



HHS Public Access

Author manuscript

Environ Int. Author manuscript; available in PMC 2020 November 01.

Published in final edited form as:

Environ Int. 2019 November ; 132: 104808. doi:10.1016/j.envint.2019.05.002.

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

Xianming Zhang¹, Robert Di Lorenzo², Paul A. Helm¹, Eric J. Reiner¹, Philip H. Howard³, Derek Muir⁴, John G. Sled², Karl J. Jobst^{1,5,*}

¹Ontario Ministry of the Environment, Conservation and Parks, 125 Resources Road, Toronto, Canada, M9P 3V6

²Mouse Imaging Centre, Hospital for Sick Children, 25 Orde Street, Toronto, Canada, M5T 3H7

³SRC, Environmental Science Center, 6502 Round Pond Road, North Syracuse, New York

⁴Canada Centre for Inland Waters, Environment and Climate Change Canada, 867 Lakeshore Rd., Burlington, ON L7S 1A1

⁵Department of Chemistry and Chemical Biology, McMaster University, 1280 Main St. W., Hamilton, Canada, L8S 4M1

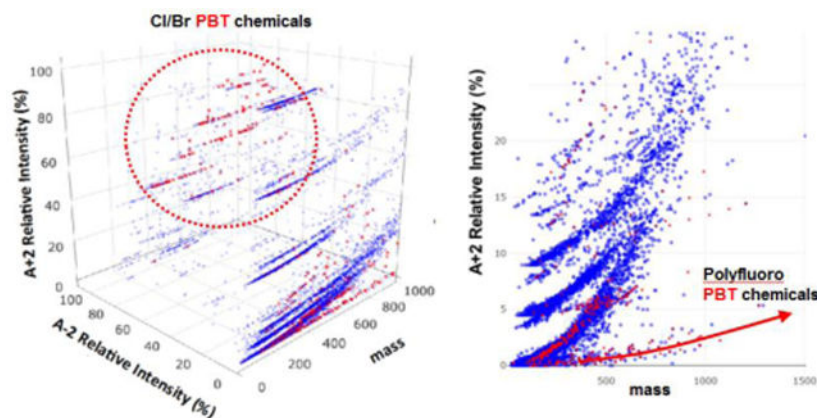
Abstract

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 chemicals.

Graphical Abstract

*Corresponding Author: Karl.Jobst@Ontario.ca.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Keywords

Persistent organic pollutants; nontargeted screening; environmental mass spectrometry; flame retardants; chemical space

INTRODUCTION

Tens of thousands of chemical substances have been introduced to the global market.^{1,2} 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.³ 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);⁴ the U.S. Toxic Substances Control Act Inventory (US TSCA);⁵ the European Inventory of Existing Commercial Chemical Substances (EINECS)⁶, and the Inventory of Existing Chemical Substances in China (IECSC)⁷]. Persistent organic pollutants (POPs) are an important subset of environmental 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 and, since 2001, the Stockholm Convention has banned or restricted the release of twenty-eight groups of halogenated POPs.⁸ It is not currently practical to conduct detailed environmental monitoring and risk assessments for all halogenated POPs, 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).⁹⁻¹⁴ 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 (very) long degradation half-lives, among other intrinsic properties. These may be predicted from quantitative structure activity relationships (QSARs)¹⁵ 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.^{16,17}

The pioneering studies of Howard and Muir,¹⁰ Wania and Brown¹⁸, Stempel *et al.*¹² and Scheringer *et al.*¹⁹ 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.¹⁰ Of these, 62% are halogenated and 8% are siloxanes. Stempel *et al.*¹² and Scheringer *et al.*¹⁹ 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 results²⁰, many of these potential POPs have been identified in environmental and biological media.²¹⁻²³

There are however drawbacks to the top-down approach. Uncertainties from QSARs can result in ~25% false positive or negative results.²⁴ 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.¹⁰ 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.^{25,26} 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.²⁷⁻³⁰

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.^{31,32} Recent developments in multi-dimensional chromatography coupled to HRMS have raised the possibility of detecting thousands of chemical components in a single analysis.³³⁻³⁶ 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 (prioritized) suspect chemicals may be accomplished using the time-honoured approach of spectral library searching and interrogating the experimental data for the presence of expected (pseudo)molecular ions, isotopic peak ratios and fragment ions.^{37,38} 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 (pseudo)molecular ions (M^{*+} , $M+H^+$ 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^{39,40}. Not every combination of elements is possible. Lobodin *et al.*⁴¹ coined the term

compositional space, noting that a boundary line defined by molecular weight (m/z) and fractional mass 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.^{27,33,42,43} 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*.⁴⁴ 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.

METHODS

Isotope clustering

A single mass-to-charge (m/z) measurement is often not sufficient to identify a compound of interest,³⁹ 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 POP¹⁰, whereas the latter is not. Their respective monoisotopic masses differ by only 4 ppm, but the isotope pattern readily distinguishes them (see SI Figure S1).

Mass defect (MD) of a molecule derived from accurate mass measurements is useful to distinguish elemental compositions. As illustrated by Figure 1 as 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. 5ppm) and chromatographic separation⁴⁵. (In practice, peak

picking, deconvolution and alignment across multiple samples can be quite challenging if the experimental conditions drift). ^{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), see Fig. 1c.

Isotopic patterns were calculated for the compounds in the chemical databases (see below) using custom Excel macros described by Kind and Fiehn.⁴⁵ Isotopic peaks in experimental data were clustered using a script tool (R code, see supplementary information) from deconvoluted mass spectra (IntelliXtract, ACD/Labs MS Workbook Suite 2012) Briefly, the R code performs four fundamental tasks : (1) In each mass spectrum, isotopic clusters are identified on the basis of isotope specific mass differences using a mass tolerance of 0.002 amu; (2) 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; (3) Selected peak ratios are calculated, viz. $(A+1.997)/A$, $(A-1.997)/A$, $(A+0.999)/A$ and $(A+1.003)$; and (4) 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.

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.⁴⁶ 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. The database is composed of (synthesized) chemicals, but many have not 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.²

Compositional spaces were also constructed using a database of 22,043 chemicals used in commerce and industry merged from the US TSCA (14,376 chemicals manufactured or imported in a quantity exceeding 25,000 pounds)⁵ and the Canadian DSL (11,317 compounds exceeding a quantity of 100kg/y).⁴ Based on this merged database, Howard and Muir¹⁰ 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.

Dust Sample Analysis

A Standard Reference Material (SRM 2585) of household dust was analyzed to demonstrate the application of compositional space filtering. The sample (50mg) was extracted with hexane ($3 \times 5\text{mL}$), centrifuged and filtered with Na_2SO_4 , and concentrated to a final volume of $500\mu\text{L}$. The extract ($1\mu\text{L}$) 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)⁴⁷ 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 $N_2^{*\bullet+}$, resulting in the formation of radical cations $M^{*\bullet+}$.^{48,49} Therefore, the theoretical mass and isotope ratios calculated from the chemical inventories will be directly comparable to the experimental measurements. Fast chromatographic separation⁴⁹ was performed using a DB-5 (15m × 0.25mm × 0.1 μm) column. The transfer line temperature was 340°C and the makeup gas flow (N_2) was 350mL/min. The injector temperature was held at 280 °C. The oven temperature was held at 90 °C for 1 minute, then ramped to 330°C at 30°C/min and held for 5 minutes. 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 data set is available in the Supporting Information.

RESULTS AND DISCUSSION

Distribution of Halogenated Organic Compounds in Compositional Space

Figure 2 displays a mass defect plot^{27,50} of the 305,135 compounds in the Kind and Fiehn⁴⁶ compilation of PubChem chemicals. Note that the standard IUPAC (International Union of Pure and Applied Chemistry) mass scale ($C = 12$) is used to construct Figure 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.*⁴¹ 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. Figure 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 fractional mass 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.^{10,19} 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 calibrated using non-standard mass scales (*e.g.* $CH_2 = 14$ amu, $CF_2 = 50$ amu, $H/Cl = 34$ amu).^{27,51-56} 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 Figures 2b and 2c 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.

Figure 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 greater than 50%) where few non-halogenated compounds are present.

Figure 4 displays an expanded view of the fine structure of "non-halogenated" region shown in Figure 3 A : 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*⁴⁶ observed that for compositions with a molecular weight below 400 amu, the number of sulphur and silicone atoms in a molecule can be derived based on the location of a chemical in this space (Figure 4b and c). Above m/z 400, the Si and S bands converge. The rising curve is a consequence of increasing contribution from ¹³C to the M+2 peak. Figure 4 was constructed assuming the ³⁴S and ¹³C² 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 ¹³C 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 (pseudo)molecular ions, fragments and isotopic peaks. The relative intensity of the M+2 isotope is diagnostic of (unknown) halogenated organic compounds (see Figure 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 more than 10 Cl atoms or 4 Br atoms, the relative abundance of M is less than 20% of the most abundant isotope peak (A).

Figure 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 Cl₁, Cl₂ and some Cl₃ compositions located on the face, where (A+2):A>30%

and $(A-2)/A=0$. For Br and $>Cl_2$ containing compounds, the intensity of the $A-2$ peak relative to A is greater than 30%. The bands are labelled on the basis of Cl and Br content in Figures 5c-e, which excludes the m/z dimension. In contrast, most (non-halogenated) compositions (c. 97%) 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 Figure 3A). Poly-/perfluoroalkyl substances (highlighted in yellow, Figures 4d and 5) are characterized by relatively low $A+2$ intensities relative to m/z . Figure 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).

Projecting the Chemical Space of Persistent and Bioaccumulative Organics onto Compositional Space

Figure 6a displays 22 043 chemicals in the North American chemical inventories (Canadian DSL and US TSCA).^{4,5} The 610 prioritized¹⁰ compounds are also displayed in Figure 6 (in red). Of the 610 compounds, 216 contain Cl and/or Br and they all distribute in the region in which the relative intensities of $A+2$ are greater than 30%. However, the majority of these Cl/Br compounds, 149 (69%), occupy the region where the relative intensities of the $A-2$ 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 relative intensities of the $A+2$ isotopes are larger than 60% and that of the $A-2$ isotopes larger than 50%. This region is predominantly occupied by compounds with more than 4 Cl/Br atoms, including most regulated POPs.

There are four sub-regions (I-IV) in Figure 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)¹⁰ via *in-silico* screening. The *in-silico* screening identified 19 P & B out of 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 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 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 or fractional mass, which is related to the number double bond equivalents (DBE)⁵⁰.

Figure 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 article⁵⁶, the most popular (and only reported) 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, Figure 6b shows that isotopic ratios (viz. $^{13}C/^{12}C$) can be used to 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 $(A+2):A = 4-6\%$, see inset of 6b (top), 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 better, as indicated using the Muir and Howard criteria¹⁰. 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.

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).⁵⁷ 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.

Figure 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), organochlorine pesticides (OCPs) such as 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 (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.⁵⁷ Fentichlor, a topical fungicide was tentatively identified, but to our knowledge, has not been reported previously in SRM 2585^{57,58}.

Searching on the basis of mass accuracy, 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.*^{31,34}. 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^{27,28}. Thus, the remaining isotopic clusters correspond to true "unknowns" of which (9) 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 C₂F₄ substitution; (ii) the abundance of these compounds is greater than those of well-known POPs, including the PBDEs.

The positive ion mode spectra of these compounds are characterized by intense molecular ion peaks (M^{+}), which shift by 19 mass units to form ions $[M-Cl+O]^{-}$ under negative ion mode conditions.⁵⁹ 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.⁶⁰

The mass spectrum (negative ion mode) of one of the C₁₃ compounds (denoted as **Cl/F-1**) is displayed in Figure 7b. The spectrum displays an intense pseudomolecular ion and the presence of two chlorine atoms in the ion (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.* C₈HCl₃NO₂⁻ at m/z 247.908 (2.8ppm) and C₆F₁₃SO₃⁻ at m/z 398.938 (4.8ppm), provide important clues on the identity of this contaminant. The C₆F₁₃SO₃⁻ fragment is unlikely to have any structure other than that of a perfluorohexane sulfonate ion, while the C₈HCl₃NO₂⁻ 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 C₈F₁₇ and between three and four Cl substituents on the phthalimide ring. Their mass spectra are displayed in Figure S3a-c and one observes that the most intense peaks in the spectra shift according to the presence of 3-4 Cl atoms and C₆F₁₃ or C₈F₁₇ moieties. Indeed, all of the peaks display in these mass spectra correspond to direct bond cleavages of the proposed phthalimide structure shown in Figure S4.

Similar to Cl/F-1,2,3 and -4, compounds Cl/F-5 and -6 display (negative ion mode) mass spectra (Figure S5) that are dominated by [M-Cl+O]⁻ ions. However, a peak corresponding to [M-H]⁻, rather than M⁻, is also present and suggests that compounds Cl/F-5 and -6 have at least one ionizable proton. This is consistent with the proposal that Cl/F-5 and -6 are phthalamides rather than phthalimides, see Figures S5 and S6. We note that Strynar et al. have previously reported the detection of Cl/F-6 in SRM2585⁶³. Cl/F-7, a Cl₃ 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 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.^{61,62} 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 Muir¹⁰. The mixed halogenated compounds identified in this study are closely related to CAS# 68815-72-5, a high production volume chemical and suspected POP¹⁰. 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* have detected compound **Cl/F-6** by LC-MS⁶³, which suggests that degradation of 68815-72-5 likely occurs during production.

Collectively, these compounds represent precursors of perfluoroalkyl sulfonates that have been restricted by the Stockholm Convention. The identification of these compounds represents a first detection in (indoor) environmental media. This example may serve to demonstrate the effectiveness of the approach to reveal unknown contaminants, but the unusually high molecular weights of these mixed halogenated phthalimides places them outside the compositional space of known and suspected POPs : the Cl₄ compounds prioritized by *in silico* screening range between 200 and 500 amu. This nevertheless highlights the challenge posed by degradation and biotransformation 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),^{64,65} 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 logKow, 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.^{65,66}

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,^{67,68,69} 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Ron 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 of Health Grant U01-HD-087177-01.

References

1. CAS Marks Multiple Milestones. Chemical & Engineering News 2015, 93, 33.
2. CAS Registry. <http://www.cas.org/support/documentation/chemical-substances> (accessed July 2018)
3. Swanson MB; Davis GA; Kincaid LE; Schultz TW; Bartmess JE; Jones SL; George EL, A screening method for ranking and scoring chemicals by potential human health and environmental impacts. Environ. Toxicol. Chem 1997, 16, 372–383.
4. Environment and Climate Change Canada, Domestic substances list. <https://http://www.canada.ca/en/environment-climate-change/services/canadian-environmental-protection-registry/substances-list/domestic.html> (accessed Jan 2017).
5. US EPA, TSCA Chemical Substance Inventory. <http://www.epa.gov/tsca-inventory> (accessed Jan 2017).
6. European Chemical Agency (ECHA), European Inventory of Existing Commercial Chemical Substances (EINECS). <https://echa.europa.eu/information-on-chemicals/ec-inventory> (7 2018).
7. Chemical Inspection and Regulation Service, Chinese Chemical Inventory of Existing Chemical Substances (IECSC). <http://cciss.cirs-group.com/> (8 2018).

8. Stockholm Convention on Persistent Organic Pollutants. <http://chm.pops.int/TheConvention/Overview/TextoftheConvention/tabid/2232/Default.aspx> (accessed July 2018).
9. Gramatica P; Cassani S; Sangion A, PBT assessment and prioritization by PBT index and consensus modeling: comparison of screening results from structural models. *Environ. Int* 2015, 77, 25–34. [PubMed: 25617903]
10. Howard PH; Muir DC, Identifying new persistent and bioaccumulative organics among chemicals in commerce. *Environ. Sci. Technol* 2010, 44, 2277–85. [PubMed: 20163179]
11. Muir DC; Howard PH, Are there other persistent organic pollutants? A challenge for environmental chemists. *Environ. Sci. Technol* 2006, 40, 7157–7166. [PubMed: 17180962]
12. Stempel S; Scheringer M; Ng CA; Hungerbühler K, Screening for PBT chemicals among the “existing” and “new” chemicals of the EU. *Environ. Sci. Technol* 2012, 46, 5680–5687. [PubMed: 22494215]
13. Reppas-Chrysovitinos E; Sobek A; MacLeod M, Screening-level exposure-based prioritization to identify potential POPs, vPvBs and planetary boundary threats among Arctic contaminants. *Emerging Contaminants* 2017, 3, 85–94.
14. Zhang X; Suhring R; Serodio D; Bonnell M; Sundin N; Diamond ML, Novel flame retardants: Estimating the physical-chemical properties and environmental fate of 94 halogenated and organophosphate PBDE replacements. *Chemosphere* 2016, 144, 2401–2407. [PubMed: 26613357]
15. US EPA, Estimation Programs Interface Suite™ v 4.11 United States Environmental Protection Agency, Washington, DC, USA 2012.
16. Wania F, Assessing the potential of persistent organic chemicals for long-range transport and accumulation in polar regions. *Environ. Sci. Technol* 2003, 37, 1344–1351.
17. Gawor A; Wania F, 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.: Process Impact* 2013, 15, 1671–1684.
18. Brown TN; Wania F, Screening chemicals for the potential to be persistent organic pollutants: A case study of Arctic contaminants. *Environ. Sci. Technol* 2008, 42, 5202–5209. [PubMed: 18754370]
19. Scheringer M; Stempel S; Hukari S; Ng CA; Blepp M; Hungerbuhler K, How many persistent organic pollutants should we expect? *Atmospheric Pollution Research* 2012, 3, 383–391.
20. Reppas-Chrysovitinos E; Sobek A; MacLeod M; In Silico Screening-Level Prioritization of 8468 Chemicals Produced in OECD Countries to Identify Potential Planetary Boundary Threats. *Bulletin of Environmental Contamination and Toxicology* 2018, 100(1):134–146. [PubMed: 29285590]
21. Sverko E; Reiner EJ; Tomy GT; McCrindle R; Shen L; Arsenault G; Zaruk D; MacPherson KA; Marvin CH; Helm PA; McCarry BE, Compounds structurally related to Dechlorane Plus in sediment and biota from Lake Ontario (Canada). *Environ. Sci. Technol*, 2010, 44, 574–579. [PubMed: 20025285]
22. D’eon JC; Crozier PW; Furdui VI; Reiner EJ; Libelo EL; Mabury SA, Observation of a commercial fluorinated material, the polyfluoroalkyl phosphoric acid diesters, in human sera, wastewater treatment plant sludge, and paper fibers. *Environ. Sci. Technol* 2009, 43, 4589–4594. [PubMed: 19603681]
23. Nyholm JR; Grabic R; Arp HPH; Moskeland T; Andersson PL, Environmental occurrence of emerging and legacy brominated flame retardants near suspected sources in Norway. *Sci. Total Environ* 2013, 443, 307–314. [PubMed: 23201697]
24. Zhang X; Brown TN; Wania F; Heimstad ES; Goss K-U, Assessment of chemical screening outcomes based on different partitioning property estimation methods. *Environ. Int* 2010, 36, 514–520. [PubMed: 20451252]
25. Shen L; Reiner EJ; Helm PA; Marvin CH; Hill B; Zhang X; MacPherson KA; Kolic TM; Tomy GT; Brindle ID, Historic trends of Dechloranes 602, 603, 604, Dechlorane Plus and other norbornene derivatives and their bioaccumulation potential in Lake Ontario. *Environ. Sci. Technol* 2011, 45, 3333–3340. [PubMed: 21434636]

26. Hoh E; Zhu L; Hites RA, Dechlorane plus, a chlorinated flame retardant, in the Great Lakes. *Environ. Sci. Technol* 2006, 40, 1184–9. [PubMed: 16572773]
27. Jobst KJ; Shen L; Reiner EJ; Taguchi VY; Helm PA; McCrindle R; Backus S, The use of mass defect plots for the identification of (novel) halogenated contaminants in the environment. *Anal. Bioanal. Chem* 2013, 405, 3289–3297. [PubMed: 23354579]
28. Shen L; Jobst KJ; Helm PA; Reiner EJ; McCrindle R; Tomy GT; Backus S; Brindle ID; Marvin CH, 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* 2012, 404, 2737–48. [PubMed: 22718042]
29. Shen L; Jobst KJ; Reiner EJ; Helm PA; McCrindle R; Taguchi VY; Marvin CH; Backus S; MacPherson KA; Brindle ID, Identification and occurrence of analogues of dechlorane 604 in Lake Ontario sediment and their accumulation in fish. *Environ. Sci. Technol* 2014, 48, 11170–11177. [PubMed: 25184555]
30. Brazeau AL; Pena-Abaurrea M; Shen L; Riddell N; Reiner EJ; Lough AJ; McCrindle R; Chittim B, Dechlorinated Analogues of Dechlorane Plus. *Environ. Sci. Technol* 2018, 52, 5619–5624. [PubMed: 29659266]
31. Hollender J; Schymanski EL; Singer HP; Ferguson PL, Nontarget screening with high resolution mass spectrometry in the environment: ready to go? *Environ. Sci. Technol* 2017, 51, 11505–11512. [PubMed: 28877430]
32. Sobus JR; Wambaugh JF; Isaacs KK; Williams AJ; McEachran AD; Richard AM; Grulke CM; Ulrich EM; Rager JE; Strynar MJ, Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. *J. Expo. Sci. Environ. Epidemiol* 2017, 1.
33. Pena-Abaurrea M; Jobst KJ; Ruffolo R; Shen L; McCrindle R; Helm PA; Reiner EJ, 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* 2014, 48, 9591–9599. [PubMed: 24999818]
34. Schymanski EL; Singer HP; Slobodnik J; Ipolyi IM; Oswald P; Krauss M; Schulze T; Haglund P; Letzel T; Grosse S, Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. *Anal. Bioanal. Chem* 2015, 407, 6237–6255. [PubMed: 25976391]
35. Fernando S; Jobst KJ; Taguchi VY; Helm PA; Reiner EJ; McCarry BE, 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* 2014, 48, 10656–10663. [PubMed: 25133985]
36. Muscalu AM; Górecki T, Comprehensive Two-Dimensional Gas Chromatography in Environmental Analysis. *TrAC Trends in Analytical Chemistry* 2018.
37. Watson JT; Sparkman OD, Introduction to mass spectrometry: instrumentation, applications, and strategies for data interpretation. John Wiley & Sons: 2007.
38. Schymanski EL; Jeon J; Guide R; Fenner K; Ruff M; Singer HP; Hollender J (2014) Identifying small molecules via high resolution mass spectrometry: communicating confidence. ACS Publications.
39. Kim S; Rodgers RP; Marshall AG, Truly “exact” mass: Elemental composition can be determined uniquely from molecular mass measurement at ~ 0.1 mDa accuracy for molecules up to 500 Da. *International Journal of Mass Spectrometry* 2006, 251, 260–265.
40. Jobst C; Thein Identisch Caffein, *Chemie und Pharmaceutische Chemie Insbesondere*, 1838, 25, 63–66.
41. Lobodin VV; Marshall AG; Hsu CS, Compositional space boundaries for organic compounds. *Anal. Chem* 2012, 84, 3410–3416. [PubMed: 22376063]
42. Liu Y; Pereira ADS; Martin JW, Discovery of C5–C17 poly- and perfluoroalkyl substances in water by in-line SPE-HPLC-Orbitrap with in-source fragmentation flagging. *Anal. Chem* 2015, 87, 4260–4268. [PubMed: 25818392]
43. Cariou R; Omer E; Leon A; Dervilly-Pinel G; Le Bizec B; Screening halogenated environmental contaminants in biota based on isotopic pattern and mass defect provided by high resolution mass spectrometry profiling. *Anal. Chim. Acta* 2016, 936, 130–138. [PubMed: 27566348]

44. Milman BL; Zhurkovich IK, The chemical space for non-target analysis. *TrAC Trends in Analytical Chemistry* 2017.
45. Byer JD; Siek K; Jobst KJ; Distinguishing the C3 vs SH4 mass split by comprehensive two-dimensional gas chromatography-high resolution time-of-flight mass spectrometry, *Anal. Chem.* 2016, 88(12), 6101–6104. [PubMed: 27269256]
46. Kind T; Fiehn O, Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 2007, 8, 105. [PubMed: 17389044]
47. Megson D; Robson M; Jobst KJ; Helm PA; Reiner EJ; 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.* 2016, 88(23), 11406–11411. [PubMed: 27934115]
48. McEwen CN, GC/MS on an LC/MS instrument using atmospheric pressure photoionization. *International Journal of Mass Spectrometry* 2007, 259, 57–64.
49. Di Lorenzo RA; Lobodin VV; Cochran J; Kolic T; Besevic S; Sled JG; Reiner EJ; Jobst KJ; Fast gas chromatography-atmospheric pressure (photo)ionization mass spectrometry of polybrominated diphenylether flame retardants, *Anal. Chim. Acta* 2019, 1056, 70–78. [PubMed: 30797463]
50. Hughey CA; Hendrickson CL; Rodgers RP; Marshall AG; Qian K, Kendrick mass defect spectrum: a compact visual analysis for ultrahigh-resolution broadband mass spectra. *Anal. Chem* 2001, 73, 4676–4681. [PubMed: 11605846]
51. Taguchi VY; Nieckarz RJ; Clement RE; Krolik S; Williams R, Dioxin analysis by gas chromatography-Fourier transform ion cyclotron resonance mass spectrometry (GC-FTICRMS). *J. Am. Soc. Mass Spectrom* 2010, 21, 1918–1921. [PubMed: 20727785]
52. Myers AL; Jobst KJ; Mabury SA; Reiner EJ, Using mass defect plots as a discovery tool to identify novel fluoropolymer thermal decomposition products. *J. Mass Spectrom* 2014, 49, 291–296. [PubMed: 24719344]
53. Ubukata M; Jobst KJ; Reiner EJ; Reichenbach SE; Tao Q; Hang J; Wu Z; Dane AJ; Cody RB, 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* 2015, 1395, 152–159. [PubMed: 25869800]
54. D'Agostino LA; Mabury SA, Identification of novel fluorinated surfactants in aqueous film forming foams and commercial surfactant concentrates. *Environ. Sci. Technol* 2013, 48, 121–129. [PubMed: 24256061]
55. Barzen-Hanson KA; Roberts SC; Choyke S; Oetjen K; McAlees A; Riddell N; McCrindle R; Ferguson PL; Higgins CP; Field JA, Discovery of 40 classes of per-and polyfluoroalkyl substances in historical aqueous film-forming foams (AFFFs) and AFFF-impacted groundwater. *Environ. Sci. Technol* 2017, 51, 2047–2057. [PubMed: 28098989]
56. Liu Y; D'Agostino LA; Qu Quangbo; Jiang G; Martin JW; 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).
57. Hilton DC; Jones RS; Sjodin A, A method for rapid, non-targeted screening for environmental contaminants in household dust. *J. Chromatogr. A* 2010, 1217, 6851–6856. [PubMed: 20864112]
58. Ouyang X; Weiss JM; de Boer J; Lamoree MH; Leonards PEG; Non-target analysis of household dust and laundry dryer lint using comprehensive two-dimensional liquid chromatography coupled with time-of-flight mass spectrometry, *Chemosphere*, 2017, 166, 431–437. [PubMed: 27705830]
59. Carroll D; Dzidic I; Stillwell R; Haegele K; Horning E, Atmospheric pressure ionization mass spectrometry. Corona discharge ion source for use in a liquid chromatograph-mass spectrometer-computer analytical system. *Anal. Chem* 1975, 47, 2369–2373.
60. Plassmann MM; Fischer S; Benskin JP; Nontarget time trend screening in human blood, *Environmental Science and Technology Letters*, 2018, 5(6), 335–340.
61. Cohnen W; Kircher K; Muller PR; Krishnan S; Neuray D (1985) Polycarbonate molding compositions having improved flame retardance. US Patent: US4552911A.

62. Avakian RW (1985) Stabilization of flame retardant polycarbonate-polyester compositions. US Patent: US4555540A.
63. Strynar M. Using point of use sampling devices and high resolution mass spectrometry techniques for characterizing drinking water exposures; National Environmental Monitoring Conference; Ascona. August 8–12, 2016;
64. Ong VS; Hites RA, Electron capture mass spectrometry of organic environmental contaminants. *Mass Spectrom. Rev* 1994, 13, 259–283.
65. Fernando S; Green MK; Organtini K; Dorman F; Jones R; Reiner EJ; Jobst KJ, Differentiation of (mixed) halogenated dibenzo-p-dioxins by negative ion atmospheric pressure chemical ionization. *Anal. Chem* 2016, 88, 5205–5211. [PubMed: 27074061]
66. Jobst KJ; De Winter J; Flammang R; Terlouw JK; Gerbaux P; Differentiation of the pyridine radical cation from its distonic isomers by ion-molecule reactions with dioxygen, *Int. J. Mass Spectrom*, 2009, 286, 83–88.
67. Peng H; Chen C; Saunders DM; Sun J; Tang S; Codling G; Hecker M; Wiseman S; Jones PD; Li A, Rockne KJ; Giesy JP, 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* 2015, 87, 10237–10246. [PubMed: 26379008]
68. Peng H; Saunders DM; Sun J; Jones PD; Wong CK; Liu H; Giesy JP, Mutagenic azo dyes, rather than flame retardants, are the predominant brominated compounds in house dust. *Environ. Sci. Technol* 2016, 50, 12669–12677. [PubMed: 27934287]
69. Liu Y; Richardson ES; Derocher AE; Lunn NJ; Lehmler H-J, Li X; Zhang Y; Yue Cui J; Cheng L; Martin JW; Hundreds of unrecognized halogenated contaminants discovered in polar bear serum, *Angew. Chem* 2018, in press. 10.1002/anie.201809906

Highlights

- Elemental composition determination is the first step to identifying a contaminant.
- Boundaries in *compositional space* enclose persistent organic pollutants (POPs).
- ~305,134 compounds (PubChem) were used to visualize the space occupied POPs.
- A script tool was developed to select potential POPs from high resolution MS data.
- Unknown Cl/F-flame retardants were identified in house dust.

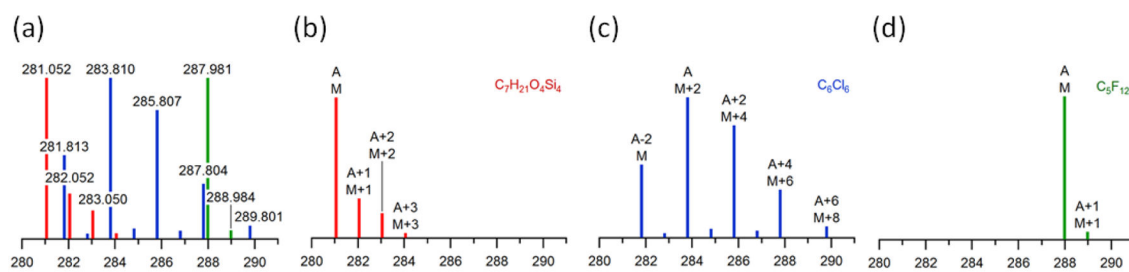


Figure 1.

(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.

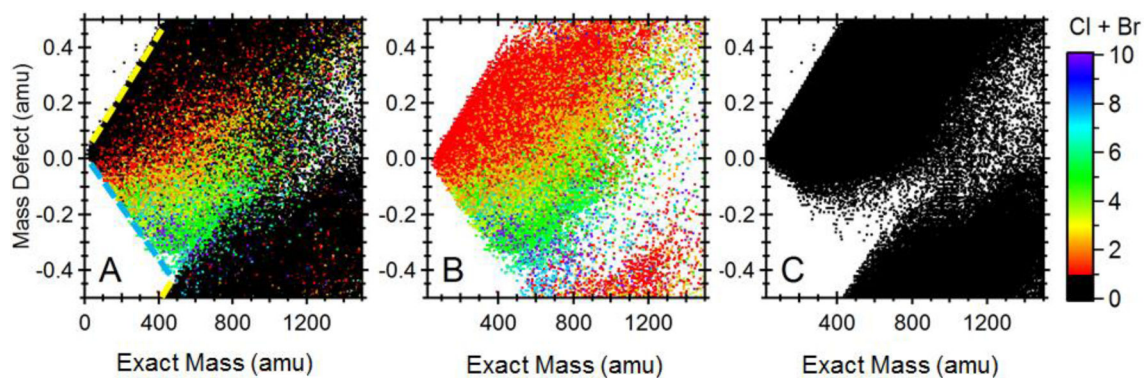


Figure 2. Mass defect plot obtained from (A) all compounds in the Kind and Fiehn compilation of PubChem⁴⁵; (B) only chlorinated and brominated chemicals, and (C) only non-halogenated chemicals. There is significant overlap with non-halogenated compounds and Cl+Br < 3 compounds, which makes discovering halogenated species solely based on mass defect difficult.

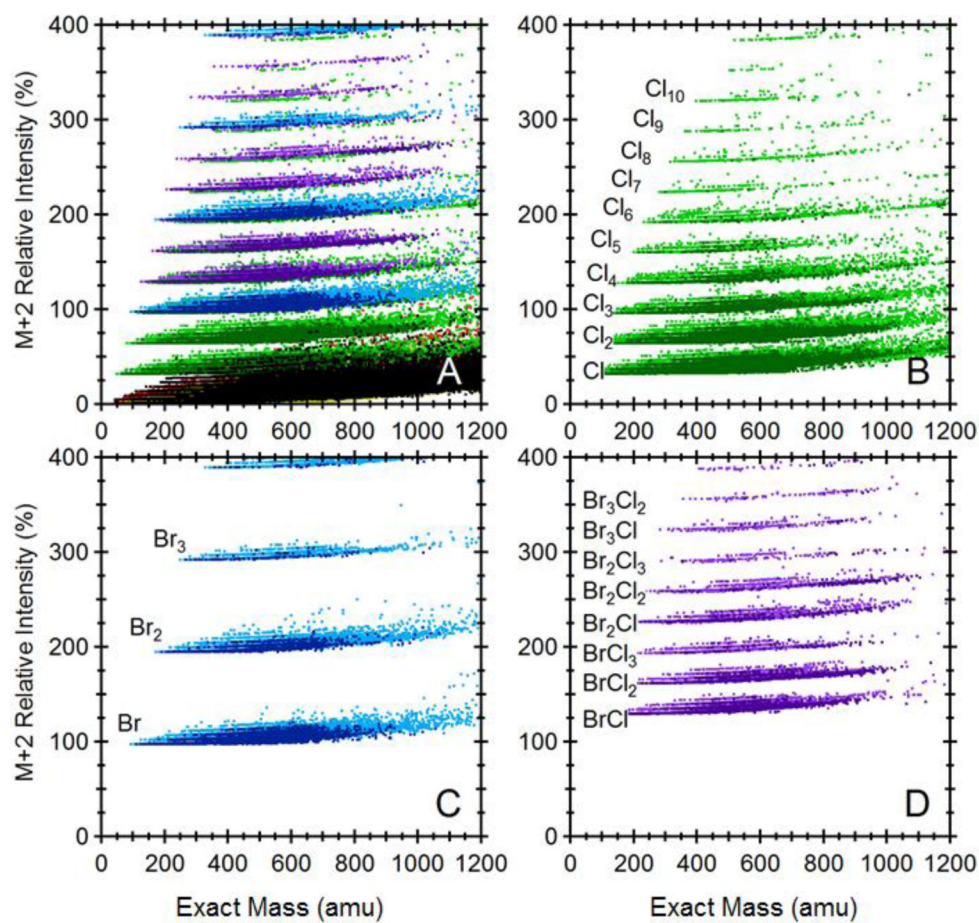


Figure 3. Isotope ratio chemical space defined by the ratio of the M+2 isotope to the monoisotopic peak, M, for (A) all compounds in the PubChem database, (B) only chlorinated compounds (green), (C) only brominated compounds (blue), and (D) mixed bromo-chloro compounds (purple). The red dots in (A) indicate Si-containing compounds.

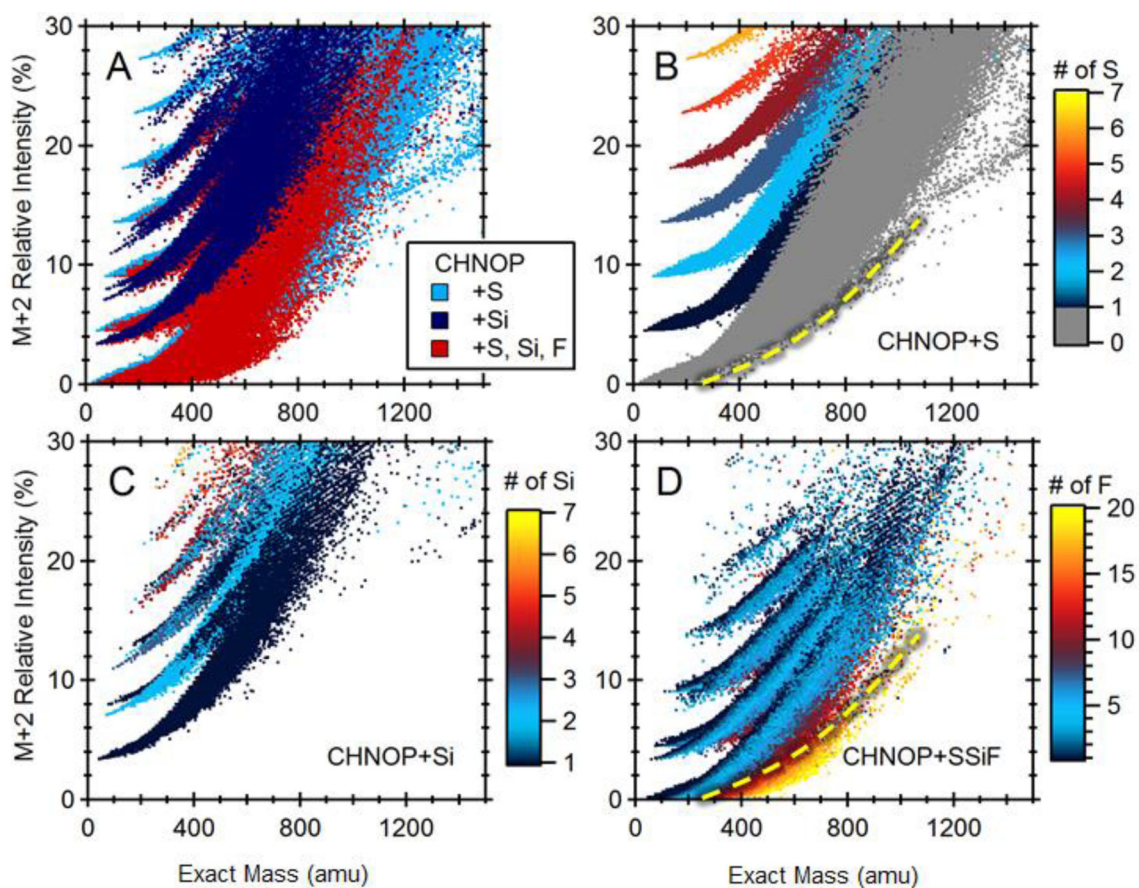


Figure 4. Fine structure of the compositional space defined by mass and isotope ratios M+2:M. (A) All non-chlorinated and -brominated compounds shown. (B) Compounds with S. (C) Compounds with Si. (D) Compounds with F.

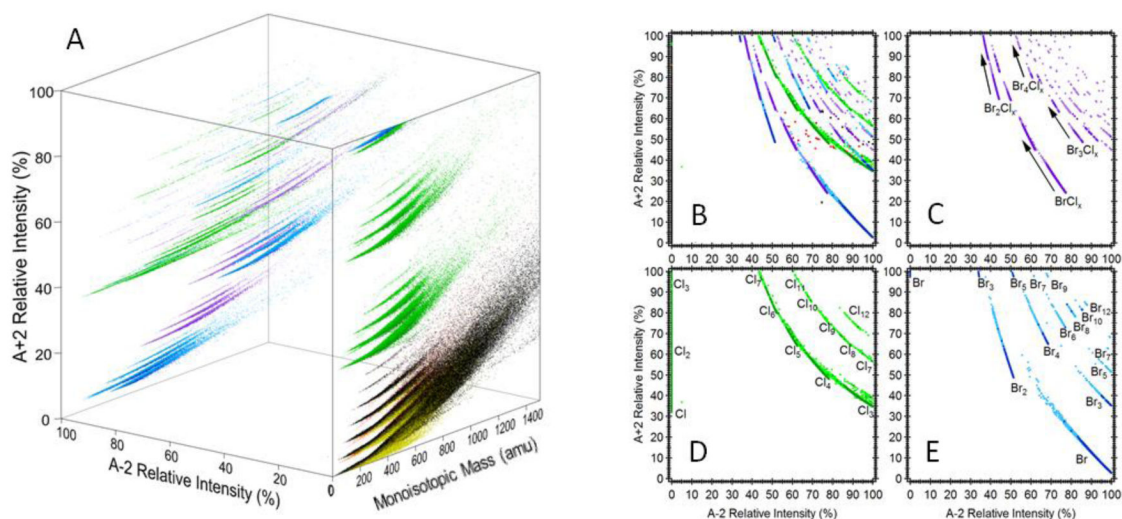


Figure 5.

(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, A-2 and A+2, 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 bromo-chloro compounds, (D) chlorinated compounds, and (E) brominated compounds;

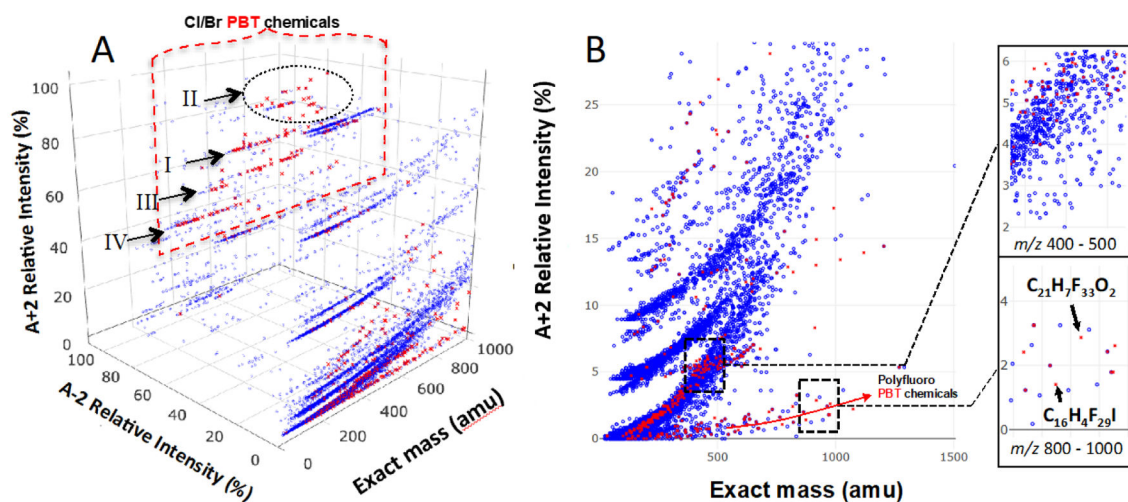


Figure 6. Distribution of 610 prioritized persistent bioaccumulative (PB) compounds (in red) and commercial chemicals (in blue) from the North American chemical inventories in the compositional spaces defined by (A) exact mass and intensity ratios (A-2):A and (A+2):A; and (B) m/z and (A+2):A. The insets of (B) display regions that are rich in non-halogenated PB compounds (top) and polyfluorinated compounds (bottom).

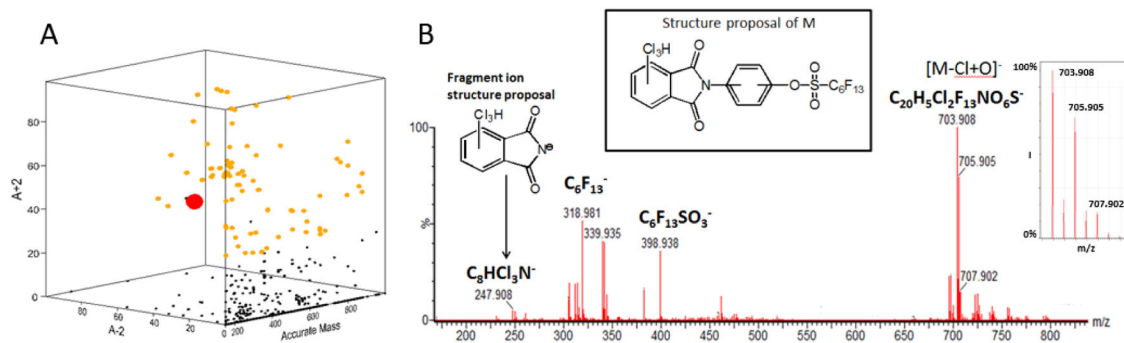


Figure 7.

(a) Approximately 18,000 chemical features (ACD/Labs MS Workbook Suite) displayed according to m/z, and the relative intensities of A+2 and A-2; 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, bioaccumulative 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.