Accepted Manuscript

Rapid screening and identification of chemical hazards in surface and drinking water using high resolution mass spectrometry and a case-control filter

Sarit L. Kaserzon, Amy L. Heffernan, Kristie Thompson, Jochen F. Mueller, Maria Jose Gomez Ramos

PII: S0045-6535(17)30775-0

DOI: 10.1016/j.chemosphere.2017.05.071

Reference: CHEM 19286

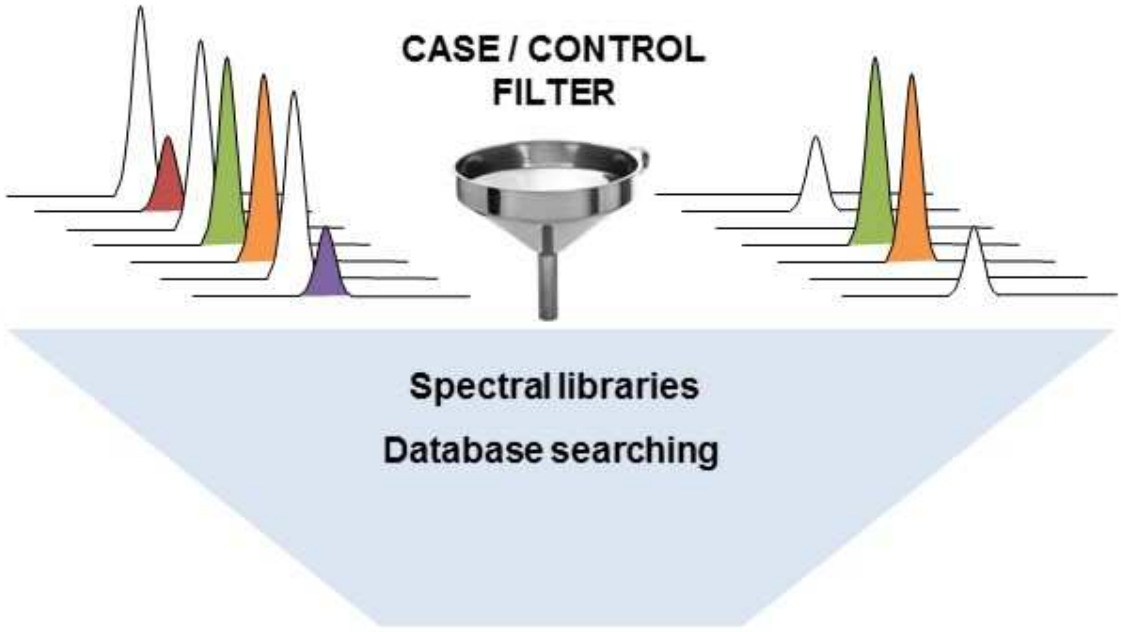
To appear in: *ECSN*

Received Date: 28 February 2017

Revised Date: 8 May 2017

Accepted Date: 11 May 2017

Please cite this article as: Kaserzon, S.L., Heffernan, A.L., Thompson, K., Mueller, J.F., Gomez Ramos, M.J., Rapid screening and identification of chemical hazards in surface and drinking water using high resolution mass spectrometry and a case-control filter, *Chemosphere* (2017), doi: 10.1016/j.chemosphere.2017.05.071.



Spectral libraries

Database searching

Rapid screening and identification of chemical hazards in surface and drinking water using high resolution mass spectrometry and a case-control filter

Sarit L. Kaserzon[a], Amy L. Heffernan[a,b], Kristie Thompson[a], Jochen F. Mueller[a] and Maria Jose Gomez Ramos[ac]*

[a]Queensland Alliance for Environmental Health Science (QAEHS), The University of Queensland, 39 Kessels Road, Coopers Plains, QLD 4108, Australia

[b]The University of Melbourne, Florey Institute of Neuroscience and Mental Health, 30 Royal Parade, Parkville, VIC 3058 Australia

[c]Agrifood Campus of International Excellence (CeiA3), Department of Chemistry and Physics, University of Almeria, European Union Reference Laboratory for Pesticide Residues in Fruit and Vegetables, Almería, Spain

[*]Corresponding author: +61 (0)415671348, k.sarit@uq.edu.au

ABSTRACT

Access to clean, safe drinking water poses a serious challenge to regulators, and requires analytical strategies capable of rapid screening and identification of potentially hazardous chemicals, specifically in situations when threats to water quality or security require rapid investigations and potential response. This study describes a fast and efficient chemical hazard screening strategy for characterising trace levels of polar organic contaminants in water matrices, based on liquid chromatography high resolution mass spectrometry with post-acquisition 'case-control' data processing. This method allowed for a rapid response time of less than 24 hours for the screening of target, suspect and non-target unknown chemicals via direct injection analysis, and a second, more sensitive analysis option requiring sample pre-concentration. The method was validated by fortifying samples with a range of pesticides, pharmaceuticals and personal care products (n=46); with >90% of target compounds positively screened in samples at 1 ng mL^{-1}, and 46% at 0.1 ng mL^{-1} when analysed via direct injection. To simulate a contamination event samples were fortified with compounds not present in the commercial library (designated 'non-target compounds'; fipronil and fenitrothion), tentatively identified at 0.2 and 1 ng mL^{-1}, respectively; and a compound not included in any known commercial library or public database (designated 'unknown' compounds; $8Cl^-$ perfluorooctanesulfonic acid), at 0.8 ng mL^{-1}. The method was applied to two 'real-case' scenarios: (1) the assessment of drinking water safety during a high-profile event in Brisbane, Australia; and (2) to screen treated, re-circulated drinking water and pre-treated (raw) water. The validated workflow

was effective for rapid prioritisation and screening of suspect and non-target potential hazards at trace levels, and could be applied to a wide range of matrices and investigations where comparison of organic contaminants between an affected and control site and or timeframe is warranted.

Key words**:** suspect screening, water monitoring, LC-QTOF, data reduction strategy, hazard identification

1. INTRODUCTION

The World Health Organization attributed an estimated 4.9 million deaths to management of, and exposure to, known chemicals in 2004[1]. Due to the large number of new chemicals registered every year, and the relatively small proportion of which are thoroughly tested, the potential risk to biota and human health is largely unknown[2]. Sources of hazardous chemicals include chemical manufacturers, service stations, hazardous materials waste sites, and common household products Environment[3]. The relatively uncharacterised nature of hazardous chemicals poses a serious challenge for regulators in charge of safeguarding human health and environmental wellbeing.

Historically, analytical methods used for aquatic monitoring typically cover only a small fraction of known, target chemicals. This approach is limited in situations where an issue of concern is identified, such as deliberate or accidental chemical spills, or extreme weather events (e.g. floods, heavy rain or droughts that can generate contaminant concentration pulses of ecotoxicological relevance to the aquatic environment) but the link to a specific chemical hazard is unclear[4]. Recently advances in high resolution mass spectrometry (HRMS) and data processing software has seen a rise in non-target analytical strategies[5-13], and particularly suspect screening[14-16], to address the need for analysis of an increasing number of analytes in complex mixtures. 'Non-target analysis' refers to detection and tentative identification of analytes for which chemical reference standards are unavailable. 'Suspect screening' is a form of non-target analysis whereby analytes are identified on the basis of accurate mass, elemental composition and structure prediction, followed by database or library searching. 'Unknown' non-target analysis is an unbiased approach, and is usually performed after targeted and suspect screening. It involves different data filtering strategies to reduce the size of the search space, followed by assignment of probable chemical formula based on MS/MS fragmentation and other strategies[17]. Non-target analyses have been used to investigate contaminants in waste[6, 14] and surface waters[10, 11, 16, 18, 19], foodstuffs[20-22], and forensic applications[11, 23, 24], but have not yet been applied in response to time-critical environmental hazard assessment.

There is a need for analytical strategies capable of rapid non-target and suspect screening for identification of hazardous chemicals, specifically in situations where exposure is unknown or involves complex chemical mixtures, and requires an immediate response. Data reduction strategies based on comparison of 'case' samples (which have an outcome of interest or concern), and 'control' samples (which do not have the observed outcome/concern) can be used to rapidly analyse the large amount of data generated during screening experiments using HRMS. The case-control approach has

been successfully used in proteomics and metabolomics studies[25, 26], but currently has limited use in environmental monitoring applications, including water quality testing. Briefly, a peak-finding algorithm is used to identify molecular features across different samples, followed by case-control comparison to identify suspect features for subsequent identification by searching against available spectral libraries, and to eliminate any matrix-specific interferences. The combination of accurate mass data and statistical evaluation of sample constituents allow for the rapid extraction and prioritisation of the most important chemical suspects for further identification.

Here we present a new approach for the rapid identification of unknown polar chemical hazards in water, based on the post-acquisition comparison of samples in a case–control setting. High resolution quadrupole time-of-flight tandem mass spectrometry (QTOF-MS/MS) is used together with "smart" data-mining software to (1) develop a rapid case-control screening method to identify the presence of potential hazardous chemicals; (2) validate the method using fortified water samples and simulate a contamination event; and (3) apply the screening strategy to raw and drinking water samples in two independent 'real-case' scenarios.

2. EXPERIMENTAL METHODS

2.1 Chemicals and standards

A standard working solution of 46 model compounds was prepared in methanol at 1 mg/L concentration (Table S1). A surrogate standard containing 12 labelled compounds at 1 mg/L was added prior to sample extraction, and used to monitor method performance; an injection standard of acetylsulfamethoxazole-d4 at 10 ng/mL was added prior to injection and used to monitor instrument performance. Calibration standard solutions were prepared in 20% methanol. All reagents and standards were high purity analytical grade (refer to Supplementary Material).

2.2 Sample preparation

Model chemicals were fortified in 1 L drinking water. Target chemicals (Table S1) were fortified at concentrations of 10, 5, 1, 0.5 and 0.1 ng mL^{-1}. Non-target chemicals (Table 1) were fortified in samples at levels of 10, 1 and 0.2 ng mL^{-1}, with the exception of 8 Cl^- PFOS, which was fortified at 40, 4 and 0.8 ng mL^{-1}. All samples underwent two treatments: (1) a 1 mL aliquot was sampled, filtered, and analysed immediately via direct injection (i.e. with no sample preconcentration); and (2) 500 mL was pre-concentrated via solid phase extraction (SPE) using 6 cc Oasis HLB cartridges (Waters) to increase sensitivity. All samples were filtered post-extraction using a 0.22 µm PTFE syringe filter (Phenomenex) and transferred to 1.5 mL glass vials prior to analysis by LC-MS/MS. A procedural

blank, instrumental blanks and calibration curves were included with each batch of samples for quality assurance/control purposes. Quantification of target compounds was performed using labelled standards. For further details refer to Supplementary Material.

2.3 LC/MS-MS parameters

Samples were analysed with a Shimadzu Nexera X2 UHPLC system equipped with a binary pump and reverse-phase XDB-C18 analytical column, coupled to a 5600 TripleTOF mass spectrometer (SCIEX, Melbourne, Australia) and equipped with electrospray ionization interface working in positive and negative ionization modes. In negative mode chromatographic separation was achieved using a reverse-phase XDB-C18 analytical column (4.6×50 mm, 1.8µm; Agilent Technologies, Santa Clara, CA) maintained at 45°C, with a flow rate of 0.6 mL min^{-1}. Mobile phases consisted of: (A) methanol:water (99:1, v/v); and (B) methanol:water (90:10, v/v); with 5mM ammonium acetate in both phases, with a gradient ramp as follows: 0min, 10%B; 0.2min, 10%B; 6.50min, 100%B; 9.50min 100%B; 9.6min, 10%B followed by equilibration at initial conditions for 2.20min. In positive mode, separation was achieved using a XDB-C18 column (2.1×100 mm, 1.8 µm; Agilent Technologies, Santa Clara, CA) maintained at 50°C, with a flow rate of 0.4 mL min^{-1}. Different chromatographic columns were used for positive and negative ionization modes that were selected following analytical column validation tests (including injection volume tests). For all the compounds investigated chromatographic performance including sensitivity and mass resolution for all analytes were very satisfactory in both ionization modes and with both analytical columns. Mobile phases consisted of: (A) 100% ultrapure water; (B) methanol, with 0.1% formic acid in both phases; with a gradient ramp as follows: 0min, 5%B; 0.2min, 5%B; 10.2min, 100%B; 14.7min 100%B; 14.9min, 10%B followed by equilibration at initial conditions for 2.20min. Injection volume was 10 µL and 5 µL in negative and positive mode, respectively. The ion source parameters were optimized as follows: source voltage, -4500 V and 5500 V for negative and positive ionization, respectively; temperature, 600ºC; curtain gas, 35 L min^{-1}; and ion source gas at 70 psi. High purity nitrogen was used as the nebulizer gas, curtain gas and collision gas.

MS and MS/MS data was acquired in high-sensitivity mode with both data-dependent (information dependent acquisition, IDA) and data independent (Sequential Window Acquisition of all Theoretical fragment-ion spectra, SWATH) modes. Data was processed with PeakView®, MS Library and MultiQuant software (SCIEX). Confirmation of target analytes was based on retention time (±0.5min), accurate mass (mass error <5ppm; mass error score >80%), isotopic distribution (isotope score >60%) and automatic MS/MS library searching (library score >70%). SWATH data was

processed both with and without product fragment ions to inform the search sequence. Extracted ion chromatogram (XIC) parameters were set to >300, corresponding to a signal-to-noise ratio of five, and XIC width was set to 0.01 Da. Confirmation of compounds not included *a priori* in the analytical method (hereon in referred to as 'non-target compounds') and structural characterization of compounds not found in spectral libraries (hereon in referred to as 'unknown' compounds) was based on accurate mass (mass error <5 ppm), elemental composition assignment, isotopic pattern distribution, ring and double bonds (RDB) factor and MS/MS spectrum interpretation. *In-silico* fragmentation with a mass tolerance threshold <10 ppm was used for structural elucidation of unknown analytes. See Supplementary Material for further description of instrument parameters, data processing and compound identification.

2.4 Suspect screening with case-control filter

A sample collection and data processing strategy based on case and control samples was developed. Control samples were collected prior to an *event*, and reflected historical or baseline chemical composition. Case samples were collected during an *event*, and captured chemical composition during the period of interest. Case and control samples were extracted and analysed simultaneously to minimize variability. Suspect screening was performed by comparing experimental spectra with a commercial MS/MS spectral library (SCIEX) containing ~2900 common aquatic organic micro-pollutants representing structurally diverse chemical classes. To create the 'filter', the intensity ratio of case to control was calculated for each suspect. For suspects where the peak intensity of case relative to control was ≤3, the suspect was 'filtered' out, as the difference between case and control was deemed negligible. This strategy allowed for a rapid and substantial reduction in the number of suspect masses investigated.

2.5 Method Application

2.5.1 Rapid screening of drinking water to identify potential hazards during an event

From 15-16 November 2014, meetings of the Group of Twenty (G20), which includes leaders from 20 major economies, took place in Brisbane, Australia. To test and safeguard the Brisbane city drinking water, potable water samples were analysed using a case-control approach and reported daily. Samples from a key water treatment plant were collected daily from 3-17 November 2014 and represented the 'case' samples (n=2 per day; n=30 total). Samples were collected from the same treatment plant prior to the event on 30 October 2014 to represent the 'control' sample (n=4).

2.5.2 Risk management of treated drinking water samples

To screen for chemicals of concern, pre- and post-treatment spot water samples were collected from nine water treatment facilities in the South East Queensland region in Australia, from 7-10 April 2015. Sampling sites were supernatant ponds containing treated, re-circulated water (case), and pre-treated (raw) water (control) (*n*=2 from each sampling site; n=36 total). Treated samples were processed through conventional means of coagulation, flocculation, sedimentation and filtration.

3. RESULTS AND DISCUSSION

3.1 Validation of the analytical method

The analytical method was optimised and validated using a range of 46 compounds covering different physico-chemical properties (Table S2), including 31 pesticides/herbicides and 15 pharmaceuticals and personal care products. Recovery was calculated by comparing samples (drinking water from a dam, and ultrapure water; n=5 replicates) fortified pre- and post-extraction at 1 ng mL^{-1}, and ranged from 20 to 109 %. Linearity was established from 0.1 to 50 ng mL^{-1} using least-squares regression, with R^2 >0.99 in all cases. Method quantification limits (MQLs) were determined as the concentration of the lowest standard that produced a peak signal ten times the background noise, and ranged from 0.1 to 1 ng mL^{-1}. Intra- and inter-day precision were determined from triplicate analyses of samples fortified at 5 ng mL^{-1}, and were <11 %RSD (Table S2).

3.2 Method development for identification of chemical hazards

3.2.1 Data acquisition and filtering strategy

A conceptual framework for the analysis and identification of potentially hazardous chemicals is presented in Figure 1. The strategy is divided into five components including sample preparation and analysis (described in Experimental Methods); and data acquisition, data processing and compound identification, discussed below.

Data was acquired in both IDA and SWATH acquisition modes, as described above and both methods successfully screened and identified all 46 compounds, with similar performance. However, it is important to note that for the compounds investigated via SWATH, relatively high concentrations with minimal matrix effects (i.e. high signal-to-noise ratio) produced high quality MS and MS/MS spectra, facilitating successful identification. In the case of lower concentration analytes, or more complex matrices, this may not always be the case as non-selective fragmentation may hamper identification[27]. Deconvolution of characteristic low m/z fragments typically used for identification of environmental pollutants could be improved, and a second injection with IDA may be required for tentative identification of some compounds (Figure 1). Additionally, because of the wider isolation

window in SWATH compared to IDA, more ions are transferred to the collision cell for fragmentation, but some ions may still be missed. For example, no MS/MS was recorded for the most intense peak (249.0244 m/z) of the non-target compound bisphenol S. In order to record all product ions for all precursor ions, a wide pass MS^{All} approach would be required. For both IDA and SWATH experiments the quality of mass spectra (and hence the closeness of the library match) decreased with decreasing chemical concentration - at fortification levels of 10, 0.5 and 0.1 ng mL^{-1}, positive library matches from direct injection samples were returned for 46, 42 and 29 compounds, respectively.

3.2.2 Rapid identification of target compounds

With non-specific screening methods, it is essential to incorporate parameters that give the greatest breadth and depth of coverage for compound identification, without comprising data quality. Collision energy and declustering potential were optimized as 35±15 eV and ±80V (for negative and positive ionisation modes), respectively, as these parameters gave the best results (high quality MS and MS/MS) for the greatest number of compounds. Notwithstanding, these generic parameters did not generate the best quality spectra for each individual analyte, and in some cases, it was necessary to perform a second injection, increasing the collision energy to obtain adequate fragmentation for subsequent identification.

Using a test solution of known composition, and percentage of detection fails as the endpoint, software parameters were systematically varied and the results assessed to evaluate the reasons for compound identification failures in each case. Failures were typically attributed to: low or no library match; low or no isotope score; and high mass error scores. Greater failures in library matching were observed at the lower concentration range (SI Figure S1). A full description of parameters used for compound identification is included in the Supplementary Material. It is important to note that identification of unknown chemicals can only be regarded as tentative until confirmed (retention time, accurate mass, and mass spectra) with an analytical standard, and tentative identification is subject to analyst bias and experience [5, 28].

3.3 Chemical hazard identification under simulated conditions

To test the capability of the developed approach, a contamination incident was simulated by fortifying water samples with known contaminants, (case), with unperturbed drinking water serving as the control. The level of difficulty and reliability of the identification approach increased from scenario (a) to scenario (d), Figure 1, and is described in detail below:

246

247 *3.3.1 Scenario A: Identification of target chemicals*

248 The first scenario was directed at identifying target compounds that are present in both commercial
249 MS/MS libraries and for which chemical standards are available in a typical laboratory setting (or can
250 be easily acquired). Using the strategy proposed in Figure 1, identification of 46 fortified model
251 compounds in drinking water samples (Table S1) by direct injection was assessed using optimised
252 SWATH and IDA acquisition methods. Results showed ≥90% of target chemicals were positively
253 matched against MS/MS library at 1 ng mL^{-1}, ≤86% at 0.5 ng mL^{-1} and ≤59% at 0.1 ng mL^{-1}. At 10 ng
254 mL^{-1} all target chemicals were positively library matched. When comparing SWATH data, processing
255 with and without the inclusion of fragment ions search function, data with fragment ion inclusion
256 slightly outperformed data without fragment ion inclusion (Figure S1).

257

258 *3.3.2 Scenario B: Suspect screening of chemicals*

259 The second scenario identified compounds that were present in a commercial MS/MS library, but for
260 which standards were not held by the investigating laboratory. The fungicide fipronil and insecticide
261 fenitrothion were chosen for this simulation. A positive identification and library match for fipronil
262 was generated for samples fortified at 0.2 ng mL^{-1}, and for fenitrothion at 1 ng mL^{-1}. Tentative
263 identification of fipronil was confirmed with mass error score (91.1), isotope score (96.6), combined
264 score (93.1), library score (100), and mass error (-0.9 ppm). An example for the identification of
265 fipronil with TOF-MS and TOF-MS/MS library match is provided in Figure 2. For fenitrothion at 1 ng
266 mL^{-1}, the relative intensity of case to control was less than 3, and thus the analyte would nominally
267 be excluded from further analysis. However, it is important to note that in a real incident scenario,
268 such as a chemical spill or toxic event, it is possible that contaminants would be present at much
269 higher concentrations.

270

271 *3.3.3 Scenario C: Identification of non target compounds*

272 The third scenario was aimed at identifying non target compounds, for which compound structures
273 were available via online chemical databases (e.g. ChemSpider, Pubchem), but not in commercial
274 spectral libraries. In this instance the flame-retardant tetrabromobisphenol A (TBBPA), the anti-
275 corrosive 4,4'-sulfonyldiphenol, and the plasticiser hexafluorobisphenol A (BPAF) were chosen for
276 the simulation, and fortified at 0.2 to 40 ng mL^{-1}. In this scenario an accurate mass was measured for
277 the chemicals, but no positive identification was possible because there was no spectral library
278 match available for these compounds. Therefore, further investigation was required.

279

An example of this process using bisphenol S is shown in Figure 3. Four theoretical elemental compositions were generated based on accurate mass (M-H ion, 249.02344 m/z), and were prioritised for further investigation by selecting compounds with the smallest mass error and the highest MS/MS rank. Probable chemical structures were generated for the highest priority elemental composition ($C_{12}H_{10}O_4S$) with the aid of ChemSpider. Each structure was fragmented *in silico*, and the theoretical fragmentation pattern compared with the experimentally-derived MS/MS for congruence. We note that a reference spectrum for bisphenol S is now available in MassBank (but was not at the time of analysis), and could be used in place of in silico fragmentation pattern. This step was critical as it allowed for tentative compound identification, despite no chemical standard being available. Generally, when searching for unknown compounds, elemental compositions that yield the highest number of results, and/or database entries that contain common chemical names are likely to be primary suspects for tentative identification, but does not always yield a successful result.

3.3.4 Scenario D: Identification unknown compounds

The final scenario was directed at 'unknown compounds', compounds that have not previously been described in the literature, and are not listed in any chemical database (i.e. chemical structures are not available). This may be the case for new chemical formulations, by-products or transformation products formed from parent species, for example, and is an important new area of research, as in some cases these secondary products are more toxic than the parent[29, 30]. For this scenario a newly discovered and synthesised by-product of the perfluorinated flame retardant perfluorooctanesulfonic acid (PFOS), $8Cl^-$ PFOS[31], was used for the simulation, and was detected at the lowest fortification level (0.8 pg mL^{-1}).

Elucidation of the chemical structure of this compound is shown in Figure 4. The 'Formula Finder' algorithm generated five possible elemental compositions. Evidence from the isotopic profile of MS and MS/MS reflected the natural isotopic abundance of a single chlorine (MS; 514.8988 m/z and 514.8956 m/z, with 30% relative intensity), and single sulfur atom (MS/MS; 79.9598 m/z and 82.9628 m/z, with 4% relative intensity), respectively; and a mass defect >0.9 confirmed the presence of fluorine[31, 32]. Together, this suggested $C_8HClF_{16}O_3S$ as the most likely chemical formula. Fragments at 329.9470, 279.9463 and 229.9489 m/z indicate sequential loss of $[CF_2]^-$, and ring/double bond factor of zero suggest a linear structure, as shown in Figure 4.

313 **Table 1. List of non-target and unknown compounds, their classification and properties.**

Compound name	Classification	Molecular formula	CAS #	Molecular weight	Speciation	Availability in chemical libraries
Tetrabromobisphenol A (TBBPA)	flame retardant	$C_{15}H_{12}Br_4O_2$	79-94-7	540.75	negative	present in online database (e.g.ChemSpider)
4,4'-Sulfonyldiphenol (BPS)	anticorrosive	$C_{12}H_{10}O_4S$	80-09-1	249.02	negative	present in online database (e.g.ChemSpider)
Hexafluorobisphenol A (BPAF)	polymer applications	$C_{15}H_{10}F_6O_2$	1478-61-1	335.05	negative	present in online database (e.g.ChemSpider)
Fipronil	fungicide	C12H4Cl2F6N4OS	120068-37-3	434.93	negative	present in commercial chemical library
Fipronil-sulfone*	fungicide degradation products	C12H4Cl2F6N4O2S	120068-36-2	450.93	negative	present in commercial chemical library
Fenitrothion	phosphorothioate insecticide	C9H12NO5PS	122-14-5	278.02	positive	present in commercial chemical library
Chlorinated PFOS (8 Cl-PFOS)	PFOS by-product, perfluorinated flame retardant	C8HO3F16SCl	*N/A*	514.90	negative	Unknown, not present in any library

* Fipronil-sulfone was not fortified in sample but detected as the degradation product of fipronil

314

315 **3.4 Application of the method in real case studies**

316 *3.4.1 Case 1: Drinking water during an incident event*

317 During the high profile G20 summit of world leaders held in Brisbane in November 2014, local water
318 authorities were charged with ensuring the integrity of the city's drinking water supplies. A rapid,
319 sensitive and effective water screening approach was therefore required to identify any substances
320 that could pose a threat to consumers.

321

322 Immediately prior to and during the event, drinking water samples (case samples) were collected
323 and screened daily, with results reported at the end of each day. Eight hundred and twenty four
324 spectral features were isolated in the control sample collected prior to the event. Of these,
325 doxylamin, fludroxycortide, mirtazapine, and morphine were detected and tentatively identified via
326 direct injection; and atrazine, caffeine, DEET, fludroxycortide, valacyclovir, PFOA, and PFOS were
327 identified in SPE pre-concentrated samples. On average, up to six suspect detects were observed in
328 case samples. However, these were chemicals that were already present in the 'control' sample, but
329 were flagged due to higher intensity ratios (>×3 ratio) in case samples, possibly the result of inter-
330 day variability in drinking water source and supply. We note that no suspicious compounds were
331 detected in samples over the G20 sampling period.

332

333 Observations from this sampling campaign highlight the importance of obtaining a control sample
334 with the exact matrix composition as the case. This allows for the most effective results, i.e.
335 optimum elimination of analytes of non-concern, and focusing of attention on a significantly reduced
336 number of analytes of potential concern. The combination of direct injection analysis and analysis of
337 pre-concentrated samples provided sensitive and comprehensive coverage.

338

339 *3.4.2 Case 2: Risk management of treated drinking water samples*

340 The case-control method was successfully applied in a water treatment risk management strategy
341 aimed at screening raw and treated/recirculated drinking water samples for identification of
342 potential hazards. Raw water samples represented the 'control' sample and treated/recirculated
343 water represented the 'case' samples to identify any hazards in treated water. The processing of the
344 samples was then reversed (i.e. treated water represented the 'control' samples while the raw water
345 samples represented the 'case') to evaluate water treatment removal efficiency. A small number of
346 non-target chemicals were detected in both raw and treated samples at one water treatment facility
347 (Table S3, Figure S2). These included the fungicide fipronil and its degradation product fipronil
348 sulfone; plasticizers dioctyl phthalate (DEHP) and dibutyl phthalate; and the altertoxin altenuene.

DEHP is the most widely used plasticiser, and was detected at five times higher levels in treated/recirculated water samples, indicating possible addition of this chemical to drinking water during the water treatment process.

4. CONCLUSIONS

Here we present a rapid, case-control screening approach to identify non-target and unknown chemical hazards, with novel application to water samples in different scenarios. This case-control strategy allowed us to focus our efforts on the identification of suspect compounds related to a case (e.g. incident event), by filtering spectral features by intensity ratio, thereby dramatically reducing data processing time. The methodology was successful in screening a large group of chemicals (n=46) with various physio-chemical properties in drinking water, with >90% and >60% of model compounds tentatively identified at ultra-trace levels of 0.5 and 0.1 ng mL^{-1}, respectively. The case-control screening approach was efficient and accurate in reducing the complexity of processing HRMS data to identify the pollutants of concern in pre- and post-treated drinking water, and surface waters, and has potential for application in many other situations and matrices including food contamination and biota.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Vitukawalu Matanitobua, Kelvin O'Halloran, Andrew Banks and Andrew Novic for field work and sample preparation; Chris Hodgkins for software support; and funding support from the Australian Research Council (FF120100546; LE140100129) and the Queensland Department of Health.

SUPPLEMENTARY MATERIAL

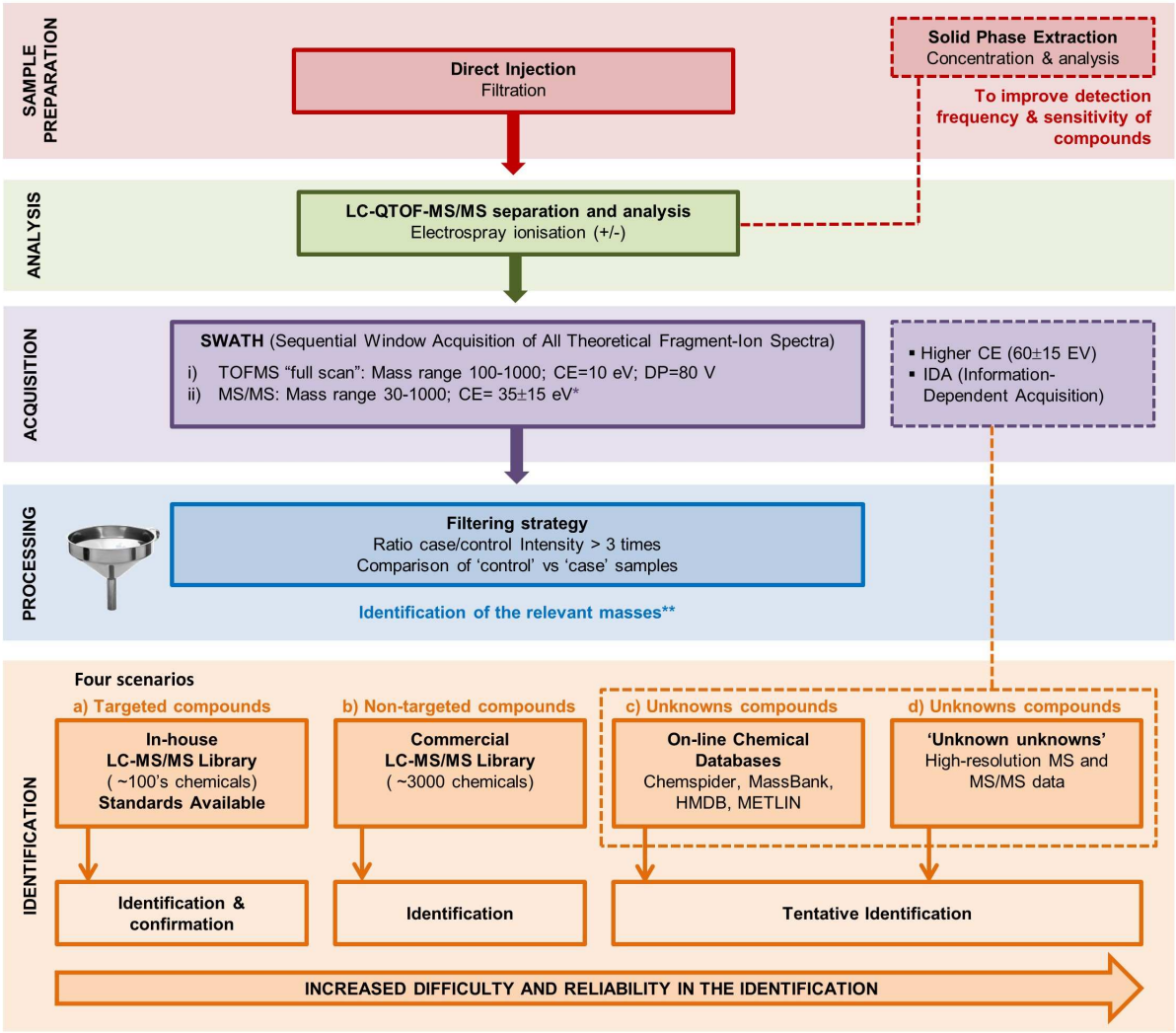
The Supplementary Material includes a detailed description of the experimental methods and compound identification, list of the 46 model compounds used and their classification and properties, validation parameters for the analytical method, and a list (and example) of the non-target and unknown chemicals detected in raw and treated/recirculated drinking water at a water treatment facility.

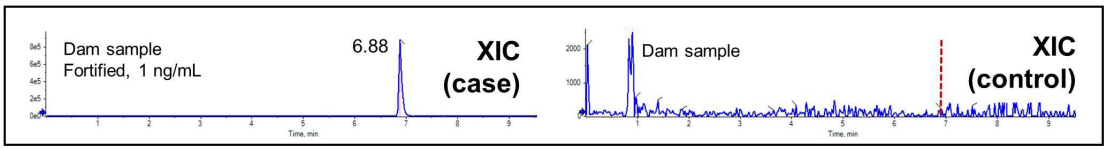
REFERENCES

1. Prüss-Ustün, A., et al., *Knowns and unknowns on burden of disease due to chemicals: a systematic review.* Environmental Health, 2011. **10**: p. 9-9.
2. Muir, D.C. and P.H. Howard, *Are there other persistent organic pollutants? A challenge for environmental chemists.* Environ Sci Technol, 2006. **40**(23): p. 7157-66.
3. European Environment Agency, E.E., *Hazardous substances in Europe's fresh and marine waters*, in *EEA Technical Report* 2011.
4. Escher, B.I., et al., *Most Oxidative Stress Response In Water Samples Comes From Unknown Chemicals: The Need For Effect-Based Water Quality Trigger Values.* Environmental Science & Technology, 2013. **47**(13): p. 7002-7011.
5. Herrera-Lopez, S., et al., *Simultaneous screening of targeted and non-targeted contaminants using an LC-QTOF-MS system and automated MS/MS library searching.* J Mass Spectrom, 2014. **49**(9): p. 878-93.
6. Gomez, M.J., et al., *Rapid automated screening, identification and quantification of organic micro-contaminants and their main transformation products in wastewater and river waters using liquid chromatography-quadrupole-time-of-flight mass spectrometry with an accurate-mass database.* J Chromatogr A, 2010. **1217**(45): p. 7038-54.
7. Gómez, M.J., et al., *Automatic Searching and Evaluation of Priority and Emerging Contaminants in Wastewater and River Water by Stir Bar Sorptive Extraction followed by Comprehensive Two-Dimensional Gas Chromatography-Time-of-Flight Mass Spectrometry.* Analytical Chemistry, 2011. **83**(7): p. 2638-2647.
8. Gomez, M.J., et al., *Spatio-temporal evaluation of organic contaminants and their transformation products along a river basin affected by urban, agricultural and industrial pollution.* Sci Total Environ, 2012. **420**: p. 134-45.
9. Krauss, M., H. Singer, and J. Hollender, *LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns* Analytical and Bioanalytical Chemistry, 2010. **397**(3): p. 943-951.
10. Ibáñez, M., et al., *Rapid non-target screening of organic pollutants in water by ultraperformance liquid chromatography coupled to time-of-light mass spectrometry.* TrAC Trends in Analytical Chemistry, 2008. **27**(5): p. 481-489.
11. Bletsou, A.A., et al., *Targeted and non-targeted liquid chromatography-mass spectrometric workflows for identification of transformation products of emerging pollutants in the aquatic environment.* TrAC Trends in Analytical Chemistry, 2015. **66** p. 32–44.
12. Schymanski, E.L., et al., *Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis.* Anal Bioanal Chem, 2015. **407**(21): p. 6237-55.
13. Baz-Lomba, J.A., M.J. Reid, and K.V. Thomas, *Target and suspect screening of psychoactive substances in sewage-based samples by UHPLC-QTOF.* Anal Chim Acta, 2016. **31**(914): p. 81-90.
14. Gago-Ferrero, P., et al., *Extended Suspect and Non-Target Strategies to Characterize Emerging Polar Organic Contaminants in Raw Wastewater with LC-HRMS/MS.* Environ. Sci. Technol., 2015. **49**(20): p. 12333-12341.
15. Pochodyloa, A.L. and D.E. Helbling, *Emerging investigators series: prioritization of suspect hits in a sensitive suspect screening workflow for comprehensive micropollutant characterization in environmental samples.* Environ. Sci.: Water Res. Technol., 2017. **3**: p. 54-65.
16. Badea, R., et al., *Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis.* Sci. Total Environ., 2015. **538**: p. 934–941.
17. Andra, S.S., et al., *Trends in the application of high-resolution mass spectrometry for human biomonitoring: An analytical primer to studying the environmental chemical space of the human exposome* Environment International 2017. **100**: p. 32-61.
18. Gomez, M.J., et al., *Photodegradation study of three dipyrone metabolites in various water systems: identification and toxicity of their photodegradation products.* Water Res, 2008. **42**(10-11): p. 2698-706.
19. Ruff, M., et al., *Quantitative target and systematic non-target analysis of polar organic micro-pollutants along the river Rhine using high-resolution mass-spectrometry – Identification of unknown sources and compounds.* Water Research, 2015. **87**: p. 145-154.

20. Gomez-Ramos, M.M., et al., *Liquid chromatography-high-resolution mass spectrometry for pesticide residue analysis in fruit and vegetables: screening and quantitative studies.* J Chromatogr A, 2013. **1287**: p. 24-37.
21. Gomez-Ramos, M.M., et al., *Screening of environmental contaminants in honey bee wax comb using gas chromatography-high-resolution time-of-flight mass spectrometry.* Environ Sci Pollut Res Int, 2016. **23**(5): p. 4609-20.
22. He, Z., et al., *Wide-scope screening and quantification of 50 pesticides in wine by liquid chromatography/quadrupole time-of-flight mass spectrometry combined with liquid chromatography/quadrupole linear ion trap mass spectrometry.* Food Chem, 2016. **196**: p. 1248-55.
23. Fels, H., et al., *Liquid chromatography-quadrupole-time-of-flight mass spectrometry screening procedure for urine samples in forensic casework compared to gas chromatography-mass spectrometry.* Drug Test Anal, 2016.
24. Montesano, C., et al., *Broad Screening and Identification of Novel Psychoactive Substances in Plasma by High-Performance Liquid Chromatography-High-Resolution Mass Spectrometry and Post-run Library Matching.* J Anal Toxicol, 2016. **40**(7): p. 519-28.
25. Zhang, A., et al., *Metabolomics in diagnosis and biomarker discovery of colorectal cancer.* Cancer Lett, 2014. **345**(1): p. 17-20.
26. Mapstone, M., et al., *Plasma phospholipids identify antecedent memory impairment in older adults.* Nat Med, 2014. **20**(4): p. 415-8.
27. Zhu, X., Y. Chen, and R. Subramanian, *Comparison of Information-Dependent Acquisition, SWATH, and MSAll Techniques in metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography–Quadrupole Time of-Flight Mass Spectrometry.* Analytical Chemistry, 2014. **86**: p. 1202-1209.
28. Hug, C., et al., *Identification of novel micropollutants in wastewater by a combination of suspect and nontarget screening.* Environ Pollut, 2014. **184**: p. 25-32.
29. Donner, E., et al., *Ecotoxicity of carbamazepine and its UV photolysis transformation products.* Sci Total Environ, 2013. **443**: p. 870-6.
30. Escher, B.I. and K. Fenner, *Recent Advances in Environmental Risk Assessment of Transformation Products.* Environmental Science & Technology, 2011. **45**(9): p. 3835-3847.
31. Rotander, A., et al., *Novel fluorinated surfactants tentatively identified in firefighters using liquid chromatography quadrupole time-of-flight tandem mass spectrometry and a case-control approach.* Environ Sci Technol, 2015. **49**(4): p. 2434-42.
32. Myers, A.L., et al., *Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products.* J Mass Spectrom, 2014. **49**(4): p. 291-6.

LIST OF FIGURES

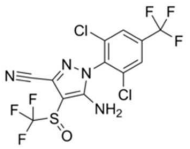

SAMPLE PREPARATION
Direct Injection
Filtration
Solid Phase Extraction
Concentration & analysis
To improve detection frequency & sensitivity of compounds
ANALYSIS
LC-QTOF-MS/MS separation and analysis
Electrospray ionisation (+/-)
ACQUISITION
SWATH (Sequential Window Acquisition of All Theoretical Fragment-Ion Spectra)
i) TOFMS "full scan": Mass range 100-1000; CE=10 eV; DP=80 V
ii) MS/MS: Mass range 30-1000; CE= 35±15 eV*
▪ Higher CE (60±15 EV)
▪ IDA (Information-Dependent Acquisition)
PROCESSING
Filtering strategy
Ratio case/control Intensity > 3 times
Comparison of 'control' vs 'case' samples
Identification of the relevant masses**
IDENTIFICATION
Four scenarios
a) Targeted compounds
b) Non-targeted compounds
c) Unknowns compounds
d) Unknowns compounds
In-house LC-MS/MS Library (~100's chemicals) Standards Available
Commercial LC-MS/MS Library (~3000 chemicals)
On-line Chemical Databases Chemspider, MassBank, HMDB, METLIN
'Unknown unknowns' High-resolution MS and MS/MS data
Identification & confirmation
Identification
Tentative Identification
INCREASED DIFFICULTY AND RELIABILITY IN THE IDENTIFICATION

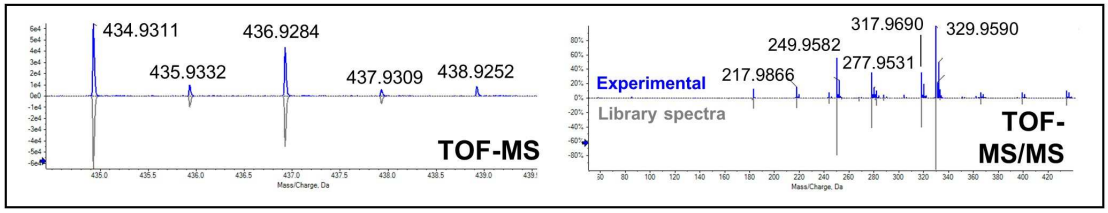


FILTERING RESULTS Ratio case/control = 3

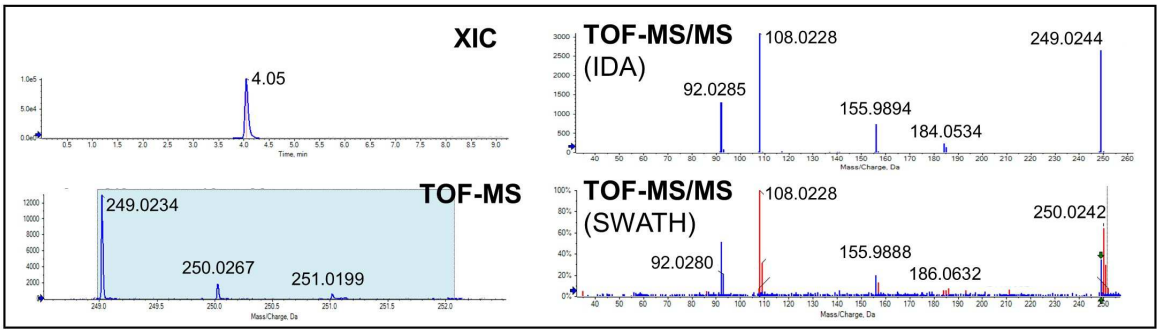
Control Intensity	Intensity	Name	Formula	Mass (Da)	Extraction Mass (Da)	Isotope	Found At Mass (Da)	Mass Error Score	Isotope Score	Combined Score	Error (ppm)	Library Score
11436	**1036814**	**Fipronil**	**C12H4Cl2F6N4OS**	**435.93872**	**434.93144**	**0**	**434.93106**	**91.1**	**96.6**	**93.1**	**-0.9**	**100**
3138	24178	Fipronil-sulfone	C12H4Cl2F6N4O2S	451.93362	450.92635	0	450.92627	98.4	83.8	81.7	-0.2	97.1
1100	114453	249.0229 / 4.04		249.02292	249.02292	0	249.02344			23	2.1	0
1022	71035	282.9833 / 4.19		282.98331	282.98331	0	282.98379			31.8	1.7	0
2479	227188	335.0509 / 6.42		335.0509	335.0509	0	335.05108			35.1	0.5	0
1573	58660	369.0110 / 6.73		369.01097	369.01097	0	369.01193			41	2.6	0
768	1036814	434.9311 / 6.88		434.93114	434.93114	0	434.93083			48.3	-0.7	0
949	714460	436.9281 / 6.88		436.92808	436.92808	0	436.92843			46.3	0.8	0
390	142121	438.9249 / 6.88		438.92495	438.92495	0	438.92529			27.1	0.8	0
825	331005	514.9000 / 6.62		514.89995	514.89995	0	514.90036			42.8	0.8	0
799	117551	516.8970 / 6.62		516.89697	516.89697	0	516.89736			39.5	0.8	0

Proposed Structure:
Fipronil





Name	Exp. RT (min)	Elemental Composition	Theoretical mass (Da)	Exp. Mass (Da)	Mass Error (ppm)	Isotope Ratio Difference (%)	Mass Error Score	Isotope Score	Library Score	Combined Score
Fipronil	6.89	C12H4Cl2F6N4OS	434.9311	434.93102	-1.0	9	90.3	87.2	96.1	90.5



ELEMENTAL COMPOSITION

Find Any | Find

Hit	Formula	m/z	RDB	ppm	MS Rank	MSMS ppm	MSMS Rank	Found
1	C7H10F4O3S	249.0214	1.0	8.0	3	7.3 (10)	1	NA/0
2	C5H10F3N2O4P	249.0258	1.0	-9.5	4	8.5 (11)	2	NA/2
3	C9H11FO5S	249.0238	4.0	-1.8	2	14.2 (11)	3	NA/1
4	C12H10O4S	249.0227	8.0	2.8	1	13.3 (10)	4	NA/45

MS Details | MSMS Details | Compound Details

Isotope cluster details — Charge: -1

Peak | Use | m/z | % Intensity | Width

Elements from:

Elements to: C40 H60 N5 O10 S2 F20 P

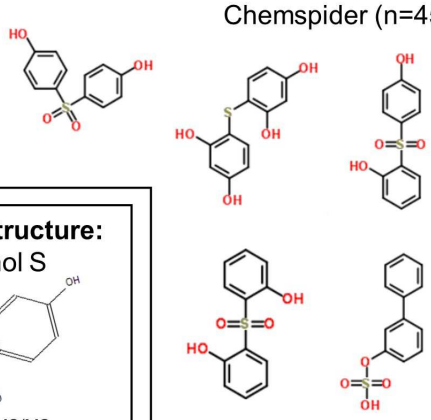
Mass tolerance (ppm): 10

Intensity tolerance (%): 10

#C/#heteroatoms greater than: 0

DATABASE MATCHES

Chemspider (n=45)



IN SILICO FRAGMENTATION

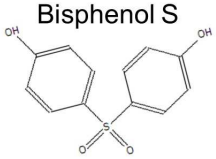
Fragments | Peaks

m/z	Num H	Broken Bonds	Bond Closure	Rad.	Error (ppm)	Composition
92.0268	0	1		✓	13.2	C6.H4.O-
93.0346	1	1			46.6	C6.H5.O-
155.9887	1	1			1.1	C6.H4.O3.S-
249.0227	0	0			6.8	C12.H9.O4.S-

Num. fragments: 4

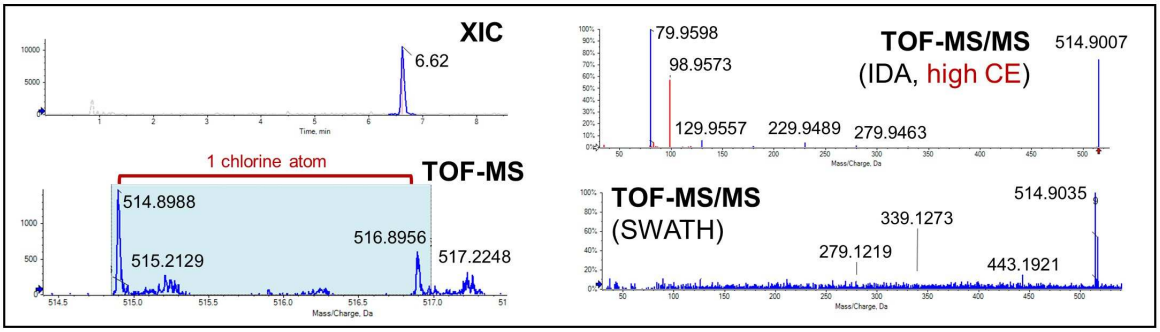
Proposed structure:

Bisphenol S



Linked to MS/MS

Theoretical mass (Da)	Exp. mass (Da)	Proposed Formula	Mass error (ppm)	RDB	Database structures found	Fragments (relative abundance)	Elemental composition (error, ppm)
249.0227	249.0234	C12H10O4S	MS: 2.8 MS/MS: 13.3	8.0	45	108.0220 (100) 249.0244 (87) 155.9887 (25) 92.0268 (42) 184.0530	C6H4O2- (2.9) [M-H]- (6.9) C6H4O3S- (4.9) C6H4O- (13.2) C12H8O2- (2.3)



ELEMENTAL COMPOSITION

Found elemental compositions

Find Any | Find

Hit	Formula	m/z	RDB	ppm	MS Rank	MSMS ppm	MSMS Rank	Found
1	C8H4ClF11N2O5S2	514.9002	2.0	-2.7	1	7.6 (19)	2	NA/0
2	C9H8ClF10O5PS2	514.9007	1.0	-3.7	2	8.3 (26)	1	NA/0
3	C8HClF16O3S	514.9007	0.0	-3.6	3	5.8 (3)	3	NA/0
4	C10H6ClF8N2O5PS2	514.8944	5.0	8.6	4	17.4 (26)	4	NA/0
5	C10H3ClF13O3PS	514.8949	3.0	7.6	5	16.9 (15)	5	NA/0

MS Details | MSMS Details | Compound Details

Isotope cluster details — Charge: -1

Elements from:

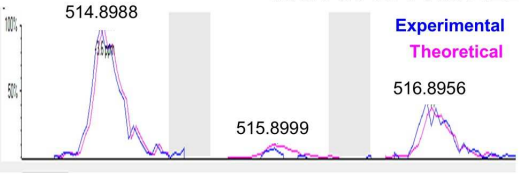
Elements to: C60 H50 N2 O5 S2 F40 P Cl

Mass tolerance (ppm): 10

Intensity tolerance (%): 10

#C/#heteroatoms greater than: 0

ISOTOPIC PROFILE



No database or library matches

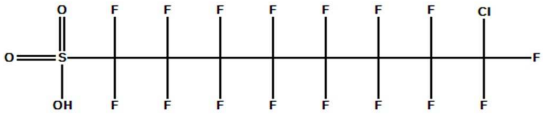
IN SILICO FRAGMENTATION

Fragments | Peaks

m/z	Num H	Broken Bonds	Bond Closure	Rad.	Error (ppm)	Composition
116.9724	2	2	☐	☐	12.6	C2.H.Cl.F3-
116.9724	2	2	☐	☐	12.6	C2.H.Cl.F3-
129.9542	0	1	☐	☑	11.7	C.F2.O3.S-
179.9510	0	1	☐	☑	15.8	C2.F4.O3.S-
229.9478	0	1	☐	☑	4.8	C3.F6.O3.S-
279.9446	0	1	☐	☑	6.3	C4.F8.O3.S-
329.9414	0	1	☐	☑	17.1	C5.F10.O3.S-

Num. fragments: 10

Proposed structure: 8Cl-PFOS



PFOS with one fluorine atom substituted by one chlorine atom

Theoretical mass (Da)	Exp. mass (Da)	Proposed Formula	Mass error (ppm)	RDB	Database structures found	Fragments (relative abundance)	Elemental composition (error, ppm)
514.9007	514.8988	C8HClF16O3S	MS: -3.6 MS/MS: 5.3	0.0	0	79.9591 (100) 514.9053 (74) 98.9565 (58) 129.9557 (6) 229.9489 (4) 279.9463 (2)	SO3- (21.1) [M-H]- (9.0) O3SF- (7.5) CF2O3S- (11.7) C3F6O3S- (4.8) C4F8O3S- (6.3)

- Non-target suspect screening of polar contaminants in water matrices
- ‘Case-control’ data processing to efficiently reduce HRMS data complexity
- Rapid, <24h response time to chemical hazard screening in real-life case studies
- >90% of target compounds (n=46) positively screened in samples at 1 µg/L