Research Article
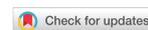
Extended characterization of the indoor dust chemical composition by retrospective suspect and non-target analysis of high resolution mass spectrometric data

Florian Dubocq[#], Thanh Wang (iD)

Man-Technology-Environment (MTM) Research Centre, Örebro University, Örebro 701 82, Sweden.
[#]Present address: University Paris Est, ANSES, Laboratory for Food Safety, 94700 Maisons-Alfort, University Paris Est Créteil, CNRS, ICMPE, UMR 7182, 2 rue Henri Dunant, Thiais 94320, France.

Correspondence to: Dr. Thanh Wang. Man-Technology-Environment (MTM) Research Centre, Örebro University, Örebro 701 82, Sweden. E-mail: thanh.wang@oru.se

Abstract

With the recent improvements in high-resolution mass spectrometry (HRMS), retrospective chemical analysis has been increasingly used in environmental sciences. This enables new insights into the chemical content of previously analysed samples with new data analysis methods or new information about emerging contaminants. This study aimed to conduct an in-depth investigation into the chemical content of various indoor dust samples using retrospective analysis. The samples were previously extracted using liquid-solid extraction without clean-up to increase the chemical coverage and thereafter analysed both using liquid chromatography (positive and negative ionisations) and gas chromatography coupled with high-resolution mass spectrometry. A retrospective data processing workflow was conducted in this new study by using both suspect screening analysis and non-target analysis. Among 30 dust samples from four different indoor settings, 298 compounds were tentatively identified with an identification confidence level of ≥ 3. The discussion was conducted on both individual compounds as well as their chemical compound groups and functional uses. Main detected chemical groups were plant natural products ($n = 57$), personal care products ($n = 44$), pharmaceuticals ($n = 44$), food additives ($n = 43$), plasticisers ($n = 43$), flame retardants ($n = 43$), colourants ($n = 42$) and pesticides ($n = 31$). Although some detected

compounds were already reported for the same samples in our previous study, this retrospective analysis enabled the tentative identification of compounds such as polyethylene glycols, per- and polyfluoroalkyl substances, pesticides, benzotriazoles, benzothiazoles, fragrances, colourants and UV stabilizers. This study showed the usefulness of retrospective analysis on indoor dust samples to further characterise the chemical content, which can help to better estimate the exposure risks of organic contaminants to humans in the indoor environment.

INTRODUCTION

Industrial chemicals are used throughout the modern society and the demand for greater functionality is a key driver for the introduction of new chemicals into the market. Some of the high production volume (HPV) chemicals, such as organophosphate esters, phthalates and parabens, have been found to cause adverse effects on human health and the environment[1,2], and new replacement chemicals have been introduced into the market. However, many of these might have unknown long-term environmental health effects. To better understand the chemical exposure risk for humans, a major obstacle is to obtain a comprehensive overview of the chemical content of various environmental matrices that are relevant for human exposure assessment. Ingestion of indoor dust could be a significant human exposure pathway to organic contaminants since we spend more than 90% of our time in various indoor microenvironments[3,4]. Although numerous studies have characterised the chemical content of indoor dust, most of these mainly focused on a preselected number of targeted compounds and thus did not fully characterise the indoor dust chemical composition[4,5].

Recent improvements in high-resolution mass spectrometry (HRMS) with faster scanning rates at higher resolution, together with novel data analysis methods, have enabled the detection of previously unknown environmental contaminants[6,7]. HRMS is often coupled with either liquid chromatography (LC) or gas chromatography (GC), and data are recorded in full-scan mode, which detects all ions within a preselected mass-to-charge ratio (m/z) range. This enables a comprehensive analysis and minimises discrimination of some specific compounds. New strategies for suspect screening analysis (SSA) and non-target analysis (NTA) using LC- or GC-HRMS have been developed to extend the coverage of chemical contaminant profiling in environmental samples without having prior knowledge of the chemical content[8,9]. SSA generally includes the comparison of detected m/z of parent ions (MS), the fragmentation pattern (MS/MS or MS2), the isotopic pattern, and in some cases, the retention time of the features (so-called suspect list) or spectral libraries.

To communicate the level of confidence for chemical identification, Schymanski *et al.* developed an LC-HRMS identification confidence scale (ICL) between 1 and 5, with Level 5 being the lowest level of confidence which only reports the m/z and Level 1 being the most confident by confirming the detected compound with its analytical standard[10]. When the analytical standard is not available, comparison of the unknown spectrum to available reference spectra in various databases can provide identification confidence level of 2a (probable structure with confirmation by library spectrum match), 2b (probable structure with confirmation by diagnostic evidence) or 3 (tentative candidate). A main challenge in SSA/NTA is identifying the thousands of features in the analysed samples at the highest possible confidence levels[6]. Some state-of-the-art SSA/NTA methods have been applied to: (a) increase the number of detected compounds as well as their identification confidence levels for various environmental/biological matrices; (b) improve analytical methods for the detection of compounds at low concentrations[11,12]; (c) set up early warning monitoring systems for contaminants by, e.g., temporal and spatial trends[7,13]; (d) detect previously

unmonitored transformation products[14,15]; (e) identify specific classes of compounds using isotopic pattern recognition or mass defect analysis[16]; (f) perform (semi-)quantification of identified compounds[17-19]; and (g) enable the retrospective analysis of previously analysed samples[20,21]. For all these reasons, SSA and NTA are now widely used for chemical analysis of environmental samples, although targeted analysis is still needed as a complementary method for unequivocal identification and more accurate quantification. However, full-scan HRMS also enables the re-analysis of data when the initial data processing workflow has been insufficient[22] or to retrospectively find newly identified or regulated compounds from other studies[23]. This becomes even more useful when the acquisition is performed using full-scan tandem mass spectrometry in data-dependent (DDA) or data-independent (DIA) analysis modes[24].

In this study, we conducted a retrospective data analysis of indoor dust samples previously analysed by Dubocq *et al.* using both LC- and GC-HRMS. The aim was to extend the list of detected and identified compounds at higher confidence identification levels[25]. The previous study reported over 600 compounds using a suspect list obtained from Rostkowski *et al.*[4]. Thus, this new retrospective analysis was conducted by comparing the m/z of the parent ion (MS) as well as the fragmentation spectrum (MS2), the isotopic pattern, and in some cases, the retention time with specific libraries such as the National Institute of Standards and Technology (NIST) MS Library for GC data and the MassBank of North America (MoNA)/Europe for LC data. Here, we demonstrate the power of retrospective analysis by re-analysing the raw data using additional strategies to discover contaminants previously overlooked and to improve the characterisation of the indoor dust chemical composition. Furthermore, the detected compounds were grouped according to their structural similarities to investigate the prevalence of various chemical groups and linked to their functional usages to elucidate potential sources in the indoor environment.

METHODS
From sample collection to chemical analysis
Dust sampling, preparation and chemical analysis were conducted as previously described by Dubocq *et al.*[25]. Briefly, 30 dust samples from four indoor settings, namely offices (O), households (H), preschools (P) and various occupational settings (IT: IT help desk department; PW: printing workshop; FS: furniture shop; ES: electronic store) were sampled in 2019 around Örebro county in Sweden using a custom-made stainless steel housing unit connected to an industrial vacuum cleaner (Alto Aero 21 Inox, Nilfisk, Sweden). Dust samples were collected on horizontal surfaces above the floor. Weights of the collected dust samples were between 33 and 100 mg, and the glass fibre filters were stored in glass Petri dishes with lids, packed in aluminium foil and thereafter stored in a -20 °C freezer at the laboratory. Samples were collected, prepared and analysed in 2019, with a detailed description provided by Dubocq *et al.*[25]. Briefly, samples were prepared using a solid-liquid extraction method adapted from Moschet *et al.*[26]. The dust was extracted first with 3 mL of a 3:1 hexane:acetone mixture and then 3 mL of acetone. Each extraction was performed by mixing dust sample and solvent followed by vortexing, sonication and centrifugation. Both extracts were combined, evaporated, filtered through a 0.2 μm PTFE filter and split in half, one fraction for LC analysis and the other one for GC analysis. The GC fraction was analysed on a GC coupled with an Orbitrap mass spectrometer (Q Exactive GC Orbitrap, Thermo Scientific, Bremen, Germany) using a DB5-MS column (L 30 m ID 0.25 mm, stationary phase 0.25 μm) in electron impact (EI, 70 eV) mode, while the LC fraction was analysed on a Waters G2-XS-QToF instrument (Milford, Massachusetts, USA) using an Acquity UPLC BEH C18 column (L 100 mm ID 2.1 mm, particles of 1.7 μm) with electrospray ionisation (ESI) mode, in both positive and negative modes. The mobile phase was 0.1% formic acid H_2O/0.1% formic acid acetonitrile for positive mode, while 1 mM NH_4F in H_2O/acetonitrile was used for negative mode. Various quality assurance and quality control (QA/QC) protocols were also implemented to investigate the robustness of the method, such as field blanks (GFF filters deployed for several minutes on the housing unit), procedural blanks (pre-burned GFF filters),

solvent blanks, NIST SRM 2585 indoor dust standard reference material and QC sample (combined aliquots from all extracts). Further information can be found in the work of Dubocq *et al.*[25].

Data pre-processing

During the initial study[25], a comparison was conducted between groups using statistical tests such as hierarchical cluster analysis (HCA) with heatmaps and dendrograms as visualising tools. Chemical identification was conducted using available standards and a comprehensive suspect list of > 600 indoor dust contaminants reported by the "Indoor Environment Substances from 2016 Collaborative Trial" from the NORMAN network[4]. However, the list only contains mass-to-charge ratio (m/z) of suspects without additional information on fragmentation, isotopic patterns or retention time, which resulted in relatively low identification levels (Level 4 or 5 according to Schymanski *et al.*[10]). To mitigate this, we re-analysed the dataset using a workflow with spectral matching against various HRMS spectral libraries that are openly available as well as an in-house library.

Openly available software was mostly used for the retrospective data analysis. The raw profile data files were first converted to centroided mzML format using MSConvert GUI (64-bit, ProteoWizard) version 3.0.19122[27]. Thereafter, the centroided LC data were processed using MS-DIAL software version 4.7[28]. The main parameters for LC positive mode were: mass slice width (0.1 Da), minimum peak height (10 000 amp), sigma window (0.5), MS/MS abundance (50 amp), adducts ($[M+H]^+$, $[M+Na]^+$, $[M+ACN+H]^+$ and $[2M+H]^+$) and alignment retention time (rt) tolerance (0.1 min). The same settings were used for LC negative mode except for adducts ($[M-H]^-$ and $[M-H_2O-H]^-$). A feature was kept when it was detected in at least 2/3 of the replicate injections. Spectral libraries from MassBank of North America and MassBank Europe in both positive and negative modes were used, which were downloaded from the MS-DIAL msp spectral database website[29]. These include organic contaminants as well as small compounds of biological relevance, such as metabolites and plant natural products. Mass tolerance was set to 0.01 Da and identification cut-off at 70% for similarity score between query and reference spectra. These thresholds were selected to include compounds with low intensity while limiting the number of false positive hits.

Centroided GC-HRMS data were also processed using MS-DIAL to allow for HRMS matching. The main parameters for GC-HRMS processing were: mass slice width (0.05 Da), minimum peak height (50,000), smoothing level (3 scans), average peak width (10 scan) and sigma window value (0.5). Suspect screening was conducted using the HRMS spectral library published by Price *et al.* combined with an in-house library using the parameters: retention index tolerance (60), m/z tolerance (0.005 Da) and EI similarity cut-off (70%)[30]. Representative spectra from the aligned peak list were also exported in msp format and queried against the NIST14 library for additional matching of compounds.

Data analysis

Data analysis was conducted in Excel and R statistical programming language. Hierarchical clustering analysis and visualisation were conducted using the heatmaply package in R (v1.3). Output data from MS-DIAL for all three ionisation modes (GC, LC$^+$ and LC$^-$) were used. For GC data, the area of the quantification ion was used to represent each detected compound, whereas MS1 data were used for LC runs to represent precursor ions (mainly $[M+H]^+$ and $[M-H]^-$). The intensities were scaled on the row direction (i.e., features/components) to normalise intensities and enable comparison between different sites as well as between the three different detection methods (GC, LC$^+$ and LC$^-$). The Euclidean distance metrics and average linkage criterion were used for hierarchical clustering.

Chemical structure classification was conducted using Classyfire[31]. The identified compounds were classified according to Superclass, Class, Subclass and Parent Levels 1-4. Chemical classification of the individual compounds was visualised using the R package sunburstR (v2.1.6).

Information about the functional uses of the detected compounds was collected from various sources such as Pubchem, the EPA CompTox Chemicals Dashboard and various online sources such as chemical trading platforms. A network graph (visNetwork v2.1 package in R) was generated to visualise the potential linkage of individual compounds to chemical groups and functional uses. The source code for the visualisations can be found in the Supplementary Materials.

RESULTS

Detected compounds

Retrospective analysis of the 30 indoor dust samples led to the tentative identification of 298 components at a ICL of ≥ 3. Some components were detected at multiple retention times, which suggested the presence of several isomers. Since standards were not available for retention time confirmation, all isomers are reported in the Supplementary Materials. However, during the discussion on chemical grouping, all isomers were considered as one compound (i.e., having the same SMILES or InChI). Figure 1 summarises the number of detected compounds with regard to the different ionisation modes.

Detection was conducted by combining SSA and NTA for both GC and LC analysis. For 46 compounds, chemical standards were available and thus were confirmed with an ICL of 1. Moreover, 200 compounds were tentatively identified with an ICL of 2 (156 at ICL of 2a and 44 at ICL of 2b) and 67 with an ICL of 3. Some of the compounds were reported at different ICLs when detected in GC and LC; thus, the sum of reported confidence levels (313) does not match the number of reported unique compounds (298). This was because some chemical standards were only compatible with only one of the injection methods (difference between Levels 1 and 2/3) or because more information (e.g., mass spectrum or retention index) was available for only one of the two methods (difference between Levels 2 and 3).

For LC-ESI, 8625 and 13,211 features were detected for all samples combined in negative and positive mode, respectively. For GC-EI mode, the number of deconvoluted components was 4741. The tentatively identified compounds for EI, LC positive mode (LC$^+$) and LC negative mode (LC$^-$) were 162, 78 and 85, respectively. The relatively high number of detected compounds in negative mode compared to positive mode was somewhat different from most NTA studies. This difference could be due to the use of different mobile phases between the two modes in our study, as well as the use of different spectral libraries for suspect screening.

The Supplementary Materials include a list of all detected compounds as well as a heatmap [Supplementary Figure 1] with dendrograms with their normalised intensities (higher intensities are coloured in red in the heatmap). Some contaminants, such as di(propylene glycol) dibenzoate (DPGDB), tributyl citrate (TBC), acetyl tributyl citrate (ATBC) and a polychlorobiphenyl congener (PCB-11), were frequently detected in the samples. Moreover, tetrabromobisphenol A (TBBPA), which is mainly used as a reactive flame retardant in printed circuit boards, was only detected in the electronic store (ES) sample. Some other compounds, such as trichlorobenzene, dichloroaniline, trichloroaniline, phthalates, toluene diisocyanates (2,4- and 2,6-TDI), personal care products and pharmaceuticals (triclocarban and diclofenac), were previously detected in the same samples reported by Dubocq *et al.*[25]. However, additional compounds were retrospectively detected, such as the pesticide fipronil (LC ESI-, IDL 2a), which was mainly detected in the H4 sample. The intensities of fipronil were also highly correlated (R = 0.94) with one of its metabolites, fipronil sulfone
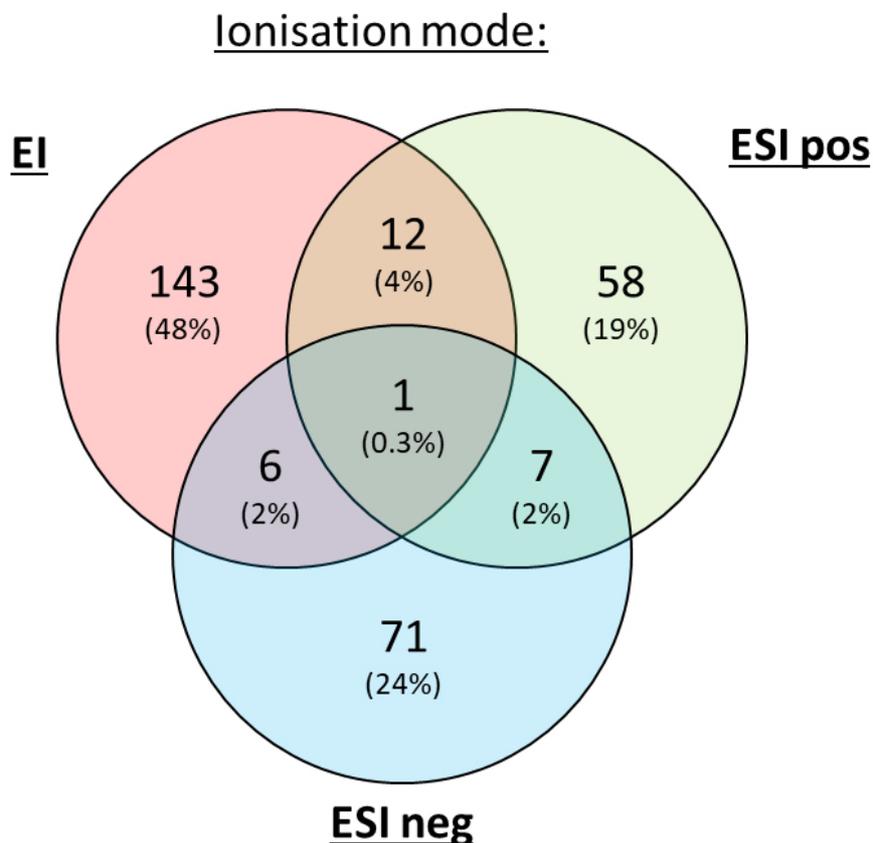
Figure 1. Venn diagram of detected unique compounds (at least Level 3). EI refers to GC-HRMS, whereas ESI refers to LC-HRMS in either positive or negative modes. EI: Electron impact; ESI: electrospray ionization; GC: gas chromatography; HRMS: high-resolution mass spectrometry; LC: liquid chromatography.

(LC ESI-, IDL 2a). Other detected pesticides and biocides were DEET, DDE, chlorpyrifos, permethrin, fludioxonil, dinoseb and terbutryn. These pesticides have the highest intensities in the NIST reference sample as well as in the household (H) samples. Many polyethylene glycols (PEGs) were also tentatively identified and were detected in most of the samples. Six benzothiazoles (1,3-benzothiazole, 2-(methylmercapto)benzothiazole, 2-benzothiazolesulfonic acid, benzisothiazolinone, 2-hydroxybenzothiazole and 2,8-dimethyldibenzothiophene) and five benzotriazoles (drometrizole, bumetrizole, octrizole, 1Hbenzotriazole and 1-(4-chlorophenyl)-1H-1,2,3-triazol-4-ylmethanol) were also detected, as well as several per- and polyfluoroalkyl substances (PFAS) including PFOS, PFUnDA, N-EtFOSAA, PFDA, PFDS, PFHxS, PFHpS, 8:2 FTSA, 6:2 FTSA and FTNA. Benzothiazoles and benzotriazoles were detected in all dust categories with lower intensities in H dust samples, while PFAS were detected with the highest intensity in the NIST sample and various samples such as H2 and FS (PFDA) and P6 (PFHxA). Other identified compound groups were: (a) ultraviolet (UV) stabilisers (octrizole, octocrylene and UV-328, in all dust categories); (b) organophosphate esters (TCP, TDCIPP and TCEP, in all dust categories with the highest intensity in samples PW and H3); (c) psychotropic substances (caffeine in O1 and P1, nicotine and cotinine in H3 and cocaine in NIST sample); (d) personal care products (parabens, triclosan and chlorocresol, in P and H samples with chlorocresol having highest intensity in FS sample); (e) pharmaceuticals (acetaminophen with the highest intensity in O1 sample); (f) dyes (C.I Disperse Red 60, C.I. Basic Red 9 and crystal violet, in P and O samples); (g) isocyanates (MDI and TDI in NIST, H3 and PW samples); (h) fragrance compounds (rosacetol, chloroatranol, chloroatranorin, galaxolide

and tonalide, in NIST, H, P and O dust categories); and (i) antioxidant for plastics (BHT, in all dust categories). Some compounds, such as some PCB congeners and some pesticides (DDE and DDT), were detected in the NIST dust sample, and the presence of these compounds was confirmed in the NIST certificate reporting the SRM chemical composition[32]. The detection and tentative identification of all these compounds can help to understand the extent of chemical contamination in indoor dust to further elucidate the exposure risk for humans.

Grouping and classification of compounds and samples

Detected compounds were classified based on their structural similarity into different chemical groups using Classyfire[31]. Figure 2 shows a sunburst chart of the detected compounds (n = 297, as one compound could not be classified). An interactive HTML version of this chart is available in the Supplementary Figure 2. The chart shows the grouping of the detected compounds into chemical classes and thus facilitates the overview of structurally similar compounds. The five main detected superclasses were benzenoids (38% of the detected compounds), phenylpropanoids and polyketides (14%), organoheterocyclic compounds (11%), lipids and lipid-like molecules (11%) and organic acids and derivatives (10%).

To investigate the relationships between chemical groups and reported functional usages, a network visualisation was conducted, as shown in Figure 3. This network graph shows the relationship between the detected compounds (small blue nodes in Figure 3) and their compound classes (yellow circles), followed by their inter-relationship with their functional uses (red circles). Some compounds can have multiple functional uses or use categories, and the network graph provides a useful overview of the complex relationships between compounds and their potential sources (e.g., some benzoic acid esters can be used both as plasticisers and in personal care products). Detected compounds are mainly plant natural products, pharmaceuticals, food additives, personal care products, colourants, plasticisers and flame retardants [Supplementary Figure 3].

DISCUSSION

The detection and tentative identification of compounds in the indoor dust samples are important to understand their sources in the indoor environment. Some reasons were already discussed for some detected compounds by Dubocq *et al.*, and therefore the following discussion emphasises more on newly identified compounds, the chemical classification and source apportionment[25].

Detected compounds

The presence of fipronil and its degradation products such as fipronil sulfone in dust samples was previously demonstrated by Mahler *et al.*[33]. Fipronil is widely used as flea-and-tick control for dogs and cats and thus can be ubiquitous in some indoor dust samples. These two compounds had the highest relative intensities in a house dust sample which has a pet cat (H4). This also implies the importance of investigating potential degradation products of detected parent compounds, and future SSA/NTA should include more transformation products into their workflows.

The detection of pesticides in indoor dust has already been reported, and their concentrations were usually higher around agricultural areas[34]. Detected compounds such as DEET, permethrin and terbutryn are used as repellent for mosquitos and pests in homes and were detected with the highest intensities in the NIST sample, but they were also detected in some household dust (such as H2 and H4)[34]. A class of ubiquitous contaminants, organophosphate ester flame retardants, was detected in most of the analysed dust samples. These are common contaminants in indoor environments and include, e.g., TCPP, TDCPP and TCEP[35].
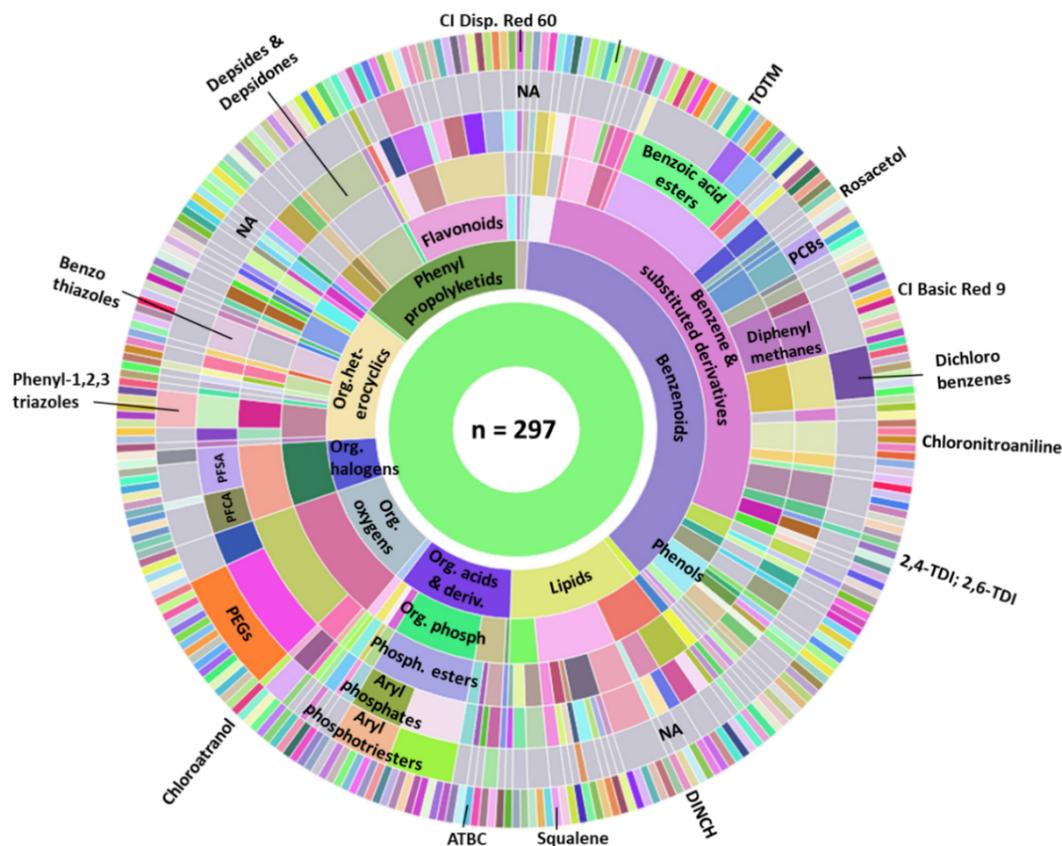
Figure 2. Sunburst chart with grouping according to the levels: Superclass > Class > Subclass > Parent Level 1 > Parent Level 2 > Common Name. The identification level of the compounds ranged from 1 to 3. Some selected compounds have been annotated in the figure. An interactive HTML version of this chart is available in the Supplementary Figure 2. ATBC: Acetyl tributyl citrate; PCBs: polychlorobiphenyls; PEGs: polyethylene glycols; TDI: toluene diisocyanates.

Several fragrance chemicals (rosacetol, chloroatranol, chloroatranorin, galaxolide and tonalide) were tentatively identified in dust samples since they can be used in many different products and thus can be sorbed to indoor dust. Chlorantranol was recently banned from use in cosmetics, and thus the occurrence of this substance and its transformation products will likely decrease with time[36]. Other detected compounds, benzothiazoles and benzotriazoles, are widely used and produced as organic corrosion inhibitors for metals as well as UV light filters in plastics and polymers and have also been previously detected in indoor environments[37]. Many other detected compounds are representative of daily usage and could be used as chemical fingerprints of the indoor environment[26]. As an example, the detection of caffeine or nicotine can be related to coffee drinkers and smokers, respectively. Moreover, usage of personal care products can lead to the release of various compounds such as parabens, triclosan and polyethylene glycols (PEGs) that can be sorbed to indoor dust. This demonstrates the need to use suspect screening and non-target analysis for the detection and identification of chemicals in the dust samples since the dust composition can vary with location and environment. Conversely, some target compounds from the previous study[25] could not be detected using the suspect and non-target screening method, which stresses the importance of combined screening strategies.

Grouping and classification of compounds and samples

The sunburst chart in Figure 2 provides a general overview of compound groups according to their chemical structure. The main detected classes were PEGs (from PEG_8 to PEG_{20}), PFAS (both sulfonic and

Figure 3. Network visualisation mapping the relationships of individual compounds (small blue nodes) to their Parent Level 1 chemical groups (blue edges/lines) and then from chemical groups to their functional uses (orange edges/lines). Names of selected compound groups (yellow nodes) and use categories (red nodes) are displayed in the graph (the relative node sizes are displayed in the inset). The functional uses or use categories include intentional uses, intermediate chemicals and potential transformation compounds from the chemicals used in the products.

carboxylic acids from C_4 to C_{11}), benzoic acid esters (parabens, phthalates and UV-stabilisers) and organophosphate esters (e.g., TPP, TCEP, TEP and TEHP). As mentioned above, PEGs have huge commercial and industrial uses[38] and thus are ubiquitous in both outdoor and indoor environments. Benzoic acid esters mainly consist of phthalates and parabens and have previously been reported in indoor dust samples[39]. As shown in the network graphs [Figure 3 and Supplementary Figures 4 and 5], these compounds are mainly used in personal care products and as plasticisers. They are also ubiquitous in the indoor environment, although newly produced products tend to decrease the use of such compounds due to environmental concerns. The third main detected group is organophosphate esters, which includes various flame retardants and plasticisers and are also ubiquitously detected in indoor dust samples[4]. Another important detected group is per- and polyfluoroalkyl substances (PFAS), which are chemicals with a fully or partially fluorinated backbone. These compounds have been widely used in various products and applications but can be very persistent in the environment. We grouped PFAS substances separately from others due to their special properties and functional usage, and because they are used in a wide range of applications (firefighting foams, kitchen utensils, apparels, *etc.*).

Relevance of detected compounds and structural classes

Source apportionment

The network graph [Figure 3] allows for visual inspection of the predominant sources of identified compounds grouped by their chemical class and could be helpful for source apportionment and exposure assessment. This study mainly focused on the functional uses of the chemicals as it is somewhat difficult to find information about all their usages in products or materials, and many compounds also have multiple

usages. This would also allow for a better understanding of the potential sources of the dust samples in the indoor environment. For some compounds, information on potential functional usage could not be found, and these were categorised as "unknown". Grouping the individual compounds to their parent levels using Classyfire and then linking the chemical groups to the functional uses could facilitate visual inspection of the linkage between the chemical structures and functional uses. For example, benzothiazoles have been found in influents of wastewater treatment plants (WWTP) from household wastewater, with benzothiazole-2-sulfonic acid (BTSA) having the highest concentrations[40]. Since they have been detected in WWTP from household water, detection in household dust samples could also be expected. Other compounds are known to be used in everyday life such as in personal care products and pharmaceuticals and thus are also expected to be found in the indoor environment[4]. UV stabilisers could be leached from plastic packaging and polymers and are also found in sunscreens[41]. Food additives, also used in food commodities, can be detected in the indoor environment for the same reason. The detection of compounds directly or indirectly used in colourants is also interesting. Some of these are also chlorinated, such as chloroanilines and PCB-11. In general, most of the compounds detected in dust are used in indoor environments and then are transferred into dust through different processes. Some other compounds, such as PCBs, have been frequently detected in dust samples, even though they have been banned for many years[42], but the presence of PCB-11 is likely associated with their unintentional presence in pigments[43].

In conclusion, the retrospective suspect screening analysis and non-target analysis of indoor dust samples enabled the detection of almost 300 compounds with higher confidence levels compared to the initial analysis. Various functional use groups, such as pesticides, organophosphate flame retardants, personal care products and colourants, were found in dust samples. Complementary visualising tools (sunburst chart, heatmap, dendrogram, network graph and Venn diagram) can facilitate the inspection of the overall chemical composition of indoor dust samples and their potential sources. Our retrospective analysis demonstrated the importance of analysing environmental samples using full-scan HRMS as well as performing comprehensive data analysis to extend the chemical coverage of environmental samples. Some identified compounds could be of interest to investigate in more detail, such as the 2,4- and 2,6-TDI isomers, since their technical products are used to produce PUF that are widely used in the built environment. Compounds used directly or as intermediates in the production of colourants could also be further investigated due to their prevalence in building materials and consumer products. Furthermore, fragrance components such as chloroatrinorin and chloroatranol are very potent contact allergens, but their distribution in indoor environments has not been investigated in detail. The main drawback of this retrospective analysis is the limited quantitation of detected compounds due to the lack of available standards and time. The proper quantitative analysis would also require calculating the recovery of the analytes during the extraction process and could be difficult unless the original samples are still available for re-analysis. Future quantitative analysis on prioritised compounds from the retrospective analysis could be conducted to increase knowledge on the human exposure risks of indoor contaminants.

DECLARATIONS

Authors' contributions
Experimental design, chemical analysis, data analysis, interpretation and writing: Dubocq F
Conceptualisation, experimental design, funding acquisition, data analysis, interpretation and writing:
Wang T

Availability of data and materials

Additional data can be found in the Supplementary Materials.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. Harbers JV, Huijbregts MA, Posthuma L, Van de Meent D. Estimating the impact of high-production-volume chemicals on remote ecosystems by toxic pressure calculation. *Environ Sci Technol* 2006;40:1573-80. DOI PubMed
2. Núñez M, Fontanals N, Borrull F, Marcé RM. Multiresidue analytical method for high production volume chemicals in dust samples, occurrence and human exposure assessment. *Chemosphere* 2022;301:134639. DOI PubMed
3. Schweizer C, Edwards RD, Bayer-Oglesby L, et al. Indoor time-microenvironment-activity patterns in seven regions of Europe. *J Expo Sci Environ Epidemiol* 2007;17:170-81. DOI PubMed
4. Rostkowski P, Haglund P, Aalizadeh R, et al. The strength in numbers: comprehensive characterization of house dust using complementary mass spectrometric techniques. *Anal Bioanal Chem* 2019;411:1957-77. DOI PubMed PMC
5. Lucattini L, Poma G, Covaci A, de Boer J, Lamoree MH, Leonards PEG. A review of semi-volatile organic compounds (SVOCs) in the indoor environment: occurrence in consumer products, indoor air and dust. *Chemosphere* 2018;201:466-82. DOI PubMed
6. 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;51:11505-12. DOI PubMed
7. Alygizakis NA, Samanipour S, Hollender J, et al. Exploring the potential of a global emerging contaminant early warning network through the use of retrospective suspect screening with high-resolution mass spectrometry. *Environ Sci Technol* 2018;52:5135-44. DOI PubMed
8. Gago-ferrero P, Schymanski E, Hollender J, Thomaidis N. Nontarget Analysis of Environmental Samples Based on Liquid Chromatography Coupled to High Resolution Mass Spectrometry (LC-HRMS). *Compr Anal Chem* 2016;71:381-403. DOI
9. Hakme E, Lozano A, Gómez-Ramos MM, Hernando MD, Fernández-Alba AR. Non-target evaluation of contaminants in honey bees and pollen samples by gas chromatography time-of-flight mass spectrometry. *Chemosphere* 2017;184:1310-9. DOI PubMed
10. Schymanski EL, Jeon J, Gulde R, et al. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* 2014;48:2097-8. DOI PubMed
11. Gómez-Ramos MM, Ucles S, Ferrer C, Fernández-Alba AR, Hernando MD. Exploration of environmental contaminants in honeybees using GC-TOF-MS and GC-Orbitrap-MS. *Sci Total Environ* 2019;647:232-44. DOI PubMed
12. Müller A, Schulz W, Ruck WK, Weber WH. A new approach to data evaluation in the non-target screening of organic trace substances in water analysis. *Chemosphere* 2011;85:1211-9. DOI PubMed
13. Gago-Ferrero P, Schymanski EL, Bletsou AA, et al. Extended suspect and non-target strategies to characterize emerging polar organic contaminants in raw wastewater with LC-HRMS/MS. *Environ Sci Technol* 2015;49:12333-41. DOI PubMed
14. López A, Dualde P, Yusà V, Coscollà C. Retrospective analysis of pesticide metabolites in urine using liquid chromatography coupled to high-resolution mass spectrometry. *Talanta* 2016;160:547-55. DOI PubMed
15. Martínez-Piernas AB, Plaza-Bolaños P, Agüera A. Assessment of the presence of transformation products of pharmaceuticals in agricultural environments irrigated with reclaimed water by wide-scope LC-QTOF-MS suspect screening. *J Hazard Mater* 2021;412:125080. DOI PubMed
16. Cariou R, Omer E, Léon 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;936:130-8. DOI

PubMed

17. Kruve A. Semi-quantitative non-target analysis of water with liquid chromatography/high-resolution mass spectrometry: how far are we? *Rapid Commun Mass Spectrom* 2019;33 Suppl 3:54-63. DOI PubMed

18. Aalizadeh R, Nikolopoulou V, Alygizakis N, Slobodnik J, Thomaidis NS. A novel workflow for semi-quantification of emerging contaminants in environmental samples analyzed by LC-HRMS. *Anal Bioanal Chem* 2022;414:7435-50. DOI PubMed

19. Alygizakis NA, Oswald P, Thomaidis NS, et al. NORMAN digital sample freezing platform: a European virtual platform to exchange liquid chromatography high resolution-mass spectrometry data and screen suspects in "digitally frozen" environmental samples. *Trac-trend Anal Chem* 2019;115:129-37. DOI

20. Günthardt BF, Wettstein FE, Hollender J, et al. Retrospective hrms screening and dedicated target analysis reveal a wide exposure to pyrrolizidine alkaloids in small streams. *Environ Sci Technol* 2021;55:1036-44. DOI PubMed

21. Creusot N, Casado-Martinez C, Chiaia-Hernandez A, et al. Retrospective screening of high-resolution mass spectrometry archived digital samples can improve environmental risk assessment of emerging contaminants: a case study on antifungal azoles. *Environ Int* 2020;139:105708. DOI PubMed

22. Hernández F, Ibáñez M, Gracia-Lor E, Sancho JV. Retrospective LC-QTOF-MS analysis searching for pharmaceutical metabolites in urban wastewater. *J Sep Sci* 2011;34:3517-26. DOI PubMed

23. Fels H, Herzog J, Skopp G, et al. Retrospective analysis of new psychoactive substances in blood samples of German drivers suspected of driving under the influence of drugs. *Drug Test Anal* 2020;12:1470-6. DOI PubMed

24. Guo J, Huan T. Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography-mass spectrometry based untargeted metabolomics. *Anal Chem* 2020;92:8072-80. DOI PubMed

25. Dubocq F, Kärrman A, Gustavsson J, Wang T. Comprehensive chemical characterization of indoor dust by target, suspect screening and nontarget analysis using LC-HRMS and GC-HRMS. *Environ Pollut* 2021;276:116701. DOI PubMed

26. Moschet C, Anumol T, Lew BM, Bennett DH, Young TM. Household dust as a repository of chemical accumulation: new insights from a comprehensive high-resolution mass spectrometric study. *Environ Sci Technol* 2018;52:2878-87. DOI PubMed PMC

27. Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol* 2012;30:918-20. DOI PubMed PMC

28. Tsugawa H, Cajka T, Kind T, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* 2015;12:523-6. DOI PubMed PMC

29. MS-DIAL. MS-DIAL metabolomics MSP spectral kit containing EI-MS, MS/MS, and CCS values. Available from: http://prime.psc.riken.jp/compms/msdial/main.html#MSP [Last accessed on 27 Sep 2022].

30. Price EJ, Palát J, Coufaliková K, et al. Open, high-resolution ei+ spectral library of anthropogenic compounds. *Front Public Health* 2021;9:622558. DOI PubMed PMC

31. Djoumbou Feunang Y, Eisner R, Knox C, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform* 2016;8:61. DOI PubMed PMC

32. NIST. Certificate of analysis, standard reference material 2585: Organic Contaminants in House Dust. 2018. Available from: https://www-s.nist.gov/srmors/view_cert.cfm?srm=2585 [Last accessed on 27 Sep 2022].

33. Mahler BJ, Van Metre PC, Wilson JT, Musgrove M, Zaugg SD, Burkhardt MR. Fipronil and its degradates in indoor and outdoor dust. *Environ Sci Technol* 2009;43:5665-70. DOI PubMed

34. Harnly ME, Bradman A, Nishioka M, et al. Pesticides in dust from homes in an agricultural area. *Environ Sci Technol* 2009;43:8767-74. DOI PubMed

35. Chupeau Z, Bonvallot N, Mercier F, Le Bot B, Chevrier C, Glorennec P. Organophosphorus flame retardants: a global review of indoor contamination and human exposure in europe and epidemiological evidence. *Int J Environ Res Public Health* 2020;17:6713. DOI PubMed PMC

36. EU. Commission Regulation (EU) 2017/1410 of 2 August 2017 amending Annexes II and III to Regulation (EC) No 1223/2009 of the European Parliament and of the Council on cosmetic products. 2017. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32017R1410&from=EN [Last accessed on 27 Sep 2022].

37. Wang L, Asimakopoulos AG, Moon HB, Nakata H, Kannan K. Benzotriazole, benzothiazole, and benzophenone compounds in indoor dust from the United States and East Asian countries. *Environ Sci Technol* 2013;47:4752-9. DOI PubMed

38. Fruijtier-Pölloth C. Safety assessment on polyethylene glycols (PEGs) and their derivatives as used in cosmetic products. *Toxicology* 2005;214:1-38. DOI PubMed

39. Tran TM, Minh TB, Kumosani TA, Kannan K. Occurrence of phthalate diesters (phthalates), p-hydroxybenzoic acid esters (parabens), bisphenol A diglycidyl ether (BADGE) and their derivatives in indoor dust from Vietnam: implications for exposure. *Chemosphere* 2016;144:1553-9. DOI PubMed

40. Kloepfer A, Jekel M, Reemtsma T. Occurrence, sources, and fate of benzothiazoles in municipal wastewater treatment plants. *Environ Sci Technol* 2005;39:3792-8. DOI PubMed

41. Liu R, Ruan T, Wang T, Song S, Guo F, Jiang G. Determination of nine benzotriazole UV stabilizers in environmental water samples by automated on-line solid phase extraction coupled with high-performance liquid chromatography-tandem mass spectrometry. *Talanta* 2014;120:158-66. DOI PubMed

42. Abb M, Breuer JV, Zeitz C, Lorenz W. Analysis of pesticides and PCBs in waste wood and house dust. *Chemosphere* 2010;81:488-93. DOI PubMed

43. Shang H, Li Y, Wang T, et al. The presence of polychlorinated biphenyls in yellow pigment products in China with emphasis on 3,3'-dichlorobiphenyl (PCB 11). *Chemosphere* 2014;98:44-50. DOI PubMed