Environment International 132 (2019) 104808

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Compositional space: A guide for environmental chemists on the identification of persistent and bioaccumulative organics using mass spectrometry

Xianming Zhang[a], Robert A. Di Lorenzo[b], Paul A. Helm[a], Eric J. Reiner[a], Philip H. Howard[c], Derek C.G. Muir[d], John G. Sled[b], Karl J. Jobst[a,e,*]

[a] Ontario Ministry of the Environment, Conservation and Parks, 125 Resources Road, Toronto M9P 3V6, Canada
[b] Mouse Imaging Centre, Hospital for Sick Children, 25 Orde Street, Toronto M5T 3H7, Canada
[c] SRC, Environmental Science Center, 6502 Round Pond Road, North Syracuse, New York, United States of America
[d] Canada Centre for Inland Waters, Environment and Climate Change Canada, 867 Lakeshore Rd., Burlington, ON L7S 1A1, Canada
[e] Department of Chemistry and Chemical Biology, McMaster University, 1280 Main St. W., Hamilton L8S 4M1, Canada

ARTICLE INFO

Handling Editor: Heather Stapleton

Keywords:
Persistent organic pollutants
Nontargeted screening
Environmental mass spectrometry
Flame retardants
Chemical space

ABSTRACT

Since 2001, twenty-eight halogenated groups of persistent organic pollutants (POPs) have been banned or restricted by the Stockholm Convention. Identifying new POPs among the hundreds of thousands of anthropogenic chemicals is a major challenge that is increasingly being met by state-of-the-art mass spectrometry (MS). The first step to identification of a contaminant molecule (M) is the determination of the type and number of its constituent elements, viz. its *elemental composition*, from mass-to-charge (m/z) measurements and ratios of isotopic peaks (M + 1, M + 2 etc.). Not every combination of elements is possible. Boundaries exist in *compositional space* that divides feasible and improbable compositions as well as different chemical classes. This study explores the compositional space boundaries of persistent and bioaccumulative organics. A set of ~305,134 compounds (PubChem) was used to visualize the compositional space occupied by F, Cl, and Br compounds, as defined by m/z and isotope ratios. Persistent bioaccumulative organics, identified by *in silico* screening of 22,049 commercial chemicals, reside in more constrained regions characterized by a higher degree of halogenation. In contrast, boundaries surrounding non-halogenated chemicals could not be defined. Finally, a script tool (R code) was developed to select potential POPs from high resolution MS data. When applied to household dust (SRM 2585), this approach resulted in the discovery of previously unknown chlorofluoro flame retardants.

1. Introduction

Tens of thousands of chemical substances have been introduced to the global market (Wang, 2015; CAS Registry, 2019) while supporting the functions of modern society, concerns have been raised about some of these chemicals based on their known or suspected ecological and health impacts (Swanson et al., 1997). In order to manage the potential risks to the environment and human health, regulatory agencies have established inventories of chemicals produced, imported and used [e.g. the Canadian Domestic Substances List (DSL) (Environment and Climate Change Canada, 2017); the U.S. Toxic Substances Control Act Inventory (US TSCA) (US EPA, 2017); the European Inventory of Existing Commercial Chemical Substances (EINECS) (European Chemical Agency, 2019), and the Inventory of Existing Chemical Substances in China (IECSC) (Chemical Inspection and Regulation Service, 2019)]. Persistent organic pollutants (POPs) are an important subset of contaminants that exhibit environmental persistence (P), bioaccumulation (B), toxicity (T) and long-range transport potential (LRTP). These chemicals have been the focus of international environmental regulations. Since 2001, the Stockholm Convention has banned or restricted the use and release of twenty-eight groups of halogenated POPs (Stockholm Convention on Persistent Organic Pollutants, 2017). It is not currently practical to conduct detailed environmental monitoring and risk assessments for all halogenated organics, let alone all anthropogenic substances. Consequently, two fundamental strategies have emerged to prioritize substances for more detailed evaluation.

The *top-down* approach is characterized by the use of *in silico* screening, *i.e.* computer modelling, to identify and prioritize chemicals on the basis of computed POP-like chemical properties (P, B, T and LRTP) (Gramatica et al., 2015; Howard and Muir, 2010; Muir and

* Corresponding author at: Ontario Ministry of the Environment, Conservation and Parks, 125 Resources Road, Toronto M9P 3V6, Canada.
E-mail address: Karl.Jobst@Ontario.ca (K.J. Jobst).

https://doi.org/10.1016/j.envint.2019.05.002
Received 15 December 2018; Received in revised form 11 April 2019; Accepted 2 May 2019
Available online 08 June 2019

Howard, 2006; Strempel et al., 2012; Reppas-Chrysovitsinos et al., 2017; Zhang et al., 2016). The environmental behaviour of a chemical substance is largely determined by K_{OW}, K_{AW}, pKa (octanol-water/air-water partition coefficients and the acid dissociation constant), and degradation half-lives, among other intrinsic properties. These may be predicted from quantitative structure activity relationships (QSARs) (US, 2012) at relatively low computational cost, thus enabling screening of thousands of chemicals. The success of this approach hinges on the fact that potential POPs occupy a well-defined region of *chemical space* defined by their intrinsic properties (Wania, 2003; Gawor and Wania, 2013).

The pioneering studies of Howard and Muir (Howard and Muir, 2010), Wania and Brown (Brown and Wania, 2008), Strempel et al. (Strempel et al., 2012) and Scheringer et al. (Scheringer et al., 2012) illustrate the power of this approach. Using *c.* 20,000 chemicals with known chemical structures from the DSL and the US TSCA inventories, Howard and Muir identified 610 priority chemicals (Howard and Muir, 2010). Of these, 62% are halogenated and 8% are siloxanes. Strempel et al. (Strempel et al., 2012) and Scheringer et al. (Scheringer et al., 2012) screened 130,000 substances from the EINECS inventory and identified 510 substances with PBT and LRTP characteristics that have not been evaluated by the Stockholm Convention. The majority of these (98%) are (mixed) halogenated compounds. Guided by these results (Reppas-Chrysovitsinos et al., 2018), a number of potential POPs have been identified in environmental and biological media (Sverko et al., 2010; D'eon et al., 2009; Nyholm et al., 2013).

There are however drawbacks to the top-down approach. Uncertainties from QSARs can result in ~25% false positive or negative results (Zhang et al., 2010). Further, to accurately link chemical hazards to environmental risk requires information on emissions and occurrence of impurities and transformation products at different stages of a contaminant's life cycle that is often unknown (Howard and Muir, 2010). For example, *in silico* screening accurately predicted the POP-like behaviour of the Dechlorane class of flame retardants, which have been identified as global contaminants (Shen et al., 2011; Hoh et al., 2006). However, several important degradation products and impurities absent from chemical inventories, such as Cl/Br analogues of Dechlorane 604, would have been overlooked had it not been for complementary analytical measurements (Jobst et al., 2013; Shen et al., 2012; Shen et al., 2014; Brazeau et al., 2018).

The *bottom-up* approach to screening is performed during the course of the analysis of environmental and biological samples, most appropriately using high resolution mass spectrometry (HRMS) (Hollender et al., 2017; Sobus et al., 2017). Recent developments in multi-dimensional chromatography coupled to HRMS have raised the possibility of detecting thousands of chemical components in a single analysis (Pena-Abaurrea et al., 2014; Schymanski et al., 2015; Fernando et al., 2014; Muscalu and Górecki, 2018). Prioritizing these components for structure elucidation, further analysis and evaluation is a challenge similar to that of prioritizing POP-like chemicals in chemical inventories.

Screening large sets of suspect chemicals may be accomplished using the time-honoured approach of spectral library searching and interrogating the experimental data for the presence of expected molecular ions, isotopic peak ratios and fragment ions (Watson and Sparkman, 2007; Schymanski et al., 2014). In contrast, identifying potential POPs without prior knowledge of their structure (e.g. from spectral libraries or chemical inventories) is considerably more difficult. This is because potential POPs must be recognized on the basis of raw mass spectral data alone. Nevertheless, such information is related to elemental composition, which in part determines the properties of chemicals and their environmental behaviour.

Mass spectrometry, hyphenated with chromatography, generates multi-dimensional data: putative contaminant molecules (M) are characterized by the mass-to-charge (m/z) and abundance (I) of (quasi) molecular ions ($M^{\cdot+}$, $M + H^+$ etc.), isotopic peaks (M + 1, M + 2 etc.) as well as dissociation and associative reaction products. Mass defect (MD), the difference between nominal mass and exact mass, measured by HRMS, represents another dimension. The first step to identification of a contaminant molecule (M) is the determination of the type and number of its constituent elements, viz. its *elemental composition*, from these data (Kim et al., 2006; Jobst, 1838). Not every combination of elements is possible. Lobodin et al. (Lobodin et al., 2012) coined the term *compositional space*, noting that a boundary line defined by molecular weight (m/z) and mass defect divides feasible and improbable compositions.

Chemical classes may also be differentiated by their positions in compositional space. Halogenated POPs for example are distinguishable from non-halogenated chemicals on the basis of mass defect and diagnostic isotope patterns (Jobst et al., 2013; Pena-Abaurrea et al., 2014; Liu et al., 2015; Cariou et al., 2016). This is in line with the view that intrinsic properties measured by mass spectrometry, i.e., m/z, isotopic patterns and MD etc., define a region in *compositional space* that is occupied by potentially bioaccumulative organics, akin to how other intrinsic properties (K_{OW}, K_{OA} etc.) enclose POP-like compounds in *chemical space* (Milman and Zhurkovich, 2017). Using these measurements, hazardous chemicals whose structures may not be known beforehand, can be prioritized for identification and analysis while discarding those components that are less likely to be of concern.

We present here a general approach which is designed to enable selective detection of unknown halogenated POPs by HRMS. Compositional spaces, defined by m/z, MD, and ratios of isotopic peaks, are visualized using a set of approximately 300,000 chemical compounds (PubChem). The compositional space boundaries of POP-like chemicals are explored by projecting the chemical space of POPs onto compositional space. Finally, a script tool (R code) was developed to filter halogenated compounds from high resolution mass spectral data collected from complex mixtures. It was applied to household dust (SRM 2585), resulting in the discovery of unknown chlorofluoro chemicals.

2. Methods

2.1. Isotope clustering

A single mass-to-charge (m/z) measurement is often not sufficient to identify a compound of interest (Kim et al., 2006), especially in the presence of interfering (matrix) compounds of a complex mixture. Consider for example two compounds listed on the Canadian DSL and US TSCA chemical inventories: 4-chloro-3,5-dimethylphenol (CAS# 88-04-0) and 4-fluoro-3-nitroaniline (CAS# 364-76-1). The former is listed as a suspected POP (Howard and Muir, 2010), whereas the latter is not. Their respective monoisotopic masses differ by only 4 ppm, but the isotope pattern readily distinguishes them (see SI Fig. S1).

Mass defect (MD) of a molecule derived from accurate mass measurements is useful to distinguish elemental compositions. Fig. 1 provides a simple example using three elemental compositions: $C_7H_{21}O_4Si_4$, C_6Cl_6, and C_5F_{12}, the spectrum of $C_7H_{21}O_4Si_4$ is characterized by a small, but positive MD (0.05) and isotopic peaks consistent with the presence of ^{29}Si and ^{30}Si (M + 1 and M + 2). In contrast, the spectrum of C_6Cl_6 is characterized by a negative MD (−0.2) and the ratios of the M + 2, M + 4, M + 6 and M + 8 peaks are consistent with the presence of six chlorine atoms. The spectrum of C_5F_{12} also displays a negative MD (−0.02), but since fluorine is monoisotopic, the spectrum is characterized by a simple isotopic pattern with a weak M + 1 peak resulting from ^{13}C.

Identifying peaks belonging to the same isotopic cluster is conceptually straightforward, even in complex mass spectra, provided the data is collected with sufficient mass resolution (R > 10,000), accuracy (c. 5 ppm) and chromatographic separation (Byer et al., 2016). (In practice, peak picking, deconvolution and alignment across multiple samples can be quite challenging if the experimental conditions drift).

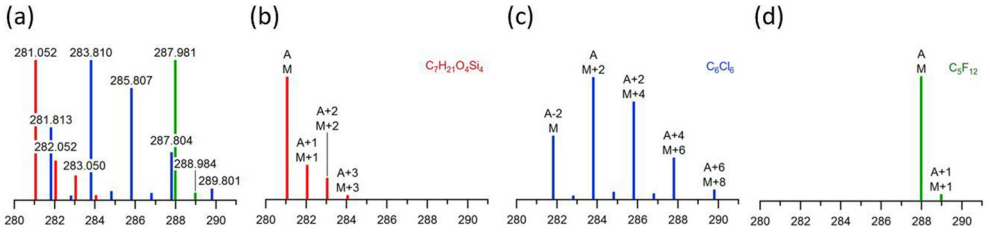

Fig. 1. (a) Isotope patterns are intrinsic to the elemental composition of a compound. (b) (Red) Si-containing and (c) (blue) Cl/Br compounds display prominent M + 2 isotope peaks, whereas (d) (Green) fluorinated compounds show relatively weak M + 1 isotope peaks. Referencing isotope peaks to the most abundant isotopic peak, A, instead of the monoisotopic peak, M, can simplify the identification of clusters of isotopic peaks. The x-axis is *m/z* and the y-axis is relative abundance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

^{13}C peaks (M + 1 and M + 2) deviate from the mass of M by 1.003 and 2.007 amu, whereas ^{34}S, ^{30}Si, ^{37}Cl and ^{81}Br isotopic peaks differ from M by 1.997 amu. Most POPs contain more than three Cl or Br atoms and, consequently, the monoisotopic peak (M) is not the most abundant peak. Significantly better detection limits are obtained by referencing the isotopic peaks relative to the most abundant isotopic peak (A), e.g. peak 'A' in Fig. 1c.

Isotopic patterns were calculated for the compounds in the chemical databases (see below) using custom Excel macros described by Kind and Fiehn (2007). Isotopic peaks in experimental data were clustered using a script tool (R code, see supplementary information) from deconvoluted mass spectra (XCMS and IntelliXtract, ACD/Labs MS Workbook Suite 2012) Briefly, the R code performs four fundamental tasks: First, in each mass spectrum, isotopic clusters are identified on the basis of isotope specific mass differences using a mass tolerance of 0.002 amu. Secondly, in each isotopic cluster, the (putative) monoisotopic peak (M) is identified as being the lowest mass peak in the cluster, whereas the intensoid peak (A) is the most intense isotopic peak. Thirdly, selected peak ratios are calculated, *viz.* (A + 1.997)/A, (A − 1.997)/A, (A + 0.999)/A and (A + 1.003). Finally the script selects Cl and Br containing compounds that meet the criteria: (A + 2):A > 25% and (A − 2):A > 30%. A more detailed description is provided in the SI.

2.2. Chemical databases

In order to visualize the compositional spaces of organic chemicals, the formulae of 305,135 compounds were retrieved from a database compiled by Kind and Fiehn (Kind and Fiehn, 2007). They represent a subset of compounds from the PubChem database that contain the elements C, H, O, N, S, P, F, Cl, Br, I or Si. Chemicals in this database may not have been produced in large volume. This set was chosen to serve as a representative sample of elemental compositions, among those of over 100 million CAS entries to date (CAS Registry, 2019). Compositional spaces were also constructed using a database of 22,043 chemicals used in commerce and industry merged from the US TSCA inventory (14,376 chemicals manufactured or imported in a quantity exceeding 25,000 pounds) (US EPA, 2017) and the Canadian DSL (11,317 compounds exceeding a quantity of 100 kg/y) (Environment and Climate Change Canada, 2017). Based on this merged database, Howard and Muir (Howard and Muir, 2010) identified 610 potential persistent bioaccumulative chemicals. The boundaries of the compositional space of POP-like chemicals were identified by projecting the 610 POP-like chemicals onto compositional space using the following dimensions, *m/z*, MD, and ratios of isotopic peaks (A, A + 1, A + 2, A − 2). The figures presented here are limited to two and three dimensions for simplicity.

2.3. Dust sample analysis

A Standard Reference Material (SRM 2585) of household dust was analyzed to demonstrate the application of compositional space filtering. The sample (50 mg) was extracted with hexane (3 × 5 mL), centrifuged and filtered with Na_2SO_4, and concentrated to a final volume of 500 µL. The extract (1 µL) was injected into an Agilent 7890B gas chromatograph (Santa Clara, CA) coupled to a Waters Xevo G2-XS quadrupole time-of-flight (QTOF) mass spectrometer (Wilmslow, UK). The mass resolution was > 20,000 FWHM (full-width half-maximum) (Megson et al., 2016). Atmospheric pressure chemical ionization (APCI), a soft ionization technique, was employed to reduce fragmentation and help preserve the molecular ions. One mechanism of APCI (positive mode) involves charge exchange between analyte molecules M and $N_2{\cdot}^+$ (and $N_4{\cdot}^+$) resulting in the formation of radical cations $M{\cdot}^+$ (McEwen, 2007; Di Lorenzo et al., 2019). Therefore, the theoretical mass and isotope ratios calculated from the chemical inventories will be directly comparable to the experimental measurements. Fast chromatographic separation (Di Lorenzo et al., 2019) was performed using a DB-5 (15 m × 0.25 mm × 0.1 µm) column. The transfer line temperature was 340 °C and the makeup gas flow (N_2) was 350 mL/min. The injector temperature was held at 280 °C. The oven temperature was held at 90 °C for 1 min, then ramped to 330 °C at 30 °C/min and held for 5 min. A constant flow (3 mL/min) of helium was used as the carrier gas. The ion source temperature was held at 150 °C. Optimum cone and auxiliary gas flows (N_2 from a Parker generator) were 100 L/h and 175 L/h respectively. Full scan mass spectra (*m/z* 50–1000) were collected at an acquisition rate of 4 Hz. Peak-picking of the HRMS data set was accomplished using XCMS and ACD/Lab MS Workbook Suite. The XCMS peak list is available in the Supporting Information.

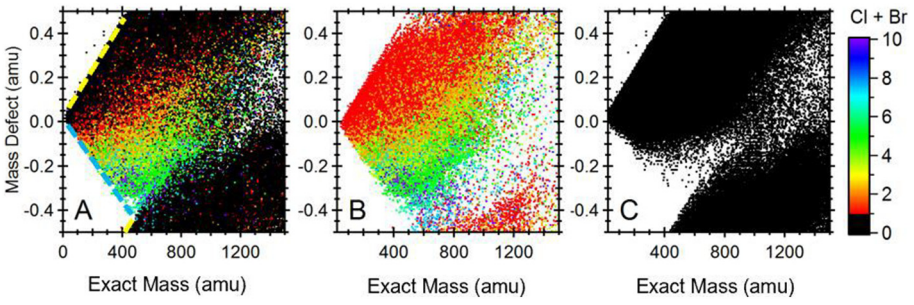

Fig. 2. Mass defect plot obtained from (A) all compounds in the Kind and Fiehn compilation of PubChem (Kind and Fiehn, 2007); (B) only chlorinated and brominated chemicals, and (C) only non-halogenated chemicals. There is significant overlap with non-halogenated compounds and Cl + Br < 3 compounds, which makes discovering halogenated species solely based on mass defect difficult. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

3. Results and discussion

3.1. Distribution of halogenated organic compounds in compositional space

Fig. 2 displays a mass defect plot (Jobst et al., 2013; Hughey et al., 2001) of the 305,135 compounds in the Kind and Fiehn (Kind and Fiehn, 2007) compilation of PubChem chemicals. Note that the standard IUPAC (International Union of Pure and Applied Chemistry) mass scale (C = 12) is used to construct Fig. 2. The mass of each composition is graphed against its corresponding (apparent) mass defect, which was calculated by subtracting the exact mass from the rounded integer mass. There is a well-defined boundary (Fig. 2 – yellow hatched line) above which, elemental compositions are absent. This is because further inclusion of elements that contribute to a positive mass defect, such as hydrogen (1.00783 Da), results in elemental compositions with no chemical significance: compounds that reside on this boundary, e.g. the alkanes, are already completely saturated. For example, the few compositions from the PubChem library that reside above the yellow hatched line (Fig. 2) are C_8H_{180}, $C_{14}H_{44}$, $C_{10}H_{40}$ and $C_{19}H_{50}$, which are not chemically feasible. Lobodin et al. (Lobodin et al., 2012) observed a similar phenomenon: the compositional space defined by carbon number and double bond equivalents (DBE) displays a boundary along which compounds are completely unsaturated.

Fig. 2 also displays a boundary (Fig. 2 – blue hatched line) that is characterized by the presence of compounds that are completely saturated with elements that contribute to a negative mass defect, such as Cl and Br. For example, perchlorinated alkanes reside on this boundary line. Between the two boundaries lie all reasonable combinations of the elements. Wrap-around occurs at *m/z* 400, when the apparent mass defect approaches the next integer mass.

The results of *in silico* screening suggest that the majority of POP-like chemicals among existing chemical inventories contain F, Cl or Br (Howard and Muir, 2010; Scheringer et al., 2012). The negative mass defect of these elements has long been exploited to reveal the presence of unknown POPs, including more recent studies that employ mass defect plots, often with non-standard mass scales (e.g. CH_2 = 14 amu, CF_2 = 50 amu, H/Cl = 34 amu) (Jobst et al., 2013; Taguchi et al., 2010; Myers et al., 2014; Ubukata et al., 2015; D'Agostino and Mabury, 2013; Barzen-Hanson et al., 2017; Liu et al., 2019). The success of this approach hinges on the assumption that most polychlorinated and polybrominated compounds (Cl and Br > 3) occupy a region of space that is devoid of other non-halogenated chemicals. A comparison of Fig. 2b and c shows that this is true to some extent, but there is still significant overlap between halogenated and non-halogenated compounds. Depending on sample complexity, it may be very difficult to identify halogenated compounds based on mass defect alone.

Fig. 3 shows the distributions of chemicals in a space defined by exact mass and the ratio of the isotopic peaks (M + 2):M. Cl/Br containing compositions are grouped in horizontal bands according to the number of Cl/Br atoms. With the exception of monochlorinated compounds, most Cl/Br compositions distribute in a region (relative intensity of M + 2 > 50%) where few non-halogenated compounds are present.

Fig. 4 displays an expanded view of the fine structure of the "non-halogenated" region shown in Fig. 3A: the compositions that contain no Cl/Br atoms (where relative intensity of M + 2 is below 30%), cluster into groups according to the number of S, and Si atoms that also contribute to the M + 2 isotopes. Kind et al. (Kind and Fiehn, 2007) observed that for a composition with a molecular weight below 400 amu, the number of sulfur and silicone atoms in a molecule can be derived based on the location of a chemical in this space (Fig. 4b and c). Above *m/z* 400, the Si and S bands converge. The rising curve is a consequence of increasing contribution from ^{13}C to the M + 2 peak. Fig. 4 was constructed assuming the ^{34}S and $^{13}C_2$ isotopic contributions to M + 2 are not resolved, although, in principle this can be achieved with a resolution of 20,000 FWHM at *m/z* 400.

Interestingly, poly−/perfluoroalkyl substances (PFASs) with F > 12 occupy a unique region below the hatched yellow line (Fig. 4b and d). Fluorine is monoisotopic and therefore it does not contribute to the M + 2 peak, which instead results from the presence of two ^{13}C atoms. Consequently, PFASs are characterized by relatively weak M + 1 and M + 2 isotopic peaks compared with other C, H, N, O organics of a similar molecular weight. Many PFASs exhibit PBT-LRTP properties and, as will be discussed below, this region of compositional space is of significant interest to environmental chemists.

The compositional space used for identification of halogenated POPs is based on the three most abundant isotopic peaks (A, A − 2, A + 2). These peaks are more convenient to be identified than the monoisotopic peak (M) and M + 2. Environmental samples are complex and it is not uncommon for peak deconvolution algorithms to reveal the presence of thousands of chemical "features", characterized by *m/z* and signal intensity measurements of the (quasi)molecular ions, fragments and isotopic peaks. The relative intensity of the M + 2 isotope is diagnostic of halogenated organic compounds (see Fig. 3). However, the correct assignment of M in an isotopic cluster may be difficult because the relative abundance of M decreases with increased halogenation: for compounds with > 10 Cl atoms or > 4 Br atoms, the relative abundance of M is < 20% of the most abundant isotope peak (A).

Fig. 5 displays the compositional space defined by monoisotopic mass and the intensity ratios (A + 2):A and (A − 2):A. Bands of Cl, Br and Cl/Br compositions are distributed in the space, with Cl_1, Cl_2 and some Cl_3 compositions located on the face, where (A + 2):A > 30% and (A − 2):A = 0. For compounds containing Br or > Cl_2, the intensity of the A − 2 peak relative to A is > 30%. The bands are labelled on the basis of Cl and Br content in Fig. 5c–e, which excludes the *m/z* dimension. In contrast, most compositions (c. 97%) are non-halogenated and occupy a densely populated region defined by (A − 2):A = 0 and (A + 2):A < 30%. (This is also the same region occupied by black markers in Fig. 3A). Poly-/perfluoroalkyl substances (highlighted in yellow, Figs. 4d and 5) are characterized by relatively low A + 2 intensities relative to *m/z*. Fig. 5b also displays several apparent outliers (in red) which correspond to compounds rich in Si (> 14 Si atoms). These potential false positives may be filtered from the Cl/Br compounds on the basis of (A + 1):A measurements (see SI).

3.2. Projecting the chemical space of persistent and bioaccumulative organics onto compositional space

Fig. 6a displays 22,043 chemicals in the North American chemical inventories (Canadian DSL and US TSCA) (Environment and Climate Change Canada, 2017; US EPA, 2017). The 610 prioritized (Howard and Muir, 2010) compounds are also displayed in Fig. 6 (in red). Of the 610 compounds, 216 contain Cl and/or Br and they all distribute in the region in which the ratios of (A + 2):A are > 30%. However, the majority of these Cl/Br compounds, 149 (69%), occupy the region where the ratios of the (A − 2):A isotopes exceed 30%, reflecting that more than two Cl or one Br atoms are present. Ninety-one of the Cl/Br containing compounds (42%) occupy the region where ratios of the (A + 2):A isotopes exceed 60% and those of the (A − 2):A isotopes exceed 50%. This region is predominantly occupied by compounds with > 4 Cl/Br atoms, including most regulated POPs.

There are four sub-regions (I–IV) in Fig. 6a in which unknown compounds are likely to be Cl/Br persistent organic pollutants. Region I, in which the relative intensities of the A + 2 and A − 2 isotopes are 65–69% and 65–69% respectively, contains compounds with five chlorine or four bromine atoms. In this region, 22 out of 25 (88%) chemicals from the DSL and TSCA inventories were identified as persistent and bioaccumulative chemicals (P & B) by Howard and Muir (2010) via *in silico* screening. The *in silico* screening identified 19 P & B out 28 (68%) in Region II where relative intensities of A + 2 and A − 2 isotopes are 65–82% and 75–93%. Compounds in this region contain 9–10 chlorine or 6–8 bromine atoms. In Region III where relative

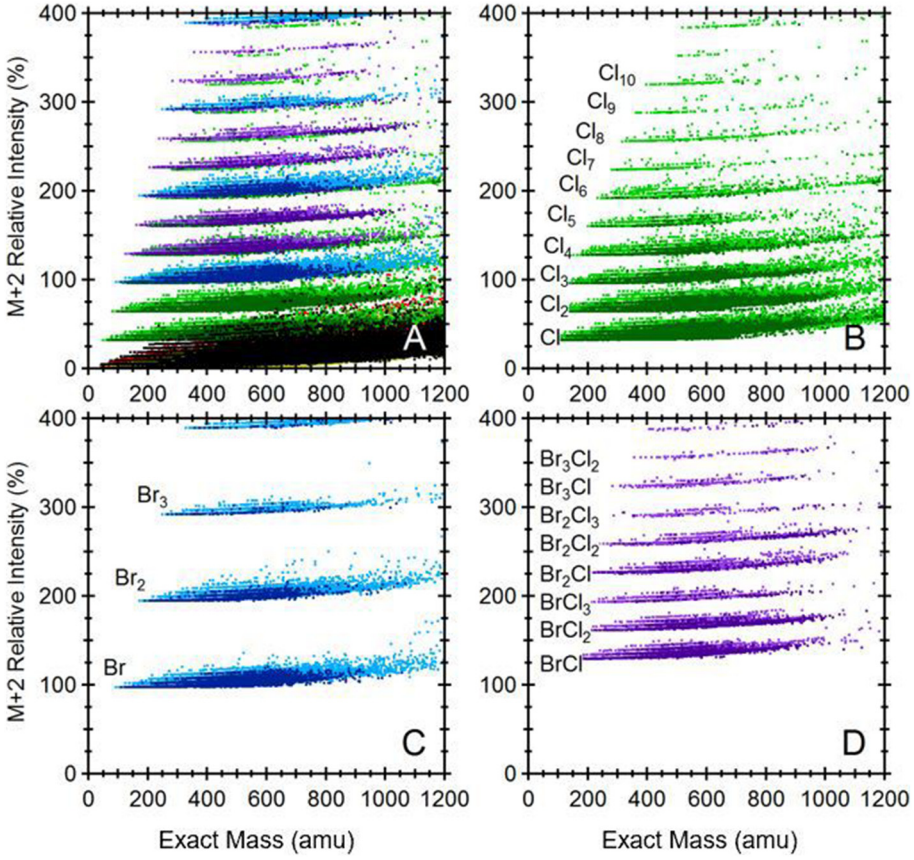

Fig. 3. Isotope ratio chemical space defined by the ratio of the M + 2 isotope to the monoisotopic peak, M, for (A) all compounds in the PubChem database, (B) only chlorinated compounds (green), (C) only brominated compounds (blue), and (D) mixed bromo-chloro compounds (purple). The red dots in (A) indicate Si-containing compounds. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

intensities of A + 2 and A − 2 isotopes are 80–83% and 50–53%, is comprised of compounds with four chlorine or three bromine atoms. Of 33 inventory chemicals in this region, 19 (58%) were identified as P & B Region IV is bound by relative intensities of A + 2 and A − 2 of 48–50% and 75–79%. Compounds in this region contain six chlorine or two bromine atoms. About 47% (27 out of 57) of the compounds from the chemical inventories are classified as P & B.

It is not surprising that PBT chemicals are characterized by a higher

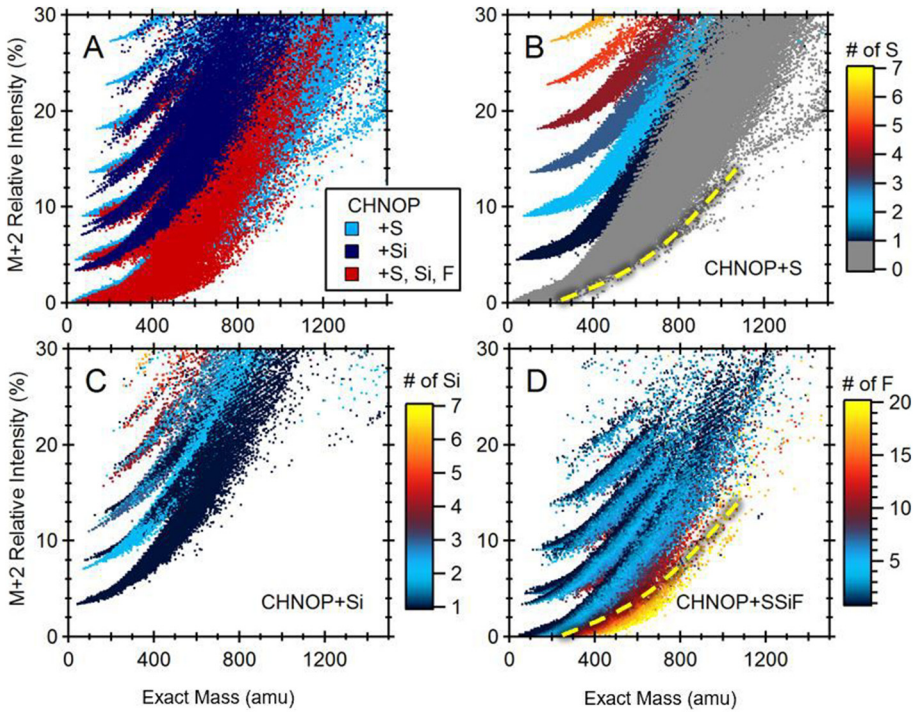

Fig. 4. Fine structure of the compositional space defined by mass and isotope ratios M + 2:M. (A) All non-chlorinated and -brominated compounds shown. (B) Compounds with S. (C) Compounds with Si. (D) Compounds with F.

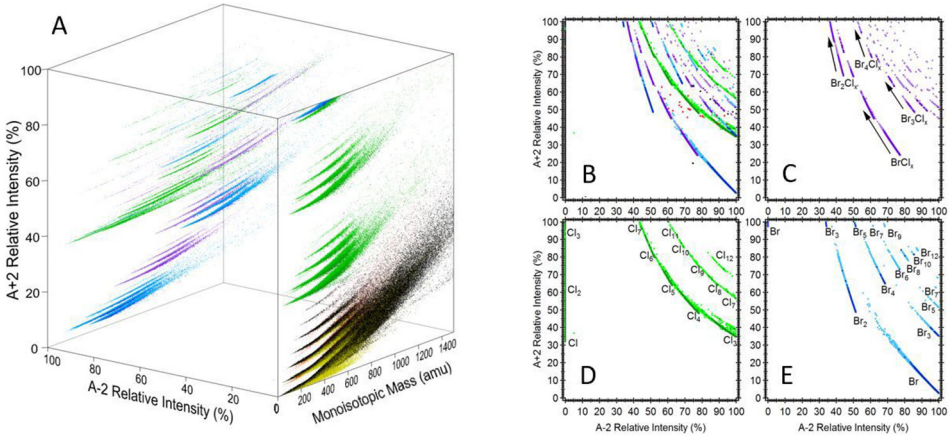

Fig. 5. (A) A composite view of the non-halogenated (black), fluorinated (yellow), chlorinated (green), brominated (blue) and mixed Br/Cl compounds (purple), defined by monoisotopic mass, and (A − 2):A and (A + 2):A ratios, among the compounds in the PubChem database. (B–E) display only the (A + 2):A vs. (A − 2):A dimensions of compositional space for the (B) Cl, Br and Cl/Br compounds; (C) mixed bromochloro compounds, (D) chlorinated compounds, and (E) brominated compounds; (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

degree of halogenation. Degree of unsaturation is also an important parameter: haloaromatics often display enhanced PBT-LRTP properties relative to saturated halocarbons (with perfluoro chemicals being an exception). One may distinguish saturated and unsaturated elemental compositions on the basis of the mass defect, which is related to the number double bond equivalents (DBE) (Hughey et al., 2001).

Fig. 6b displays only those compounds whose A − 2 and A + 2 peak intensities are 0 and < 30% relative to A. The area bound by *m/z* 400–1000 and (A + 2):A < 5% is also enriched with prioritized persistent and bioaccumulative compounds (35 out of 73 or 48%). Different from other regions, the chemicals present here are predominantly PFASs with some also containing iodine. As shown in the inset of Fig. 6b (bottom), approximately half of the compounds in this region are expected to be persistent and bioaccumulative. According to a recent review article (Liu et al., 2019), the most popular approaches to discover fluorinated compounds involve either mass defect filtering or monitoring preselected fragment ions. Unlike, Cl and Br, which are easily identified by their isotopic signatures, F is characterized by a single stable isotope. However, Fig. 6b shows that isotopic ratios (viz. $^{13}C/^{12}C$) can be used for discovery and may well be a novel approach to PFAS discovery using LC-MS (and GC–MS). On the other hand, the region bound by *m/z* 400–500 and (A + 2):A = 4–6% (top of inset of Fig. 6b) is occupied by potential POPs that are mostly non-halogenated, apart from some sulfur containing PFASs. The region is densely populated with other compounds not expected to have PBT-LRTP properties.

Chemical inventories document those produced or imported for commercial uses above certain quantities (e.g. 25,000 pounds at a single site in the US and 100 kg/y in Canada). These chemicals are more relevant to environmental chemists than most chemicals in the PubChem database. Since an environmental sample is unlikely to contain all the chemicals in the inventories, the likelihood that a compound located in one of the regions described above poses a concern may be quite high, generally 50% or higher, as indicated using the Muir and Howard criteria (Howard and Muir, 2010). Therefore, monitoring a few key intensity ratios [e.g. (A + 2):A and (A − 2):A], will help prioritize the thousands of chemical features detected by a GC- or LC-HRMS analyses. An example of this data reduction/prioritization strategy will be presented in the section below.

3.3. Screening Unknown Persistent Bioaccumulative Organics in High Resolution Mass Spectra

The approach described in the above sections is demonstrated using a sample of house dust reference material (NIST 2585) (Hilton et al., 2010). The analysis involved using the following steps: (1) First, the extract was injected into a GC-HRMS while performing data acquisition in the full-scan mode; (2) The resulting data set was deconvoluted into 18,865 *m/z* values (using XCMS), each of which being characterized by accurate mass, retention time and intensity; (3) A script tool (R code) was applied to group mass spectral peaks into isotopic clusters.

Fig. 7a displays these chemical features in a space defined by *m/z* and the ratios of isotopic peaks (A + 2):A and (A − 2):A. Cl and Br containing compounds, highlighted in orange, meet the following criteria: (A + 2):A > 30% and (A − 2):A > 30%, and are thus more likely to be possible persistent, bioaccumulative and toxic compounds.

Review of the filtered data revealed 191 isotopic clusters, including those corresponding to well-known classes of POPs, such as polychlorinated biphenyls (PCBs) and organochlorine pesticides (OCPs), including chlordane and degradation products of DDT. Polychlorinated and polybrominated diphenyl ethers (PCDEs and PBDEs) and other brominated flame retardants were also detected, including tetrabromobisphenol A (TBBPA) and bis(2-ethylhexyl)tetrabromophthalate

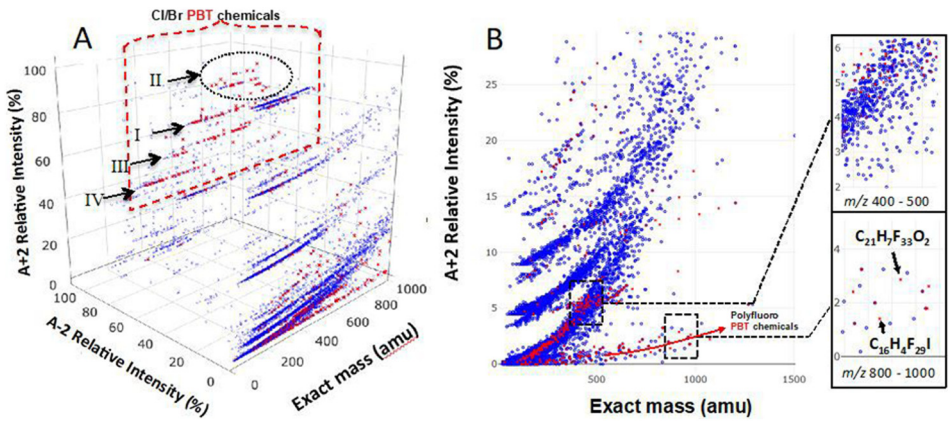

Fig. 6. Distribution of 610 prioritized persistent bioaccumulative (PB) compounds (in red) and commercial chemicals (in blue) from the North American chemical inventories in the compositional spaces defined by (A) exact mass and intensity ratios (A − 2):A and (A + 2):A; and (B) *m/z* and (A + 2):A. The insets of (B) display regions that are rich in non-halogenated PB compounds (top) and polyfluorinated compounds (bottom). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

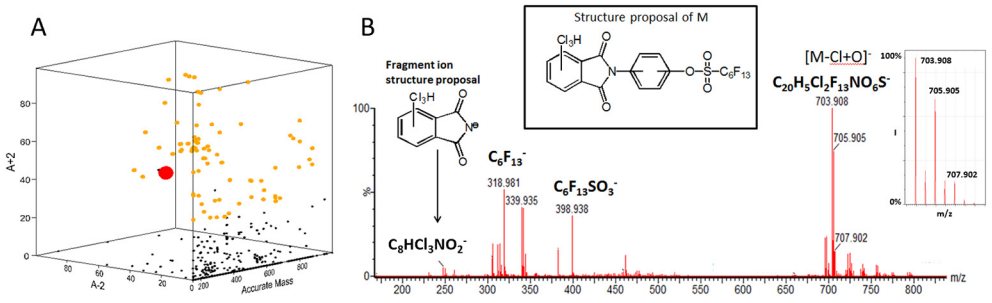

Fig. 7. (a) Approximately 18,000 chemical features detected in SRM 2585 are displayed according to *m/z*, and the intensity ratios (A + 2):A and (A − 2):A; The orange dots signify those compounds that meet the following criteria: (A + 2):A > 30% and (A − 2):A > 30%, and are thus more likely to be possible persistent, bioaccumalative and toxic compounds. The red dot signifies the compound whose MS/MS spectrum, obtained in the negative ionization mode, is shown in panel (b). An expanded view of the quasimolecular ion isotope pattern and a proposed structure of the unknown F/Cl compound are shown in the insets of B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(BEHTBP), a major component of Firemaster 550. The detection of the chlorine-containing organophosphate flame retardants tris(1,3-dichloro-2-propyl)phosphate (2 isomers), tris(2-chloroisopropyl)phosphate and tris(2-chloroethyl)phosphate is consistent with the observations reported by Hilton et al. (Hilton et al., 2010). Fentichlor, a topical fungicide was tentatively identified, but to our knowledge, has not been reported previously in SRM 2585 (Hilton et al., 2010; Ouyang et al., 2017).

Searching on the basis of accurate mass, though useful, has its limitation. Using the Environmental Protection Agency's CompTox Chemicals Dashboard, potential elemental compositions were found for approximately ~120 of the 191 isotopic clusters. However, only a subset of 24 could be assigned with a confidence level greater than four, using the scheme devised by Schymanski et al. (Hollender et al., 2017; Schymanski et al., 2015). These assignments are summarized in Table S1. We were quite interested in the remaining 71 isotopic clusters, which appeared to have no matches in the CompTox structure library. Indeed, a number of potential POPs have been discovered that are not listed in any library because of their origin as degradation and/or transformation products (Jobst et al., 2013; Shen et al., 2012). Thus, the remaining isotopic clusters correspond to true “unknowns”, of which nine attracted attention because: (i) they were obviously homologous chlorofluoro compounds, related to one another by 33.9610 amu and 99.9936 amu increments that correspond to H/Cl and C_2F_4 substitution; (ii) the abundance of these compounds is greater than those of well-known POPs, including the PBDEs. We note that the development of criteria to prioritize unknown features for identification is a topic of current interest, and such criteria may include abundance, frequency of detection, as well as temporal and spatial trends in environmental and biological samples (Plassmann et al., 2018).

The positive ion mode spectra of these compounds are characterized by intense molecular ion peaks ($M\cdot^+$), which shift by 19 mass units to form ions $[M - Cl + O]^-$ under negative ion mode conditions (Carroll et al., 1975).

The mass spectrum (negative ion mode) of one of the Cl_3 compounds (denoted as **Cl/F-1**) is displayed in Fig. 7b. The spectrum displays an intense quasimolecular ion and indicates the presence of two chlorine atoms (following O/Cl exchange ionization) as witnessed by the 9:6:1 intensity ratios of the isotopic peaks. The elemental compositions of two key in-source fragment ions, viz. $C_8HCl_3NO_2^-$ at *m/z* 247.908 (2.8 ppm) and $C_6F_{13}SO_3^-$ at *m/z* 398.938 (4.8 ppm), provide important clues on the identity of this contaminant. The $C_6F_{13}SO_3^-$ fragment is unlikely to have any structure other than that of a perfluorohexane sulfonate ion, while the $C_8HCl_3NO_2^-$ fragment ion is proposed to have the structure of the chloro phthalimide ion shown in Fig. 7b. The 92.027 Da difference between *m/z* 339.935 and *m/z* 247.908 strongly suggests that a phenoxy moiety links the phthalimide and perfluorohexane sulfonate groups as shown in the structure proposal (inset, Fig. 7b). Other compounds in the series (**Cl/F-2,3** and **-4**) appear to be related by presence of perfluoro moieties up to C_8F_{17} and between three and four Cl substituents on the phthalimide ring. Their mass spectra are displayed in Fig. S3a-c and one observes that the most intense peaks in the spectra shift according to the presence of 3–4 Cl atoms and C_6F_{13} or C_8F_{17} moieties. Indeed, all of the peaks displayed in these mass spectra correspond to direct bond cleavages of the proposed phthalimide structure shown in Fig. S4.

Similar to **Cl/F-1,2,3** and **-4**, compounds **Cl/F-5** and **-6** yield (negative ion mode) mass spectra (Fig. S5) that are dominated by $[M-Cl + O]^-$ quasimolecular ions. However, a peak corresponding to $[M–H]^-$, rather than $M^{\cdot-}$, is also present and suggests that compounds **Cl/F-5** and **Cl/F-6** have at least one ionizable proton. This is consistent with the proposal that **Cl/F-5** and **Cl/F-6** are amides rather than phthalimides, see Figs. S5 and S6. We note that Strynar *et al.* (2016) have previously reported the detection of **Cl/F-6** in SRM 2585. **Cl/F-7**, a Cl_3 congener of **Cl/F-6** was tentatively identified on the basis of its molecular ion (see Table S1), but its mass spectrum was too weak to provide further diagnostic information. The remaining two Cl/F isotopic clusters in Table were determined to originate from in-source fragmentation.

The structures of these chemicals are not registered in any databases, but we note that chlorinated and brominated phthalimides modified with perfluoro moieties have been used in the production of flame resistant milk-white polycarbonate plastics used in computer cases and other electronics since the 1980s (Cohnen et al., 1985; Avakian, 1985). We also note that the proposed structure in Fig. 7b (inset) is very similar to that of 2,3,4,5-Tetrachloro-6-((3-(tridecafluorohexyl)sulfonyloxy)phenylaminocarbonyl)benzoic acid, whose potassium salt (CAS: 68815–72-5) is currently listed in the 610 compounds prioritized by Howard and Muir (Howard and Muir, 2010). Compounds **Cl/F-1,2,3** and **-4** may well ring-close upon heating whereas compounds **Cl/F-5** and **-6** may decarboxylate under the same conditions. While such degradation may occur in a heated GC injector, Strynar *et al* (2016) have detected compound **Cl/F-6** by LC-MS, which suggests that degradation of 68815-72-5 likely occurs during molding of thermoplastic.

Collectively, these compounds represent precursors of perfluoroalkane sulfonates that have been restricted by the Stockholm Convention and the results presented above represent a first detection in (indoor) environmental media. This example may serve to demonstrate the effectiveness of our approach to reveal unknown contaminants. However, the unusually high molecular weights of these mixed halogenated phthalimides places them outside the compositional space of known and suspected POPs: the Cl_4 compounds prioritized by *in silico* screening range between 200 and 500 amu. This highlights the challenge posed by degradation and (bio)transformation of anthropogenic chemicals. Degradation of the unknown chlorofluoro chemicals may well lead to the release of perfluorooctane and perfluorohexane sulfonic acids, which are listed as POPs and a proposed addition to the list, respectively, by the Stockholm Convention.

Bottom-up screening, using HRMS in combination with compositional space filtering, may reveal the presence of chemicals that are persistent and bioaccumulative but not intentionally produced and thus

not in a chemical inventory (e.g. impurities and degradation/transformation products). To our knowledge, the development of selection criteria based on chemical inventories and the results of *in silico* screening has never been reported. Furthermore, the general approach will aid in the development of nontargeted screening methodologies focused on other classes of environmental contaminants. The approach introduced in this paper is effective in selecting any (mixed) Cl, Br compounds (as well as some fluorine and iodine containing compounds) from HRMS data obtained from any combination of ionization and separation techniques. However, non-halogenated compounds of environmental concerns, including polycyclic aromatic compounds, organophosphorous compounds and organosiloxanes, cannot be identified on the basis of *m/z*, mass defect and isotope ratios alone. Identification of these groups of compounds requires additional dimensions afforded by complementary mass spectrometric techniques (e.g. collision-induced dissociation, selective ionization and chromatography separation) (Ong and Hites, 1994; Fernando et al., 2016), which are beyond the scope of this study. Finally, not all halogenated compounds are POPs: for example, the (positional) isomer of chloronitroaniline selected using the script tool (see Fig. S7) is unlikely to bioaccumulate because of its low octanol-water partition coefficient (Kow), which reflects polar functional groups. Thus, molecular structure also plays a critical role in the environmental fate of organic molecules. Such structure information is accessible from the dissociation behaviour of the corresponding ion, usually obtained by MS/MS, and/or structure diagnostic ion-molecule reactions (Fernando et al., 2016; Jobst et al., 2009).

The emphasis of this contribution is that fundamental measurements obtained by mass spectrometry, e.g. *m/z*, mass defect and isotopic peak ratios of both parent and progeny ions (Peng et al., 2015; Peng et al., 2016; Liu et al., 2018), can be related to the structures and the potential environmental impact of anthropogenic chemicals *without* prior knowledge of their structure or occurrence. The challenge of identifying new POPs among the thousands of chemical features detected in a single HRMS analysis (the bottom-up approach) is akin to that of screening chemical inventories guided by intrinsic properties of contaminants (the top-down approach). Together, the top-down and bottom-up approaches complement each other and will guide environmental chemists to identify new POPs.

Acknowledgements

We thank Dr. Ronald A. Hites and the anonymous reviewers for helpful comments on an earlier draft of the paper. KJJ thanks Dr. Graham McGibbon for valuable discussions that inspired our exploration of compositional space. Funding for this work was provided in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant U01-HD-087177-01.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envint.2019.05.002.

References

Avakian, R. W. (1985) Stabilization of flame retardant polycarbonate-polyester compositions. US Patent: US4555540A.

Barzen-Hanson, K.A., Roberts, S.C., Choyke, S., Oetjen, K., McAlees, A., Riddell, N., McCrindle, R., Ferguson, P.L., Higgins, C.P., Field, J.A., 2017. Discovery of 40 classes of per-and polyfluoroalkyl substances in historical aqueous film-forming foams (AFFFs) and AFFF-impacted groundwater. Environ. Sci. Technol. 51, 2047–2057.

Brazeau, A.L., Pena-Abaurrea, M., Shen, L., Riddell, N., Reiner, E.J., Lough, A.J., McCrindle, R., Chittim, B., 2018. Dechlorinated analogues of dechlorane plus. Environ. Sci. Technol. 52, 5619–5624.

Brown, T.N., Wania, F., 2008. Screening chemicals for the potential to be persistent organic pollutants: a case study of Arctic contaminants. Environ. Sci. Technol. 42, 5202–5209.

Byer, J.D., Siek, K., Jobst, K.J., 2016. Distinguishing the C3 vs SH4 mass split by comprehensive two-dimensional gas chromatography-high resolution time-of-flight mass spectrometry. Anal. Chem. 88 (12), 6101–6104.

Cariou, R., Omer, E., Leon, A., Dervilly-Pinel, G., Le Bizec, B., 2016. Screening halogenated environmental contaminants in biota based on isotopic pattern and mass defect provided by high resolution mass spectrometry profiling. Anal. Chim. Acta 936, 130–138.

Carroll, D., Dzidic, I., Stillwell, R., Haegele, K., Horning, E., 1975. Atmospheric pressure ionization mass spectrometry. Corona discharge ion source for use in a liquid chromatograph-mass spectrometer-computer analytical system. Anal. Chem. 47, 2369–2373.

CAS Registry (2019). http://www.cas.org/support/documentation/chemical-substances (accessed July 2018).

Chemical Inspection and Regulation Service, 2013. Chinese Chemical Inventory of Existing Chemical Substances (IECSC). (August 2018). http://cciss.cirs-group.com/.

Cohnen, W.; Kircher, K.; Muller, P. R.; Krishnan, S.; Neuray, D. (1985) Polycarbonate molding compositions having improved flame retardance. US Patent: US4552911A.

D'Agostino, L.A., Mabury, S.A., 2013. Identification of novel fluorinated surfactants in aqueous film forming foams and commercial surfactant concentrates. Environ. Sci. Technol. 48, 121–129.

D'eon, J.C., Crozier, P.W., Furdui, V.I., Reiner, E.J., Libelo, E.L., Mabury, S.A., 2009. Observation of a commercial fluorinated material, the polyfluoroalkyl phosphoric acid diesters, in human sera, wastewater treatment plant sludge, and paper fibers. Environ. Sci. Technol. 43, 4589–4594.

Di Lorenzo, R.A., Lobodin, V.V., Cochran, J., Kolic, T., Besevic, S., Sled, J.G., Reiner, E.J., Jobst, K.J., 2019. Fast gas chromatography-atmospheric pressure (photo)ionization mass spectrometry of polybrominated diphenylether flame retardants. Anal. Chim. Acta 1056, 70–78.

Environment and Climate Change Canada, Domestic substances list. https://http://www.canada.ca/en/environment-climate-change/services/canadian-environmental-protection-act-registry/substances-list/domestic.html (accessed Jan 2017).

European Chemical Agency (ECHA), European Inventory of Existing Commercial Chemical Substances (EINECS) (2019). https://echa.europa.eu/information-on-chemicals/ec-inventory (July 2018).

Fernando, S., Jobst, K.J., Taguchi, V.Y., Helm, P.A., Reiner, E.J., McCarry, B.E., 2014. Identification of the halogenated compounds resulting from the 1997 Plastimet Inc. fire in Hamilton, Ontario, using comprehensive two-dimensional gas chromatography and (ultra) high resolution mass spectrometry. Environ. Sci. Technol. 48, 10656–10663.

Fernando, S., Green, M.K., Organtini, K., Dorman, F., Jones, R., Reiner, E.J., Jobst, K.J., 2016. Differentiation of (mixed) halogenated dibenzo-p-dioxins by negative ion atmospheric pressure chemical ionization. Anal. Chem. 88, 5205–5211.

Gawor, A., Wania, F., 2013. Using quantitative structural property relationships, chemical fate models, and the chemical partitioning space to investigate the potential for long range transport and bioaccumulation of complex halogenated chemical mixtures. Environ. Sci.: Proecess Impact 15, 1671–1684.

Gramatica, P., Cassani, S., Sangion, A., 2015. PBT assessment and prioritization by PBT index and consensus modeling: comparison of screening results from structural models. Environ. Int. 77, 25–34.

Hilton, D.C., Jones, R.S., Sjödin, A., 2010. A method for rapid, non-targeted screening for environmental contaminants in household dust. J. Chromatogr. A 1217, 6851–6856.

Hoh, E., Zhu, L., Hites, R.A., 2006. Dechlorane plus, a chlorinated flame retardant, in the Great Lakes. Environ. Sci. Technol. 40, 1184–1189.

Hollender, J., Schymanski, E.L., Singer, H.P., Ferguson, P.L., 2017. Nontarget screening with high resolution mass spectrometry in the environment: ready to go? Environ. Sci. Technol. 51, 11505–11512.

Howard, P.H., Muir, D.C., 2010. Identifying new persistent and bioaccumulative organics among chemicals in commerce. Environ. Sci. Technol. 44, 2277–2285.

Hughey, C.A., Hendrickson, C.L., Rodgers, R.P., Marshall, A.G., Qian, K., 2001. Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. Anal. Chem. 73, 4676–4681.

Jobst, C., 1838. Thein Identisch Caffein. Chemie unde Pharmaceutische Chemie Insbesondere 25, 63–66.

Jobst, K.J., De Winter, J., Flammang, R., Terlouw, J.K., Gerbaux, P., 2009. Differentiation of the pyridine radical cation from its distonic isomers by ion-molecule reactions with dioxygen. Int. J. Mass Spectrom. 286, 83–88.

Jobst, K.J., Shen, L., Reiner, E.J., Taguchi, V.Y., Helm, P.A., McCrindle, R., Backus, S., 2013. The use of mass defect plots for the identification of (novel) halogenated contaminants in the environment. Anal. Bioanal. Chem. 405, 3289–3297.

Kim, S., Rodgers, R.P., Marshall, A.G., 2006. Truly "exact" mass: Elemental composition can be determined uniquely from molecular mass measurement at ~ 0.1 mDa accuracy for molecules up to 500 Da. Int. J. Mass Spectrom. 251, 260–265.

Kind, T., Fiehn, O., 2007. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics 8, 105.

Liu, Y., Pereira, A.D.S., Martin, J.W., 2015. Discovery of C5–C17 poly-and perfluoroalkyl substances in water by in-line SPE-HPLC-Orbitrap with in-source fragmentation flagging. Anal. Chem. 87, 4260–4268.

Liu, Y., Richardson, E.S., Derocher, A.E., Lunn, N.J., Lehmler, H.-J., Li, X., Zhang, Y., Yue Cui, J., Cheng, L., Martin, J.W., 2018. Hundreds of unrecognized halogenated contaminants discovered in polar bear serum. Angew. Chem. 57, 16401–16406. https://doi.org/10.1002/anie.201809906. in press.

Liu, Y; D'Agostino, L.A.; Qu, Quangbo; Jiang, G.; Martin, J.W.; High-resolution mass spectrometry (HRMS) methods for nontarget discovery and characterization of poly-and per- fluoroalkyl substances (PFASs) in environmental and human samples. Trends in Analytical Chemistry, 2019, (in press).

Lobodin, V.V., Marshall, A.G., Hsu, C.S., 2012. Compositional space boundaries for

organic compounds. Anal. Chem. 84, 3410–3416.
McEwen, C.N., 2007. GC/MS on an LC/MS instrument using atmospheric pressure photoionization. Int. J. Mass Spectrom. 259, 57–64.
Megson, D., Robson, M., Jobst, K.J., Helm, P.A., Reiner, E.J., 2016. Determination of halogenated flame retardants using gas chromatography with atmospheric pressure chemical ionization (APCI) and a high resolution quadrupole time-of-flight mass spectrometer (HRqTOFMS). Anal. Chem. 88 (23), 11406–11411.
Milman, B.L., Zhurkovich, I.K., 2017. The chemical space for non-target analysis. TrAC Trends Anal. Chem. 97, 179–187.
Muir, D.C., Howard, P.H., 2006. Are there other persistent organic pollutants? A challenge for environmental chemists. Environ. Sci. Technol. 40, 7157–7166.
Muscalu, A.M., Górecki, T., 2018. Comprehensive Two-Dimensional Gas Chromatography in Environmental Analysis. TrAC Trends Anal. Chem. 106, 225–245.
Myers, A.L., Jobst, K.J., Mabury, S.A., Reiner, E.J., 2014. Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products. J. Mass Spectrom. 49, 291–296.
Nyholm, J.R., Grabic, R., Arp, H.P.H., Moskeland, T., Andersson, P.L., 2013. Environmental occurrence of emerging and legacy brominated flame retardants near suspected sources in Norway. Sci. Total Environ. 443, 307–314.
Ong, V.S., Hites, R.A., 1994. Electron capture mass spectrometry of organic environmental contaminants. Mass Spectrom. Rev. 13, 259–283.
Ouyang, X., Weiss, J.M., de Boer, J., Lamoree, M.H., Leonards, P.E.G., 2017. Non-target analysis of household dust and laundry dryer lint using comprehensive two-dimensional liquid chromatography coupled with time-of-flight mass spectrometry. Chemosphere 166, 431–437.
Pena-Abaurrea, M., Jobst, K.J., Ruffolo, R., Shen, L., McCrindle, R., Helm, P.A., Reiner, E.J., 2014. Identification of potential novel bioaccumulative and persistent chemicals in sediments from Ontario (Canada) using scripting approaches with GC× GC-TOF MS analysis. Environ. Sci. Technol. 48, 9591–9599.
Peng, H., Chen, C., Saunders, D.M., Sun, J., Tang, S., Codling, G., Hecker, M., Wiseman, S., Jones, P.D., Li, A., Rockne, K.J., Giesy, J.P., 2015. Untargeted identification of organo-bromine compounds in lake sediments by ultrahigh-resolution mass spectrometry with the data-independent precursor isolation and characteristic fragment method. Anal. Chem. 87, 10237–10246.
Peng, H., Saunders, D.M., Sun, J., Jones, P.D., Wong, C.K., Liu, H., Giesy, J.P., 2016. Mutagenic azo dyes, rather than flame retardants, are the predominant brominated compounds in house dust. Environ. Sci. Technol. 50, 12669–12677.
Plassmann, M.M., Fischer, S., Benskin, J.P., 2018. Nontarget time trend screening in human blood. Environmental Science and Technology Letters 5 (6), 335–340.
Reppas-Chrysovitsinos, E., Sobek, A., MacLeod, M., 2017. Screening-level exposure-based prioritization to identify potential POPs, vPvBs and planetary boundary threats among Arctic contaminants. Emerging Contaminants 3, 85–94.
Reppas-Chrysovitsinos, E., Sobek, A., MacLeod, M., 2018. In Silico Screening-Level Prioritization of 8468 Chemicals Produced in OECD Countries to Identify Potential Planetary Boundary Threats. Bull. Environ. Contam. Toxicol. 100 (1), 134–146.
Scheringer, M., Strempel, S., Hukari, S., Ng, C.A., Blepp, M., Hungerbuhler, K., 2012. How many persistent organic pollutants should we expect? Atmospheric Pollution Research 3, 383–391.
Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying Small Molecules Via High Resolution Mass Spectrometry: Communicating Confidence. ACS Publications.
Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., 2015. Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal. Bioanal. Chem. 407, 6237–6255.
Shen, L., Reiner, E.J., Helm, P.A., Marvin, C.H., Hill, B., Zhang, X., MacPherson, K.A., Kolic, T.M., Tomy, G.T., Brindle, I.D., 2011. Historic trends of Dechloranes 602, 603, 604, Dechlorane Plus and other norbornene derivatives and their bioaccumulation potential in Lake Ontario. Environ. Sci. Technol. 45, 3333–3340.
Shen, L., Jobst, K.J., Helm, P.A., Reiner, E.J., McCrindle, R., Tomy, G.T., Backus, S., Brindle, I.D., Marvin, C.H., 2012. Identification and determination of the dechlorination products of Dechlorane 602 in Great Lakes fish and Arctic beluga whales by gas chromatography-high resolution mass spectrometry. Anal. Bioanal. Chem. 404, 2737–2748.
Shen, L., Jobst, K.J., Reiner, E.J., Helm, P.A., McCrindle, R., Taguchi, V.Y., Marvin, C.H., Backus, S., MacPherson, K.A., Brindle, I.D., 2014. Identification and occurrence of analogues of dechlorane 604 in Lake Ontario sediment and their accumulation in fish. Environ. Sci. Technol. 48, 11170–11177.
Sobus, J.R., Wambaugh, J.F., Isaacs, K.K., Williams, A.J., McEachran, A.D., Richard, A.M., Grulke, C.M., Ulrich, E.M., Rager, J.E., Strynar, M.J., 2017. Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. J. Expo. Sci. Environ. Epidemiol. 1.
Stockholm Convention on Persistent Organic Pollutants, 2017. http://chm.pops.int/TheConvention/Overview/TextoftheConvention/tabid/2232/Default.aspx, Accessed date: July 2018.
Strempel, S., Scheringer, M., Ng, C.A., Hungerbühler, K., 2012. Screening for PBT chemicals among the "existing" and "new" chemicals of the EU. Environ. Sci. Technol. 46, 5680–5687.
Sverko, E., Reiner, E.J., Tomy, G.T., McCrindle, R., Shen, L., Arsenault, G., Zaruk, D., MacPherson, K.A., Marvin, C.H., Helm, P.A., McCarry, B.E., 2010. Compounds structurally related to Dechlorane Plus in sediment and biota from Lake Ontario (Canada). Environ. Sci. Technol. 44, 574–579.
Swanson, M.B., Davis, G.A., Kincaid, L.E., Schultz, T.W., Bartmess, J.E., Jones, S.L., George, E.L., 1997. A screening method for ranking and scoring chemicals by potential human health and environmental impacts. Environ. Toxicol. Chem. 16, 372–383.
Taguchi, V.Y., Nieckarz, R.J., Clement, R.E., Krolik, S., Williams, R., 2010. Dioxin analysis by gas chromatography-Fourier transform ion cyclotron resonance mass spectrometry (GC-FTICRMS). J. Am. Soc. Mass Spectrom. 21, 1918–1921.
Ubukata, M., Jobst, K.J., Reiner, E.J., Reichenbach, S.E., Tao, Q., Hang, J., Wu, Z., Dane, A.J., Cody, R.B., 2015. Non-targeted analysis of electronics waste by comprehensive two-dimensional gas chromatography combined with high-resolution mass spectrometry: Using accurate mass information and mass defect analysis to explore the data. J. Chromatogr. A 1395, 152–159.
US EPA, 2012. Estimation Programs Interface Suite™ v 4.11. United States Environmental Protection Agency, Washington, DC. In: USA.
US EPA, TSCA Chemical Substance Inventory. http://www.epa.gov/tsca-inventory (accessed Jan 2017).
Wania, F., 2003. Assessing the potential of persistent organic chemicals for long-range transport and accumulation in polar regions. Environ. Sci. Technol. 37, 1344–1351.
Wang, L., 2015. CAS Marks Multiple Milestones. Chem. Eng. News 93, 33.
Watson, J.T., Sparkman, O.D., 2007. Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation. John Wiley & Sons.
Zhang, X., Brown, T.N., Wania, F., Heimstad, E.S., Goss, K.-U., 2010. Assessment of chemical screening outcomes based on different partitioning property estimation methods. Environ. Int. 36, 514–520.
Zhang, X., Suhring, R., Serodio, D., Bonnell, M., Sundin, N., Diamond, M.L., 2016. Novel flame retardants: Estimating the physical-chemical properties and environmental fate of 94 halogenated and organophosphate PBDE replacements. Chemosphere 144, 2401–2407.

Supplementary Information

Compositional space: A guide for environmental chemists on the identification of persistent and bioaccumulative organics using mass spectrometry

Xianming Zhang[a], Robert A. Di Lorenzo[b], Paul A. Helm[a], Eric J. Reiner[a], Philip H. Howard[c], Derek C. G. Muir[d], John G. Sled[b], Karl J. Jobst[a,e]*

[a]*Ontario Ministry of the Environment, Conservation and Parks, 125 Resources Road, Toronto, Canada, M9P 3V6*
[b]*Mouse Imaging Centre, Hospital for Sick Children, 25 Orde Street, Toronto, Canada, M5T 3H7*
[c]*SRC, Environmental Science Center, 6502 Round Pond Road, North Syracuse, New York*
[d]*Canada Centre for Inland Waters, Environment and Climate Change Canada, 867 Lakeshore Rd., Burlington, ON L7S 1A1*
[e]*Department of Chemistry and Chemical Biology, McMaster University, 1280 Main St. W., Hamilton, Canada, L8S 4M1*

A single mass-to-charge (m/z) measurement is often not sufficient to identify a compound of interest, especially in the presence of interfering (matrix) compounds of a complex mixture. Consider for example two compounds listed on the Canadian DSL and US TSCA chemical inventories : 4-chloro-3,5-dimethylphenol (CAS# 88-04-0) and 4-fluoro-3-nitroaniline (CAS# 364-76-1). The former is listed as a suspected POP, whereas the latter is not. Their respective monosotopic masses differ by only 4 ppm, but the isotope pattern readily distinguishes them (see Fig. S1).

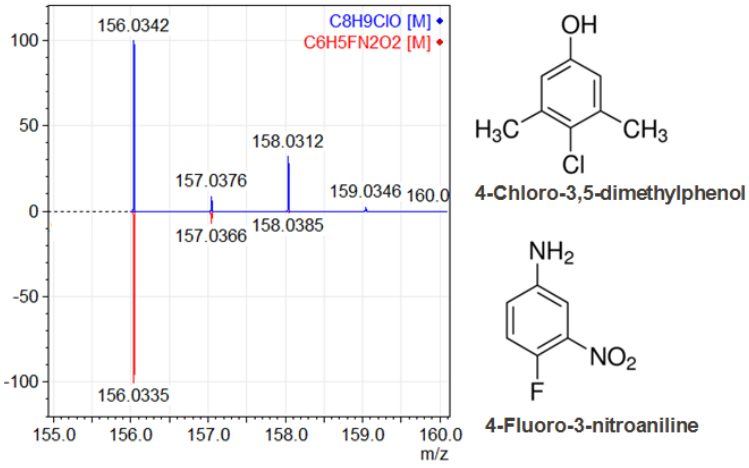

Figure S1: Partial theoretical mass spectra of 4-chloro-3,5-dimethylphenol (top) and 4-fluoro-3-nitroaniline (bottom).

While most Cl/Br containing compounds can be separated from non-halogenated compounds based on their distributions on the chemical space composed with mass, relative intensities of A-2 and A+2 isotope peaks, some compounds containing Si can cause false positives. The ratio of A+1 relative to A can be used to minimize false positives. Figure S2 displays all compounds containing Cl and Si, plotted according mass and the ratio (A+1):A. Most Si containing compounds have higher A+1 relative intensities than chlorinated compounds of similar masses. Removing compounds above the boundary depicted in Fig. S2 (represented by the equation $y = 6.5\times10^{-2}\,x + 6.8$) significantly reduces the likelihood of false positives.

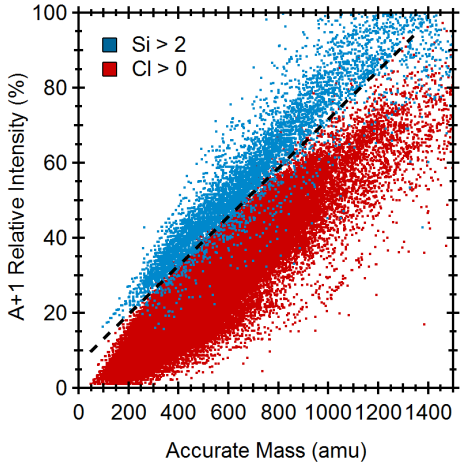

Figure S2: Isotope ratio chemical space defined by the relative intensity of the A+1 isotope for compounds with Si>2 (blue) and Cl>0 (red). Potential false positives from compounds with high Si content can be minimized.

Description of the Script

(1) First, the chemical feature data (in .csv or .xlsx format) are imported into R. The first four columns must be named “id”, “rt”, “mz”, and “intb”, corresponding to feature id number, retention time, accurate mass (m/z) and intensity.

(2) Chemical features that share the same retention time (tolerance = 0.03 min) are grouped together as a component (a pseudonym for an unknown mass spectrum). The retention time tolerance may be modified in the code in order to be compatible with different experimental data (GC- vs LC-MS data e.g.).

(3) For each component, isotopic clusters are identified on the basis of isotope specific mass differences. For example, exchange of ^{35}Cl with ^{37}Cl or ^{79}Br with ^{81}Br results in a mass shift of 1.997 amu. The exchange of 12C with 13C results in a mass shift of 1.003 amu, and the exchange of ^{28}Si with ^{29}Si results in a mass shift of 0.999 amu. A mass tolerance of 0.002 amu is employed for this search.

(4) For each isotopic cluster, the (putative) monoisotopic peak (M) is identified as peaking the lowest mass peak in the cluster. The intensoid peak (A) is the most intense isotopic peak in the cluster.

(5) Selected isotopic peak ratios are calculated, viz., (A+1.997):A, (A-1.997):A, (A+0.999):A and (A+1.003):A.

(6) The script selects Cl and Br containing compounds that meet the following criteria :

   a. (A+1.003 < A+1.997) AND (A+0.999 < A+2) - The ^{13}C and ^{29}Si isotopic peak intensity must not exceed that of the $^{37}Cl/^{81}Br$ peaks.

   b. [(A-1.997)/A=0 and (A+1.997)/A > 25%] OR [(A-1.997) > 30 and (A+1.997) >0] – Either the A-2 peak must exist or the A+2 peak must exceed 25% the intensity of A.

   c. (A+0.999)/A < 0.064*m/z + 6.8 – The ^{29}Si isotopic peak must not exceed the boundary line displayed in Fig. S2.

Table S1. Summary of (33) *m/z* values selected by the script and assigned identities using the confidence level schemed devised by Schymanksi et al. (e.g. Refs. 31 and 34). The remaining m/z values (158) were assigned confidence level 5, reflecting our interest in these as corresponding to potential POPs, but whose identities were not be proposed.

RT (min)	m/z (M)	Int. (M)	Int. (M+2)	Rel Int. (%)	Proposed formula		Proposed identity or class	Conf. Level [a]	References
7.55	955.1771	1.60E+04	2.93E+04	0.050	$C_{12}Br_{10}O$	$M^{\cdot+}$	**BDE209**	**1**	NIST 2585 CofA [b]
5.21	856.8328	3.10E+04	4.86E+04	0.097	$C_{22}H_4NSO_5Cl_3F_{17}$	$M^{\cdot+}$	**Cl/F-4**	**2b**	(Fig. SI-3a)
4.90	830.8555	2.20E+05	3.73E+05	0.685	$C_{21}H_6NSO_4Cl_4F_{17}$	$M^{\cdot+}$	**Cl/F-5**	**2b**	(Fig. SI-5a)
4.82	822.8719	5.57E+03	5.15E+03	0.017	$C_{22}H_5NSO_5Cl_3F_{17}$	$M^{\cdot+}$	**Cl/F-3**	**2b**	(Fig. SI-3b)
4.78	796.8933	1.01E+04	1.19E+04	0.031	$C_{21}H_7NSO_4Cl_3F_{17}$	$M^{\cdot+}$	**Cl/F-7**	**4**	
5.05	756.8404	1.72E+04	2.48E+04	0.053	$C_{20}H_4NSO_5Cl_4F_{13}$	$M^{\cdot+}$	**Cl/F-2**	**2b**	(Fig. SI-3c)
4.77	730.8599	4.19E+04	6.14E+04	0.130	$C_{19}H_6NSO_4Cl_4F_{13}$	$M^{\cdot+}$	**Cl/F-6b**	**2b**	(Fig. SI-5b)
4.74	730.8597	2.48E+05	3.57E+05	0.773	$C_{19}H_6NSO_4Cl_4F_{13}$	$M^{\cdot+}$	**Cl/F-6a**	**2b**	(Fig. SI-5b)
4.65	722.8773	4.18E+03	4.41E+03	0.013	$C_{20}H_5NSO_5Cl_3F_{13}$	$M^{\cdot+}$	**Cl/F-1**	**2b**	(Fig. 7b and Fig. SI-3d)
4.74	695.8914	3.56E+04	3.77E+04	0.111	$C_{19}H_5NSO_4Cl_3F_{13}$	$[M\text{-}HCl]^+$	**Cl/F-6a (fragment ion)**	**2b**	
5.89	662.4656	6.07E+06	2.75E+06	18.883	$C_{42}H_{63}PO_4$	$[M+H]^+$	**Tris(di-tert-butylphenyl)phosphate**	**3**	[a]
4.74	639.5402	3.85E+03	9.19E+03	0.012	$C_{12}H_4Br_6O$	$M^{\cdot+}$	**Br_6-BDE**	**3**	NIST 2585 CofA [b]
4.22	559.6287	1.09E+04	5.59E+04	0.034	$C_{12}H_5Br_5O$	$M^{\cdot+}$	**BDE99**	**1**	NIST 2585 CofA [b]
4.75	541.7604	2.76E+03	3.24E+03	0.009	**TBBPA**	$M^{\cdot+}$	**Tetrabromobispheno A (TBBPA)**	**1**	Megson et al. [47]
3.67	481.7159	1.66E+04	7.25E+04	0.052	$C_{12}H_6Br_4O$	$M^{\cdot+}$	**BDE47**	**1**	NIST 2585 CofA [b]
5.68	460.6688	2.59E+03	1.10E+04	0.008	C_8HBr_4O	$[M\text{-}C_{16}H_{33}O]^+$	**Bis(2-ethylhexyl)tetrabromophthalate (BEHTBP)**	**1**	Megson et al. [47]
3.28	428.8912	9.74E+04	1.91E+05	0.303	$C_9H_{16}Cl_6PO_4$	$[M+H]^+$	**Tris(1,3-dichloro-2-propyl)phosphate**	**3**	Hilton et al. [57]
3.38	428.8893	2.72E+03	5.96E+03	0.008	$C_9H_{16}Cl_6PO_4$	$[M+H]^+$	**Tris(1,3-dichloro-2-propyl)phosphate**	**3**	Hilton et al. [57]
3.18	407.7943	5.28E+03	6.72E+03	0.016	$C_{10}H_6Cl_8$	$M^{\cdot+}$	**Chlordane**	**1**	NIST 2585 CofA [b]
3.67	405.8010	3.54E+03	3.50E+03	0.011	$C_{12}H_7Br_3O$	$M^{\cdot+}$	**Br_3-BDE**	**3**	NIST 2585 CofA [b]
3.21	357.8440	4.44E+03	7.75E+03	0.014	$C_{12}H_4Cl_6$	$M^{\cdot+}$	**Cl_6-PCB**	**1**	NIST 2585 CofA [b]
3.79	354.8883	2.24E+04	3.85E+04	0.070	$C_{12}H_6Cl_5NO$		**Unknown**	**4**	
2.13	327.0071	1.64E+05	1.59E+05	0.510	$C_9H_{19}Cl_3PO_4$	$[M+H]^+$	**Tris(2-chloroisopropyl)phosphate (TDCP)**	**3**	Hilton et al. [47]; [b]
3.22	325.8823	3.28E+03	2.26E+03	0.010	$C_{12}H_5Cl_5$	$M^{\cdot+}$	**Cl_5-PCB**	**3**	NIST 2585 CofA [b]
2.95	315.9368	4.70E+04	6.20E+04	0.146	$C_{14}H_8Cl_4$	$M^{\cdot+}$	**Dichlorodiphenyldichloroethylene (DDE)**	**1**	NIST 2585 CofA [b]
4.38	311.9617	1.38E+04	1.32E+04	0.043	$C_{13}H_7Cl_3N_2O$			**4**	
4.51	305.9958	4.95E+03	3.95E+03	0.015	$C_{14}H_8Cl_2N_2O_4$		**Unknown**	**4**	
2.04	285.9633	2.63E+04	2.52E+04	0.082	$C_{12}H_8Cl_2O_2S$	$M^{\cdot+}$	**Fentichlor**	**3**	
2.04	284.9602	4.11E+05	3.97E+05	1.279	$C_6H_{12}Cl_3PO_4$	$[M+H]^+$	**Tris(2-chloroethyl)phosphate**	**3**	Hilton et al. [47]
2.79	281.9756	5.04E+03	4.79E+03	0.016	$C_{14}H_9Cl_3$	$M^{\cdot+}$	**1,1-bis(p-chlorophenyl)-2-chloroethene (DDMU)**	**4**	NIST 2585 CofA [b]
1.89	173.0094	1.47E+04	5.17E+03	0.046	$(^{13}C_1)\ C_6H_5N_2O_2Cl$	$M^{\cdot+}$	**Chloronitroaniline**	**3**	
4.63	127.0166	4.15E+04	1.29E+04	0.129	C_6H_6NCl			**4**	
4.58	125.0140	3.91E+04	1.34E+04	0.122	C_7H_6Cl			**4**	Hilton et al. [57]

Notes : [a] Confidence levels were assigned according to the scheme proposed by Schymanski et al., see e.g. refs. 31 and 34; [b] Reported in Certificate of Analysis for SRM 2585.

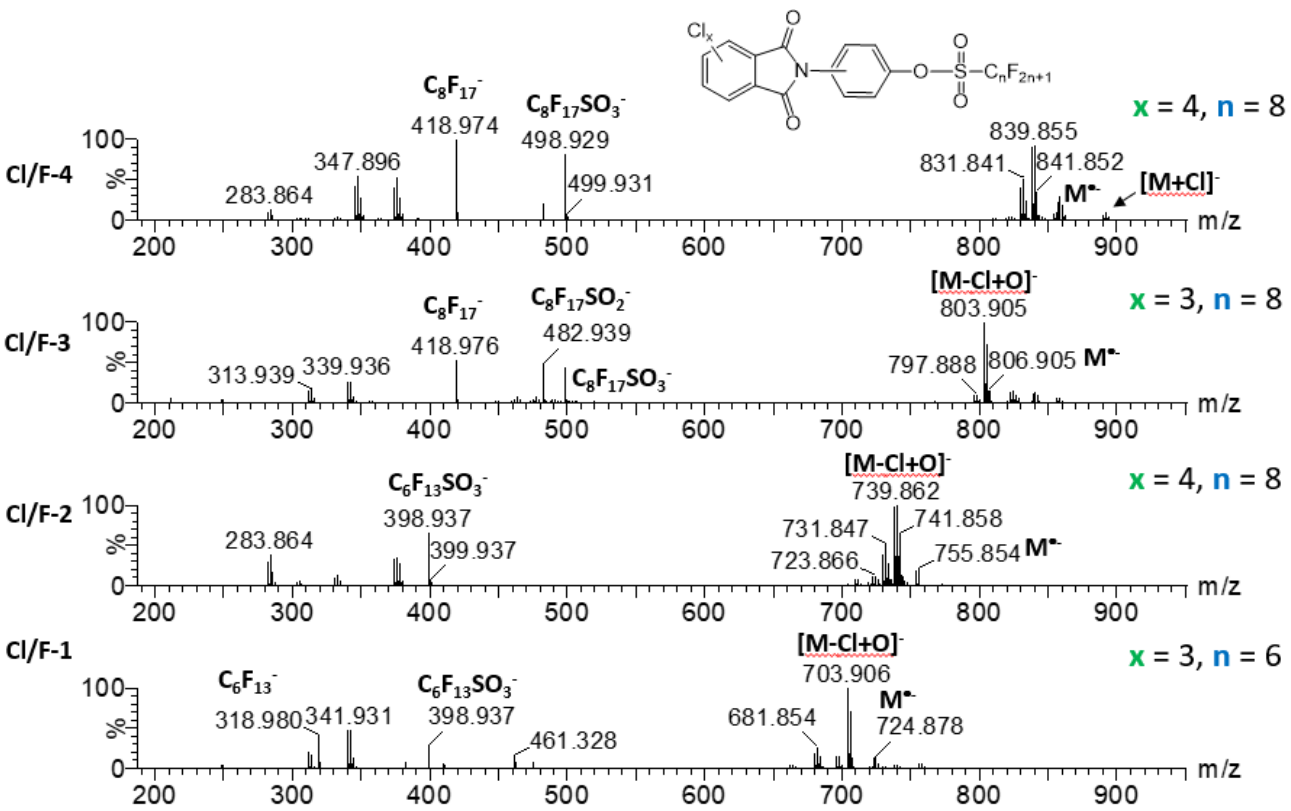

Figure S3: Negative ion mode APCI mass spectra obtained from the mixed chlorofluoro phthalimides **(Cl/F-1 – 4)** reported in this study.

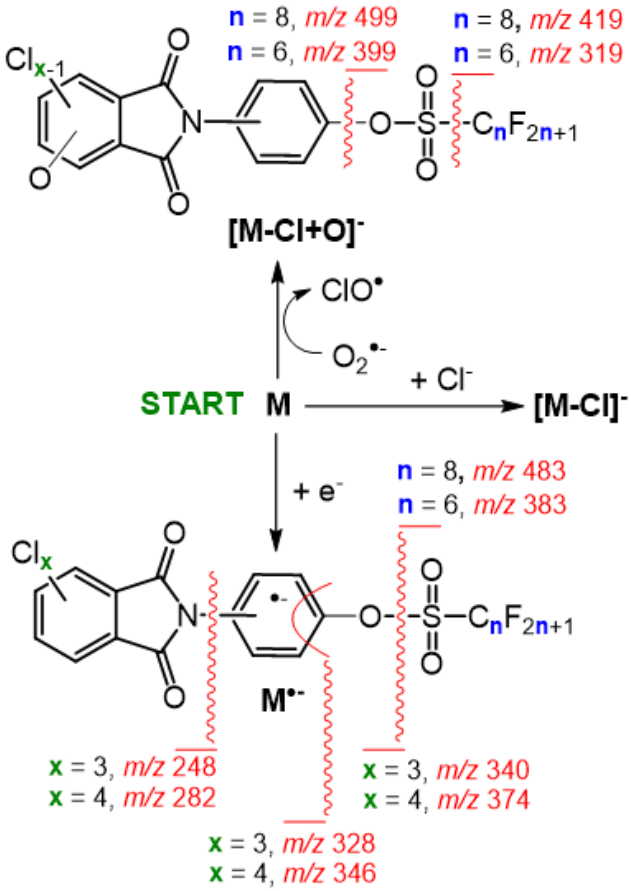

Figure S4: Negative mode ionization of compounds **Cl/F-1,2,3** and **-4** result in the formation of $M^{-\bullet}$, $[M\text{-}Cl+O]^-$ and $[M+Cl]^-$. Direct bond cleavages of these (quasi)molecular ions readily explain the main fragmentations observed in the mass spectra shown in Figure S3.

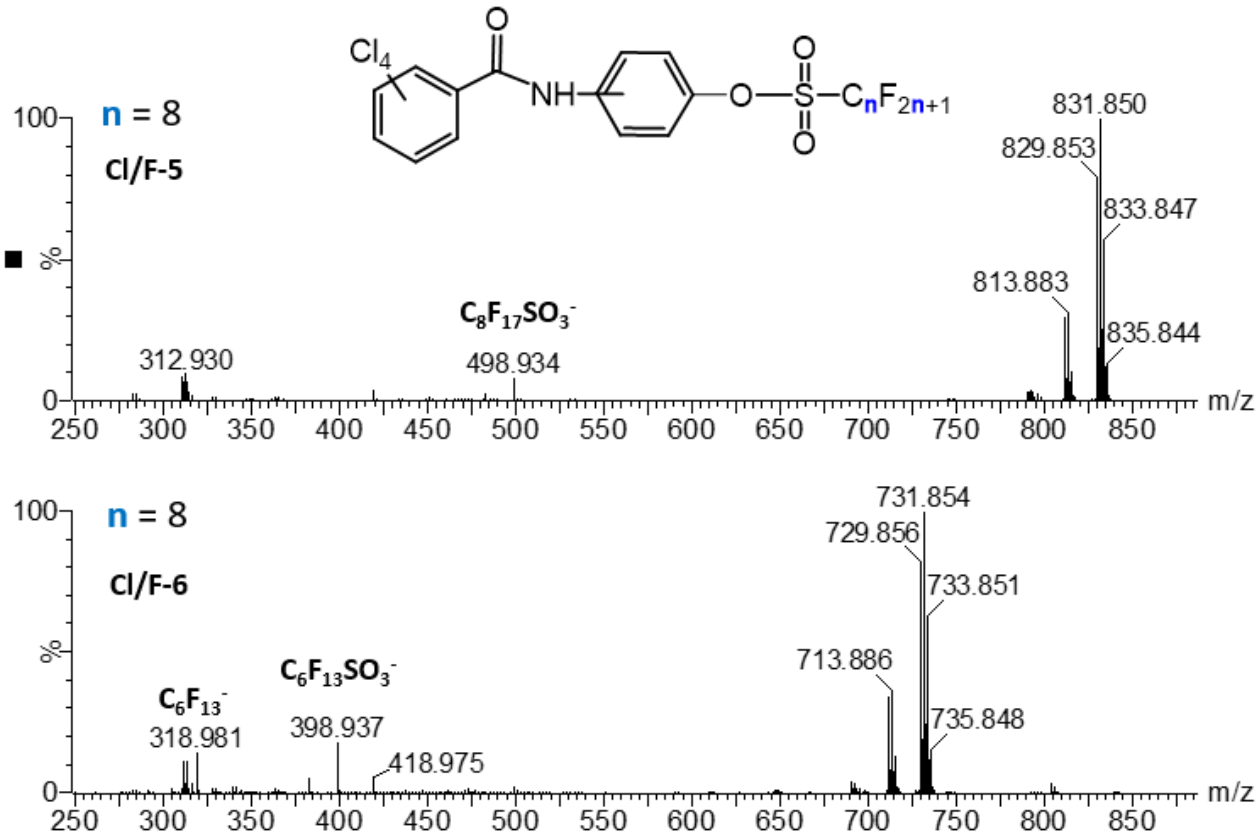

Figure S5: Negative ion mode APCI mass spectra obtained from the mixed chlorofluoro compounds **Cl/F-5** and **-6.**

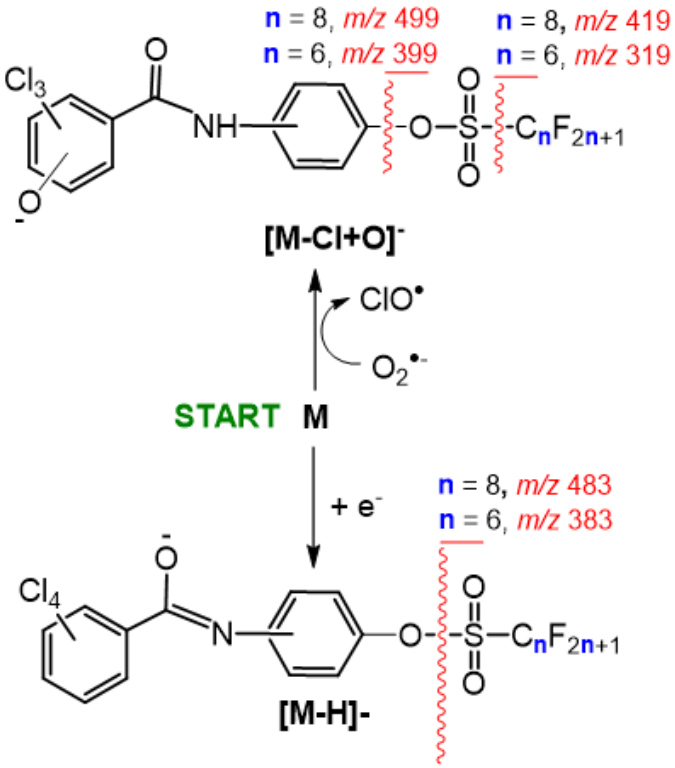

Figure S6: Negative mode ionization of compounds **Cl/F-5** and **-6** result in the formation of $[M\text{-}Cl+O]^-$ and $[M\text{-}H]^-$. Direct bond cleavages of these (quasi)molecular ions readily explain the main fragmentations observed in the mass spectra shown in Figure S5.

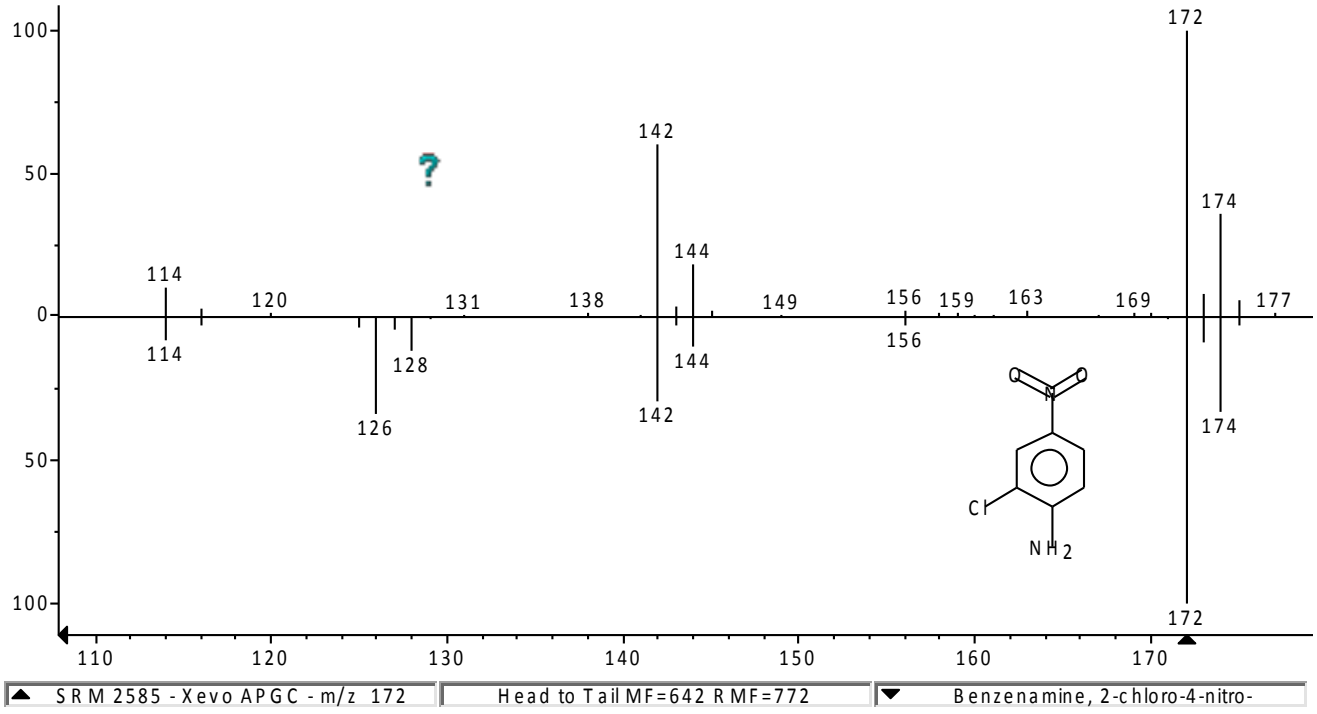

Figure S7: Positive ion mode APCI mass spectrum obtained from a $C1_1$ ion selected by the filter. A comparison with the NIST library suggests that the compound may be a positional isomer of chloronitroaniline.