21</sup>

## 2.4. Toxicity/Hazard Assessment Using Proof-of-Concept Cheminformatic Modules.

A hazard report was generated using a software tool which has been developed inside the EPA. The HCM is one module of a series of cheminformatics modules that have been developed to test capabilities, functionality, and workflows in proof-of-concept implementations prior to migrating those capabilities to production software applications such as the CompTox Chemicals Dashboard.<sup>26</sup> The HCM is a web-based implementation of the capabilities described in the original work of Vegosen and Martin and extended with additional functionality to integrate to other modules.<sup>23</sup> The HCM represents a compilation of data generated within the agency and sourced from public databases, literature, and real-time quantitative structure-activity relationship predictions. The information assembled in the output of the HCM describes human health effects, environmental hazards to aquatic organisms, and environmental fate properties of each chemical. Hazard information is converted into scores of inconclusive, low, medium, high, or very high (I, L, M, H, or VH, respectively) based on a modified version of Design for the Environment criteria with final scores assigned based on the "trumping method", which selects the highest score from the most authoritative source as the integrated score.<sup>23</sup> In the proof-of-concept tool, three specific profiles are presently available: the full data, a site-specific screening profile (with a bias to repeat exposures, persistence, and bioaccumulation), and an emergency response profile with a bias to acute toxicity and single exposures.

The HCM allows for chemical identifier inputs based on CAS RNs, chemical names, DTXSIDs (DSSTox substance identifiers available on the Dashboard), and SMILES. In this way, any candidate chemicals identified can be input to build a hazard profile for any single chemical or support a profile comparison for a set of chemicals. The resulting profile can then be exported in multiple formats, with the most generally consumable format being an Excel file (see Supporting Information Tables S2 and S3 for example, files).

A hazard profile report exported from the HCM represents the type of hazard assessment that can be provided to an on-scene coordinator after NTA work is used for initial sample and chemical characterization. For the current work, the HCM reports generated using data from each of the three mock scenarios considered only emergency response profiles, allowing a focus on the acute toxicity of tentatively identified chemical stressors.



### 3. RESULTS

#### 3.1. Mock Scenario 1: Alcoholic Beverage Spiked with Nerve Agents.

Based on the information received about the scenario from analyst 1, analyst 2 was searching for a single chemical of interest during rapid range finding. A single peak became observable in ESI+ in the 50-fold dilution sample that was not observable in the blank and increased in intensity in the 10-fold dilution sample. A screenshot of the ESI+ chromatogram in Agilent's Qualitative Analysis 10.0 of the 1000-, 50-, and 10-fold matrix blank/sample pair dilutions is shown in Supporting Information Figure S1. The peak of interest in these chromatographic spectra emerged in the 50-fold sample dilution (B, bottom) near RT = 6 min and was more apparent in the 10-fold sample dilution (C, bottom) near the same RT (6 min). It was determined that the 10-fold dilution was the best concentration of the sample to use, and the peak of interest was present in the LC-MS ESI+ mode. From the rapid range finding method, the most abundant  $m/z$  in the extracted mass spectrum of the chromatographic peak of interest was 331.0437. Following this, the longer MS method and the DDA MS/MS method, with  $m/z$  331.0437 included in the preferred ion list, were both used.

MPP matched the feature of interest to the MS-Ready formula  $C_{10}H_{19}O_6PS_2$  (match score 89.2, out of a maximum possible score of 100; SI 5.1). The molecular formula generator (MFG) tool on Agilent's Qualitative Analysis 10.0 (SI 5.2) yielded many potential formulae for the MS peak of interest and ranked  $C_{10}H_{19}O_6PS_2$  as the highest, with a score of 99.11 (out of a maximum possible score of 100). The WebApp search by mass (SI 5.3) yielded 49 potential hits for the feature of interest, with the top three matches based on number of data sources being malathion ( $C_{10}H_{19}O_6PS_2$ ), isomalathion ( $C_{10}H_{19}O_6PS_2$ ), and becampanel ( $C_{10}H_{14}N_4O_7P$ ). Because the number of data sources for malathion ( $n = 250$ ) was much greater than the next two potential matches of isomalathion ( $n = 33$ ) and becampanel ( $n = 17$ ), and because the formulae from MS-Ready formula matching and MFG were the same as the formula for malathion, malathion was chosen as the top candidate from the WebApp mass search.

Matching MS/MS data to PCDLs (SI 5.4) using Agilent's Qualitative Analysis 10.0 yielded two potential matches, neither of which was malathion, and both were scored very low (25.48 and 27.32, out of a maximum possible score of 100). Of note, malathion was returned as a match from the PCDL, but the PCDL did not contain an experimental malathion mass spectrum, so it did not receive a match score, nor was it ranked in comparison to the other two candidates. Matching to the CFM-ID *in silico* database using the WebApp (SI 5.5) yielded 55 potential matches. Of these matches, it was noted that malathion was one of the potential matches but scored very low. However, of all 55 of the potential matches returned from the Webapp's MS2 tool, malathion had the greatest number of data sources ( $n = 250$ ) compared to the next highest scoring candidates on the list ( $n = 28$ ,  $n = 16$ , etc.). Upon review of the data, the low scores from the MS/MS data processing approaches were most likely caused by a low abundance of fragment ions using our instrumental conditions (further discussion on MS/MS instrumental parameters can be found in Supporting Information Section S4). Because the chemical of interest did not fragment

well, the MS/MS spectra acquired were not of sufficient quality to produce a high-scoring match. Based on the evidence gathered from all five data processing approaches, analyst 2 correctly reported that the chemical identification was malathion at a level 2B identification.

### 3.2. Mock Scenario 2: Contamination of the Home with Illicit Drugs.

Spiked samples (and matrix blanks) used for this scenario were a surface wipe conducted over a benchtop in the laboratory and a carpet square. From rapid range finding, it was determined there were multiple peaks of interest in the chromatogram collected *via* ESI+ and the 50-fold dilution was the preferred concentration. Unlike the first mock scenario, where a single peak in the sample chromatogram that did not exist in the matrix blank was apparent, the existence of multiple peaks made it unrealistic to choose individual peaks of interest by visual inspection alone and prioritizing features by abundance after data collection was necessary. After rapid range finding, samples were acquired with the longer MS and DDA MS/MS methods, and the ion at  $m/z$  325.0928 was included in the preferred ions list for DDA MS/MS based on the results of rapid range finding. Features of interest for data processing were then selected by sorting the MPP output (SI 5.1) of data collected during the longer MS run by decreasing feature abundance after blank subtraction. Out of the top 10 most abundant features, the three features that eluted prior to chromatographic re-equilibration (RT < 20 min) were prioritized for further investigation. The top 10 features sorted by blank subtracted abundance are shown in the Supporting Information (Table S4). The formula from MPP and the experimental accurate masses of these three features were C<sub>17</sub>H<sub>13</sub>ClN<sub>4</sub>O at 324.0783 Da, C<sub>23</sub>H<sub>36</sub>N<sub>2</sub>O<sub>2</sub> at 372.2718 Da, and C<sub>11</sub>H<sub>15</sub>NO<sub>2</sub> at 193.1110 Da.

The details of the data processing for this scenario can be found in Supporting Information Section S6 and are very similar to the steps described in Section 3.1. Briefly, the three MS approaches were performed first, with priority given to the chemical with the greatest number of data sources, and the two MS/MS approaches were done after. The results from all five approaches (SI 5.1–5.5) were considered when proposing a putative identification. The first feature investigated (C<sub>17</sub>H<sub>13</sub>ClN<sub>4</sub>O at 324.0783 Da) was correctly reported by analyst 2 as  $\alpha$ -hydroxy alprazolam at level 2B. The second feature (C<sub>23</sub>H<sub>36</sub>N<sub>2</sub>O<sub>2</sub> at 372.2718 Da) was correctly reported as finasteride at level 2A. A screenshot of the MS/MS compound identification PCDL results for finasteride is shown in the Supporting Information as an example of how the PCDL results are shown (Supporting Information Figure S2). The third feature (C<sub>11</sub>H<sub>15</sub>NO<sub>2</sub> at 193.1110 Da) was reported as parbenate at level 2A (further discussed in Section 4).

### 3.3. Mock Scenario 3: Industrial Spill in Surface Water.

From rapid range finding, it was determined that there were many peaks of interest, both in ESI+ and ESI-data. It was also determined that both the 50-fold and 10-fold matrix blank/sample dilutions would need to be analyzed *via* the general MS method and the DDA MS/MS method. There were 10 ions of interest in ESI+ and 12 in ESI-manually selected from rapid range finding peaks of interest, and these were added to the preferred ion list for fragmentation during the MS/MS instrumental run.

Like the second mock scenario, the larger number of chemicals in this sample mixture made it unfeasible to choose individual peaks of interest by visual inspection alone. Features of interest for further analysis were selected by processing MS data output *via* MPP (SI 5.1) from the longer MS run and then sorting the combined results from both ESI+ and ESI- modes by decreasing feature abundance after matrix blank subtraction. A total of 14 features across both ESI+ and ESI- modes were selected for further analysis.

From both ESI+ and ESI- results, there were a total of four features identified at level 2 (A or B). These features are shown in Figure 3. Chemicals (A), (B), and (C) were identified *via* ESI- results, and chemical (D) was identified *via* ESI+ results. Chemical (A) is octyl hydrogen sulfate ( $C_8H_{18}O_4S$ , [DTXSID7042433](#)); (B) is decyl hydrogen sulfate ( $C_{10}H_{22}O_4S$ , [DTXSID8042428](#)); (C) is 6:2 fluorotelomer sulfonic acid ( $C_8H_5F_{13}O_3S$ , [DTXSID6067331](#)); and (D) is 2-(2-butoxyethoxy)ethanol ( $C_8H_{18}O_3$ , [DTXSID8021519](#)).

There were a total of two features identified (one from ESI+ and one from ESI-) at level 3. Interestingly, both features corresponded to the same unique chemical (it ionized in both ESI+/ESI-). Candidates were narrowed down to two very similar isomers, shown in Figure 3E,F. No evidence was observed to indicate which of the two isomers was the correct identity, so this was considered a level 3 identification. Chemical (E) is *N,N*-dimethyl-3-((perfluorohexyl)-ethylsulfonyl) aminopropanamine *N*-oxide ( $C_{13}H_{17}F_{13}N_2O_3S$ , [DTXSID80880983](#)) and (F) is *N*-[3-(dimethylamino)propyl]-3,3,4,4,5,5,6,6,7,7,8,8,8-tridecafluoro-*N*-hydroxyoctane-1-sulfonamide ( $C_{13}H_{17}F_{13}N_2O_3S$ , [DTXSID10868577](#)). The remaining eight features were identified at either level 4 or 5. More details on all features (polarity, measured accurate mass, RT, and ultimate identification levels), including those identified at level 4 or 5, are provided in Supporting Information Section S7 and in Table S5.

Based on the evidence gathered from all five data processing approaches for features investigated, there were six features (corresponding to five unique chemicals) assigned a structure and eight features/chemicals with no structure assignment. After analyst 1 delivered results to analyst 2, the specific AFFF mixture used in this scenario was revealed, and all five unique chemicals' assigned structures were confirmed from the literature.<sup>20,27</sup>

## 4. DISCUSSION

To summarize, the spiked chemicals of interest were correctly identified in mock scenarios 1 and 2, and major components of the AFFF mixture were identified in mock scenario 3. In mock scenario 1, the spiked chemical malathion was identified from a pure ethanol sample. In mock scenario 2, the spiked surrogates for alprazolam and fentanyl ( $\alpha$ -hydroxy alprazolam and finasteride, respectively) were identified from a surface wipe and carpet sample. During this scenario, based on the information provided to them by analyst 1, analyst 2 assumed there was likely more than one chemical of interest. During data analysis, they further investigated three features. The third feature investigated was identified as parbenate. While this was not an intentionally spiked chemical in the sample matrices, it is worth noting that the surface wipe sample used was intentionally performed over an area in the laboratory that was covered in dust to ensure that sufficient background was introduced into the sample. Therefore, this identification was not necessarily a false-positive,

but instead, likely an instance of a chemical structure assigned to a feature that was not a designed “chemical of interest” for the scenario. In mock scenario 3, major components of the AFFF mixture were identified, and all structural assignments were further confirmed *via* comparison to the literature.<sup>20,27</sup>

Performance was assessed with respect to the four previously defined metrics: (1) speed of analysis, (2) confidence in the chemical identifications, (3) degree of toxicity or hazard assessment that could be provided, and (4) transferability of the technique from analyst to analyst. Regarding the first metric (*i.e.*, speed), for all three mock scenarios, the speed of analysis and time required to reach an assigned chemical identification were within a timeframe typically allotted for targeted analysis (within 1–3 days after sample receipt; 24–72 h). The time required for the first mock scenario was 1 h of research prior to sample receipt plus 13 active hours of analysis (total time spent for sample preparation, instrumental run time, data collection, data processing, and data analysis) over the course of two days to arrive at the correct identification. The time required for the second and third mock scenarios was greater than that for the first (30 active hours over four days for mock scenario 2 and 68 active hours over 10 days for mock scenario 3), seeing as the complexity of the sample and its components increased with each mock scenario performed.

Regarding the second metric (*i.e.*, confidence), considering all mock scenarios performed, a structure was assigned to most features investigated (some features from mock scenario 3 were unable to have structures assigned). For features assigned a structure, all were level 2 per the identification confidence scale, except for the level 3 assignment in mock scenario 3. The structural assignments were proven post-analysis to be correct (either by confirmation of chemicals spiked during mock scenario 1 and 2 by analyst 1, or by consulting the literature for identified components of the commercially available AFFF mixture used in mock scenario 3).

While each of the three scenarios presented in this work mostly involved chemicals present in common databases such as NIST or DSSTox at medium/high concentrations, the applicability of this approach in the field of rapid response is still viable, as we believe many real-world rapid response scenarios would also fit this description. It is important to note that other situations (*i.e.*, with chemicals that have never been previously identified, not present in the databases used, and/or present at trace concentration levels are described further in Supporting Information Section S8) are not impossible using NTA but would require more time (weeks to months) and a slightly altered workflow, and the resulting identifications would potentially be less certain. This was exemplified in scenario 3, where some of the chosen chemicals of interest had never been identified, considering how more advanced *de novo* NTA data interpretation was used in the literature to determine structures of chemicals in the same AFFF mixture used in this work.<sup>20,27</sup> Had more time been allowed in this scenario, the features not assigned a structure may have been identified using a similar approach. In a real rapid response situation involving a complex mixture of unknowns, an analyst could very well provide initial results from the first 24–72 h of analysis, then spend more time further elucidating structures if requested. However, in these situations, increased analysis time must be weighed against the importance of identifying truly unknown compounds or narrowing down the list of candidate compounds to one

singular structural assignment. Furthermore, the expertise required to identify undocumented chemicals increases, while the methods used in this study are more easily transferable to federal, regional, state, and local laboratories that already perform analyses in rapid response situations. In a situation where only a formula could be assigned, the hazards of known structures with that formula could be considered for a worst-case scenario. The results of level 3 identifications described in mock scenario 3 serve as a good example of not narrowing the list of candidate compounds to a single structural assignment while still gathering useful information. Even though the NTA workflow failed to select a single structure as the chemical of interest, it narrowed down the list of possible candidates to two isomers, which were extremely similar and differed only at the location of an oxygen (either an –OH or =O group).

To address the third metric for success, for each mock scenario, a “hazard report” was generated using the HCM for each structural assignment. This newly developed tool can also predict 1–5 generations of transformation products *via* hydrolysis, abiotic reduction, and human biotransformation using the Chemical Transformation Simulator, a cheminformatic tool developed and hosted by the U.S. EPA (<https://qed.epa.gov/cts/>).<sup>28</sup> The hazard report for mock scenario 3 is shown in Figure 4, with the hazard reports for mock scenario 1 and 2 provided in the Supporting Information (Figures S3 and S4). The hazard report for mock scenario 3 includes structural assignments and 1 generation of breakdown products.

These reports take less than a few minutes to generate based on the number of compounds and generations of breakdown products included. The information provided in this report would aid on-scene coordinators in terms of identifying immediate/acute toxicity concerns for a variety of potential exposures. In this scenario, the main human health effects of concern for all compounds were oral acute mammalian toxicity, genotoxicity, mutagenicity, skin irritation, and eye irritation, and the ecotoxicity concern was acute aquatic toxicity. Because the chemical release in this scenario occurred in a body of water, responders would likely restrict human access to this area until remediation efforts were finished, due to the oral toxicity, skin irritation, and eye irritation concerns of the identified compounds. The flora/fauna present in this body of water could also be negatively impacted due to the compounds’ acute aquatic toxicity. A summary of the main exposure concerns from the hazard reports of mock scenarios 1 and 2 can be found in Supporting Information Sections S9.1 and S9.2. While the chemicals used in this work were stable (*i.e.*, not readily degradable), the HCM could be used to identify potential metabolites and breakdown products of readily degradable chemicals of interest.

To assess the final metric and showcase the transferability of this approach, a different individual assumed the role of analyst 2 for mock scenario 2 than for the other scenarios. This individual was a trained analytical chemist familiar with NTA approaches, but not intricately familiar with specific methods/workflows used in this project. This “new” analyst 2 received a briefing on the project a few days before the scenario began, not lasting more than a few hours. And even so, in each mock scenario, analyst 2, regardless of which individual was assuming that role, arrived at a confident identification for the chemical(s) of interest, in a timeframe consistent with rapid response situations. We believe this demonstrates the methods and workflows developed in this study could ultimately

be transferred to federal, regional, state, and local laboratories with some training to incorporate NTA into existing rapid response analyses. Transferability of the method could be further demonstrated in future work by having multiple analytical chemists from multiple laboratories performing the same scenarios using the same workflows to show within-method reproducibility.

Using the novel NTA method described in this work and focusing on identifying a small number of chemicals at a high level of confidence, the chemicals of interest in three rapid response mock scenarios were identified. The three mock scenarios presented herein showcase the applicability of NTA as an additional tool that laboratories responding to unknown chemical releases into the environment could utilize. The mock scenarios conducted used a variety of probable chemicals of interest or structurally similar surrogates and sample media meant to represent three simple real-world situations in which a rapid response would be necessary. The success of each mock scenario against the identified metrics for success was discussed, in which the level of success in each scenario increased as the complexity of the specific scenario decreased.

While making claims on the true concentration of the chemical release requires confirmation *via* comparison to an analytical standard, methods are being developed to estimate concentrations of compounds without the use of chemical standards *via* quantitative NTA (qNTA), which could improve this approach by providing environmental concentration estimates and a better risk (toxicity/hazard) assessment of the chemical(s).<sup>29,30</sup> The authors believe the work shown here demonstrates the successful use of NTA in the field of rapid response. Attempting this NTA approach in an actual scenario is the next step in further proving this method's applicability. We believe it is important to fully demonstrate this method as feasible for federal, regional, state, and local laboratories that already perform rapid response analyses.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This work was supported in part by an appointment to the Oak Ridge Institute for Science and Education (ORISE) Research Participation Program at the Office of Research and Development, U.S. EPA, through an interagency agreement between the U.S. EPA and the U.S. Department of Energy.

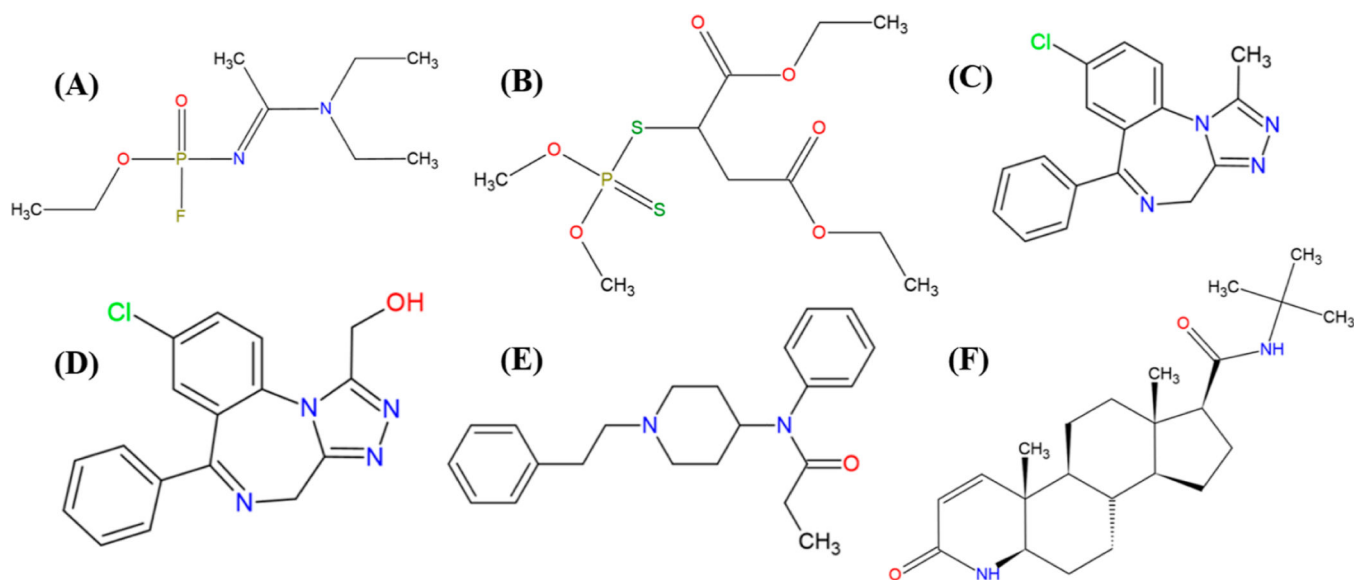
## REFERENCES

- (1). Phillips AL; Williams AJ; Sobus JR; Ulrich EM; Gundersen J; Langlois-Miller C; Newton SR A Framework for Utilizing High-Resolution Mass Spectrometry and Nontargeted Analysis in Rapid Response and Emergency Situations. *Environ. Toxicol. Chem* 2022, 41, 1117–1130. [PubMed: 34416028]
- (2). Agency, U. S. E. P. National Response Center. <https://www.epa.gov/emergency-response/national-response-center>.
- (3). Agency, U. S. E. P. Deepwater Horizon-BP Gulf of Mexico Oil Spill. <https://www.epa.gov/enforcement/deepwater-horizon-bp-gulf-mexico-oil-spill>.

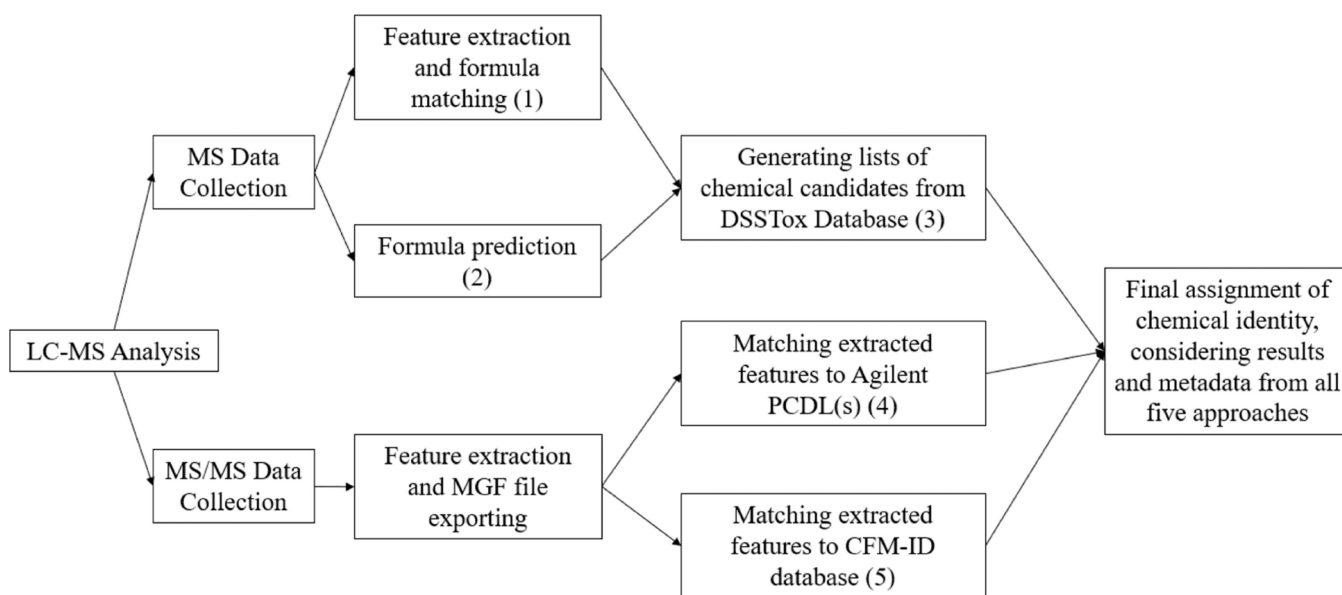
- (4). Du J; Park K; Yu X; Zhang YJ; Ye F Massive pollutants released to Galveston Bay during Hurricane Harvey: Understanding their retention and pathway using Lagrangian numerical simulations. *Sci. Total Environ* 2020, 704, 135364.
- (5). Government, U. S., 42 U.S.C. § 9604, United States Code, 2006 Edition, Supplement 4, Title 42–The public health and welfare. States, U., Ed., 2011.
- (6). Agency, U. S. E. P. OSC Warrant Officer Training: OSC Toolbox Guide; Center, C. E., Ed., 2015.
- (7). Agency, F. E. M. Emergency Support Function #10–Oil and Hazardous Materials Response Annex, 2016.
- (8). Analysis, B. a. P. f. N.-t. Glossary <https://www.nontargetedanalysis.org/glossary/> .
- (9). Overdahl KE; Sutton R; Sun J; DeStefano NJ; Getzinger GJ; Ferguson PL Assessment of emerging polar organic pollutants linked to contaminant pathways within an urban estuary using non-targeted analysis. *Environmental Science: Processes & Impacts* 2021, 23, 429–445. 3083 [PubMed: 33656498]
- (10). Pszczol ska K; Perkons I; Bartkevics V; Drzewiecki S; Plonka J; Shakeel N; Barchanska H Targeted and non-targeted analysis for the investigation of pesticides influence on wheat cultivated under field conditions. *Environ. Pollut* 2023, 316, 120468.
- (11). Fraser K; Lane GA; Otter DE; Harrison SJ; Quek SY; Hemar Y; Rasmussen S Non-targeted analysis by LC-MS of major metabolite changes during the oolong tea manufacturing in New Zealand. *Food Chem.* 2014, 151, 394–403. [PubMed: 24423549]
- (12). Jacob P; Wang R; Ching C; Helbling DE Evaluation, optimization, and application of three independent suspect screening workflows for the characterization of PFASs in water. *Environmental Science: Processes & Impacts* 2021, 23, 1554–1565. [PubMed: 34550138]
- (13). Rager JE; Strynar MJ; Liang S; McMahan 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, 88, 269–280. [PubMed: 26812473]
- (14). Newton S; McMahan R; Stoeckel JA; Chislock M; Lindstrom A; Strynar M Novel Polyfluorinated Compounds Identified Using High Resolution Mass Spectrometry Downstream of Manufacturing Facilities near Decatur, Alabama. *Environ. Sci. Technol* 2017, 51, 1544–1552. [PubMed: 28084732]
- (15). Newton SR; McMahan 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, 234, 297–306. [PubMed: 29182974]
- (16). 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, 412, 4221–4233. [PubMed: 32335688]
- (17). Lynch KL Toxicology: liquid chromatography mass spectrometry. In *Mass Spectrometry for the Clinical Laboratory*; Nair H, Clarke W, Eds.; Academic Press: San Diego, 2017; pp 109–130, Chapter 6.
- (18). Valdiviezo A; Aly NA; Luo Y-S; Cordova A; Casillas G; Foster M; Baker ES; Rusyn I Analysis of per- and polyfluoroalkyl substances in Houston Ship Channel and Galveston Bay following a large-scale industrial fire using ion-mobility-spectrometry-mass spectrometry. *J. Environ. Sci* 2022, 115, 350–362.
- (19). Roman-Hubers AT; McDonald TJ; Baker ES; Chiu WA; Rusyn I A Comparative Analysis of Analytical Techniques for Rapid Oil Spill Identification. *Environ. Toxicol. Chem* 2021, 40, 1034–1049. [PubMed: 33315271]
- (20). Luo Y-S; Aly NA; McCord J; Strynar MJ; Chiu WA; Dodds JN; Baker ES; Rusyn I Rapid Characterization of Emerging Per- and Polyfluoroalkyl Substances in Aqueous Film-Forming Foams Using Ion Mobility Spectrometry–Mass Spectrometry. *Environ. Sci. Technol* 2020, 54, 15024–15034. [PubMed: 33176098]
- (21). Schymanski EL; Jeon J; Gulde R; Fenner K; Ruff M; Singer HP; Hollender J Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol* 2014, 48, 2097–2098. [PubMed: 24476540]

- (22). Minucci JS; Deron; Purucker T; Boyce M; Grulke CM EPA NTA WebApp. [https://github.com/quanted/nta\\_app](https://github.com/quanted/nta_app).
- (23). Vegosen L; Martin TM An automated framework for compiling and integrating chemical hazard data. *Clean Technologies and Environmental Policy* 2020, 22, 441–458. [PubMed: 33867908]
- (24). Peter KT; Phillips AL; Knolhoff AM; Gardinali PR; Manzano CA; Miller KE; Pristner M; Sabourin L; Sumarah MW; Warth B; Sobus JR Nontargeted Analysis Study Reporting Tool: A Framework to Improve Research Transparency and Reproducibility. *Analytical Chemistry* 2021, 93, 13870–13879. [PubMed: 34618419]
- (25). BP4NTA NTA Study Reporting Tool (PDF). [https://figshare.com/articles/online\\_resource/NTA\\_Study\\_Reporting\\_Tool\\_PDF\\_/19763482](https://figshare.com/articles/online_resource/NTA_Study_Reporting_Tool_PDF_/19763482).
- (26). Grulke CM; Williams AJ; Thillanadarajah I; Richard AM EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. *Comput. Toxicol* 2019, 12, 100096.
- (27). Ruyle BJ; Thackray CP; McCord JP; Strynar MJ; Mauge-Lewis KA; Fenton SE; Sunderland EM Reconstructing the Composition of Per- and Polyfluoroalkyl Substances in Contemporary Aqueous Film-Forming Foams. *Environ. Sci. Technol Lett* 2021, 8, 59–65. [PubMed: 33628855]
- (28). Wolfe K; Pope N; Parmar R; Galvin M; Stevens C; Weber EJ; Flaishans J; Purucker T Chemical Transformation System: Cloud Based Cheminformatic Services to Support Integrated Environmental Modeling, 2016.
- (29). McCord JP; Groff LC; Sobus JR Quantitative non-targeted analysis: Bridging the gap between contaminant discovery and risk characterization. *Environ. Int* 2022, 158, 107011. 3084
- (30). Groff LC; Grossman JN; Krueve A; Minucci JM; Lowe CN; McCord JP; Kapraun DF; Phillips KA; Purucker ST; Chao A; Ring CL; Williams AJ; Sobus JR Uncertainty estimation strategies for quantitative non-targeted analysis. *Anal. Bioanal. Chem* 2022, 414, 4919–4933. [PubMed: 35699740]

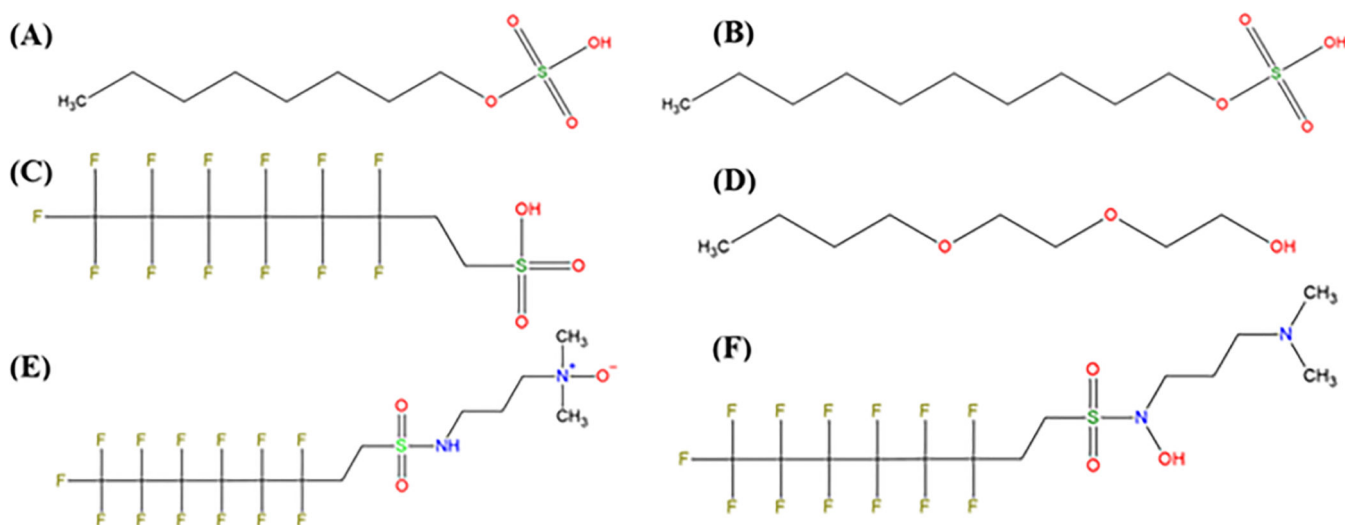




**Figure 1.** Structures of chemicals involved in mock scenarios 1 and 2: Novichok A-234 (A), malathion (B), alprazolam (C),  $\alpha$ -hydroxy alprazolam (D), fentanyl (E), and finasteride (F).



**Figure 2.** Schematic of the overall data processing workflow using MS and MS/MS data (MGF, “mascot generic format”; PCDL, “personal compound database and library”; CFM-ID, “competitive fragmentation modeling for metabolite identification”). Each of the five data processing approaches is labeled (1–5).



**Figure 3.**

Four chemicals identified at level 2 and two candidates at level 3 in mock scenario 3.

Chemicals shown in (A–C) were identified *via* ESI-results, and chemical (D) was identified *via* ESI+ results. Chemicals (E,F) were candidates of chemical identity for an ESI+ and ESI-feature that corresponded to the same unique chemical.

Chemicals: 5 Toxicity: VH - Very High H - High M - Medium L - Low I - Inconclusive N/A - Not Applicable Authority: **Authoritative** Screening QSAR Model

CAS Name	Human Health Effects										Ecotoxicity	
	Acute Mammalian Toxicity			Genotoxicity Mutagenicity	Neurotoxicity		Systemic Toxicity		Skin Sensitization	Skin Irritation	Eye Irritation	Acute Aquatic Toxicity
	Oral	Inhalation	Dermal		Single Exposure	Single Exposure	Single Exposure	Single Exposure				
<a href="#">110-11-2</a> Octyl hydrogen s...	M			VH						H		M
<a href="#">142-98-3</a> Decyl hydrogen s...	M			H						H		H
<a href="#">112-34-5</a> <small>GBTP</small> 2-(2-Butoxyethox...	M	I	L	L				I	M	H		L
<a href="#">80475-32-7</a> <small>GBT</small> N,N-Dimethyl-3-(...	I			I								I
<a href="#">29765-95-5</a> 3,3,4,4,5,5,6,6,7,... <span style="color: green;">LIKELY</span>	M			I								L

**Figure 4.** Hazard report generated *via* the HCM for mock scenario 3. Chemical structural assignments and 1 generation of breakdown products for those listed are shown, based on the “emergency response” hazard assessment profile.