Cod: 2.1015

NON-TARGETED SCREENING AS A TOOL FOR THE IDENTIFICATION OF CHEMICALS IN DUST AND PROFILING OF THE DIFFERENCES BETWEEN INDOOR ENVIRONMENTS

A.C. Ionas[1], P. Booij[1], L. Melymuk[1], P. Karaskova[1], G.P. Codling[1], M. Vernier[2], M. Diamond[3], J. Klanova[1]

[1]*Research Centre for Toxic Compounds in the Environment (RECETOX), Masaryk University, Brno*

[2]*School of Public and Environmental Affairs, Indiana University, Bloomington*

[3]*Department of Earth Sciences, University of Toronto, Toronto*

Introduction

Over the last few decades, a number of correlations have been made between our daily exposure to specific compounds and adverse health effects. However, much of this is anecdotal and direct monitoring of source pathways is fundamental to our understanding. Besides diet, which is a major pathway for exposure to various contaminants, another matrix which has proven to be a particularly useful indicator of the human exposure to (potentially) harmful chemicals is dust (Mercier et al., 2011). Most chemicals which migrate from consumer products, furnishings and building materials can be found in house dust. Since the general population spends most of their time indoors, inhalation and ingestion of indoor dust can constitute a major exposure pathway to such chemicals.
The number of chemicals that we may be exposed to from dust is very high (>10,000 just in a non-polar hexane fraction, (Hilton et al., 2010)) and only a small fraction of them is being monitored in targeted multi-residue methods.
For this reason, it is important to try to identify other chemicals that we might be exposed to (through non-target screening), in order to assess their risk to human health.
The aim of this study was to therefore establish a workflow to study more polar chemicals to which the population might be exposed to in the indoor environment and to determine whether there are major differences between indoor environments from different countries.

Materials and methods

Sample collection
A number of 62 house dust samples were collected between April and August 2013 from households from the Czech Republic (Brno, n=21), Canada (Toronto, n=20) and the United States (Bloomington, n=21). The samples were collected from at least one room in each house, typically the main bedroom and/or the living room. The samples were collected using a household vacuum cleaner fitted with polyester socks (pre-cleaned by Soxhlet extraction: 8 h in acetone followed by 8 h in toluene), on the hose attachment. The largest possible area was vacuumed and the samples were wrapped in clean aluminium foil, sealed and stored at -20 °C until analysis. Information such as electronic equipment and furniture in the sampled rooms, number of occupants, and cleaning and ventilation habits were obtained from questionnaires administered to the volunteers.

Chemicals and materials

LC-MS grade methanol was purchased from Biosolve (Valkenswaard, Netherlands) and HPLC Gradient Grade water from Fisher Scientific (Loughborough, UK). Ammonium acetate (purity ≥98.0%) from Fluka Chemie (Buchs, Germany) was used as additive to the extraction solvent and LC mobile phase. 50 mL Falcon tubes were purchased from Alpha Laboratories Ltd. (Eastleigh, United Kingdom).

Sample preparation
Amounts of approximately 100 mg of sieved dust (500 µm) were submitted to 3 × 15 min cycles of ultrasound-assisted extraction, using methanol with 5 mM ammonium acetate as extraction solvent. After each cycle, the supernatants were collected in a separate Falcon tube. The resulting extracts were blown down almost to dryness and reconstituted into the starting composition of the mobile phase – a mixture of equal parts ammonium acetate in water (5 mM) and methanol. The resulting extracts were filtered through nylon membrane syringe filters (13 mm diameter and 0.45 µm pore size). Aliquots underwent a 25 × dilution, to a volume of ~0.5 mL and were then submitted to LC-QTOFMS analysis.

Instrumental analysis
The instrument employed in the present study was an Agilent 1290 UPLC coupled to a 6550 Q-TOF MS (Agilent Technologies, Palo Alto, CA, USA), with a mass accuracy <1 ppm and a resolving power of up to 40000, equipped with a dual Jetstream electrospray ionisation source. As source parameters, the gas temperature was set at 150 °C, gas flow at 11 L/min, nebuliser pressure at 40 psig, sheath gas temperature at 300 °C and sheath gas flow at 11 L/min.
A volume of 2 μL of extract was injected, and separation was achieved using a Waters Acquity BEH C18 column (100 mm × 2.1 mm i.d., 1.7 μm particle size), fitted with a 5 mm matching Vanguard pre-column, with the same characteristics. The mobile phases employed were A: 90% milliQ water/10% MeOH and B: 100% MeOH, both with 0.1 % formic acid added. The flow rate was 0.3 mL/min and a linear gradient from 10% to 100% methanol in 15 min, followed by a 7.5min hold before returning to the initial conditions (2.5 min posttime).
The instrumental parameters were as follows: the gas temperature for the source was set at 150 °C, gas flow was 11 L/min, nebuliser pressure was 40 psig, sheath gas temperature was 300 °C and the sheath gas flow was 11 L/min. The same source parameters were employed in both polarities. As scan parameters, the voltages for the capillary (Vcap), nozzle, fragmentor, skimmer 1 and octopole RF peak were 3000, 300, 380, 65 and 750 V, respectively.

Data processing
An overview of the data processing procedure is described in Figure 1. The data files were processed by batch recursive feature extraction and the resulting peaks that matched certain filter values were grouped according to retention time and accurate mass. The peak group list was further refined according to set criteria. The differences and similarities between the samples from different countries were determined by employing orthogonal projections to latent structures discriminant analysis (OPLS-DA). Tentative identification was achieved by a combination of molecular formula generation, searches in various databases (by accurate mass and formula) and confirmation by MS/MS data in Molecular Structure Correlator (MSC).

Figure 1: Data processing procedure employed in the present study, using Mass Profiler Professional, Profinder and SIMCA software programs.

Auto MS2 acquisition
Instead of the classical (MS only) acquisition, the Auto MS/MS mode was employed. This helped streamline the data processing by providing MS/MS data, which was useful in confirming database hits or molecular formulas. The downside is that in small number of cases, the ions which later proved to be of interest were not selected for further fragmentation (no MS/MS data generated) or for labile high molecular mass analytes, the molecular ion might not be visible anymore. The collision energy was calculated as a linear function of the m/z value, with a value of 5 for the slope and an offset of 2.5. The maximum number of precursors per cycle considered was 3 and the abundance thresholds were 2000 and 0.001%, absolute and relative, respectively.

Batch recursive feature extraction
The data then underwent a batch recursive feature extraction procedure, using the Agilent MassHunter Profinder B.06 software. This is an automated process, which starts with molecular features being extracted, by searching for closely eluting peaks which are then deconvoluted according to various isotope clustering rules, by using complex fuzzy-logic algorithms. Next, the initial features are aligned by retention time and mass across the sample data, creating a smaller list of unique features (“binning”). Then, the retention time and mass data pairs of the aligned and binned features are used as input criteria to more accurately find the features in the samples files. With the help of post-processing filters as well, all the individual ion signals for a compound are grouped together and represented as a single compound feature.

Differential analysis
An important step leading to the differential analysis is the pre-processing. Normalisation serves to minimise systematic variations such as potential differences in the sample preparation, instrumental response, or even sampling location and methodology. For this purpose, 75th percentile shift normalisation was employed. As part of this process, the values also underwent a log2 transformation to facilitate further mathematical operations involved in statistical data analysis. Baseline transformation

to median of all samples was also performed, in order to reduce the weight of very large and very small compound features on later statistical analyses.
To sieve out the less relevant compounds from the chromatograms, only the peaks present in at least 30% of the samples from one group were considered. To emphasise the compounds which are most dissimilar in between sample groups, the compounds with a maximum fold change value of 16 were further investigated.
Venn diagrams were then employed to highlight the key compounds which set the different sample groups apart, but also the common compounds.
Hierarchical clustering analysis was then employed to get a notion about the degree of dissimilarity between sample sets and dendrogram tree were used to visualise the results.

Orthogonal projections to latent structures discriminant analysis
This technique was employed to investigate the clustering patterns of the samples and to emphasise and prioritise the compounds which are most relevant for such patterns.

Tentative identification and confirmation
This part of the study starts with library and database searches, based on high resolution / accurate mass, followed by molecular formula generation so as to verify whether they correlate with the MS/MS spectra of the compounds of interest or not.

Results and discussion

Critical parameters and thresholds
To get the most out of batch recursive feature extraction procedure, it is important to optimise a number of critical parameters, such as: peak filters (signal-to-noise is recommended; alternatively peak height can also be used – optimal values vary according to sample complexity; a value of 6000 was chosen for this study); allowed ion species (for non-targeted screening, it is recommended to allow all the adducts that the software can process); isotope model (an unspecific model is recommended, such as "Unbiased" or "Common organic molecules", which was chosen for this study). An exclusion list (.csv) was obtained by running molecular feature extraction on procedural blanks. The binning and alignment tolerances (RT window and mass windows) were reduced to half of the standard values. In the minimum filter matches section, a minimum detection frequency of 30% across all of the samples files was specified. In the matching criteria, it was set to not match a compound if its score is below 50 (out of 100). The score is calculated based on mass score, isotope abundance, isotope spacing and retention time and to each of the 4 parameters a weight from 0 to 100 can be assigned.
After the software generates the list of compounds extracted with the aforementioned parameters, it is recommended for it to be visually inspected. Certain thresholds can be employed to filter out the less relevant peaks, such as: retention time - peaks eluting earlier than 1.2 min (column dead volume) can be removed. A retention time span higher than 0.05-0.06 min within a group of peaks can also indicate a bad peak shape. The same goes for compounds with relative standard deviation of target scores >15-20% or with a peak width > 1 min.

Differences between the profiles of samples from different cities/countries
Although 162 of 453 (negative mode) and 248 of 601 (positive mode) compounds were common to all dust samples (Figure 2), the variability of the other compounds was great enough to identify chemical fingerprints unique to each region of study (Figure 3). The dust from the two cities in Canada and the USA is more similar in composition as compared to the one from Brno / Czech Republic (Figure 3 & 4 – clustering in HCA).

Common chemicals in the studied countries
Among the compounds found to be similar and were tentatively identified, a few categories are to be noted, such as: fatty acids from skin (2-hydroxy palmitic acid, 15-hydroxy stearic acid, etc.) chemicals used in dermatological products (oleandolide, tetrahymanol, octyl salicylate, etc.), vitamin D derivatives and small peptides from skin.

Figure 2: (top left) Venn diagrams of compounds detected in the samples from the 3 countries; Figure 3: (bottom left) OPLS-DA score scatter plots; Figure 4: Hierarchical clustering analysis (HCA) chart -

Blue: average abundance increase / Yellow: average abundance decrease / Grey: no change; A: positive/ B: negative.

Dissimilar chemicals in the studied countries

The identification of unique and/or particularly relevant chemicals in dust from each country is still ongoing. So far, the tentatively identified chemicals which were most dissimilar in the samples from Canada, as compared to the rest of the samples were: 2-methyl-undecanoic acid, eudesmic acid and ensulizole; similarly, for the US samples we have: vanillylmandelic acid, rutin, norharman and epi-4'-hydroxyjasmonic acid and for the Czech Republic: piperine/chavicine, 11-hydroxy palmitic acid and halovir B/D. It is also important to note that the analytes detected and identified in the samples might not necessarily be representative for the whole country, but just for indoor environment in the cities where the samples were collected.

Limitations

A positive identification with absolute confidence ("gold standard") requires matching the retention time, MS and MS/MS spectra of the analyte with the reference standard of the suspected chemical (Level 1 confidence, (Schymanski et al., 2014)). Since this is not feasible to do for every analyte and sometimes MS2 spectra are not available from commercial libraries and databases, in some cases a tentative identification is the most that can be generated using this procedure. The difficulty in this case lies mainly in determining the appropriate structural isomer or the correct position of one or more functional groups.

Other important factors to consider are that in the absence of adequate libraries and databases, the process can be very time-consuming, especially if working with large data sets and especially for the confirmation of structures with MS/MS data.

Acknowledgements

This work was financially supported by the National Sustainability Programme of the Czech Ministry of Education, Youth and Sports (LO1214) and the RECETOX research infrastructure (LM2015051).

References

Hilton, D.C., Jones, R.S., Sjödin, A., 2010. A method for rapid , non-targeted screening for environmental contaminants in household dust. J. Chromatogr. A 1217, 6851–6856. doi:10.1016/j.chroma.2010.08.039

Mercier, F., Glorennec, P., Thomas, O., Bot, B. Le, 2011. Organic Contamination of Settled House Dust, A Review for Exposure Assessment Purposes. Environ. Sci. Technol. 45, 6716–6727. doi:10.1021/es200925h

Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J., 2014. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. Environ. Sci. Technol. 48, 2097–2098. doi:10.1021/es5002105

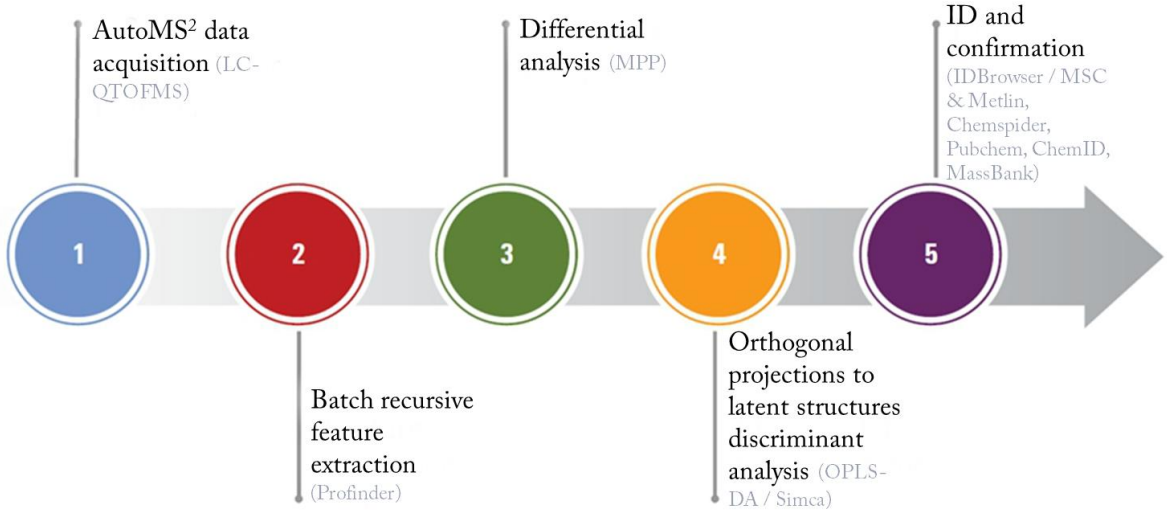

Figure 1: Data processing procedure employed in the present study, using Mass Profiler Professional, Profinder and SIMCA software programs.

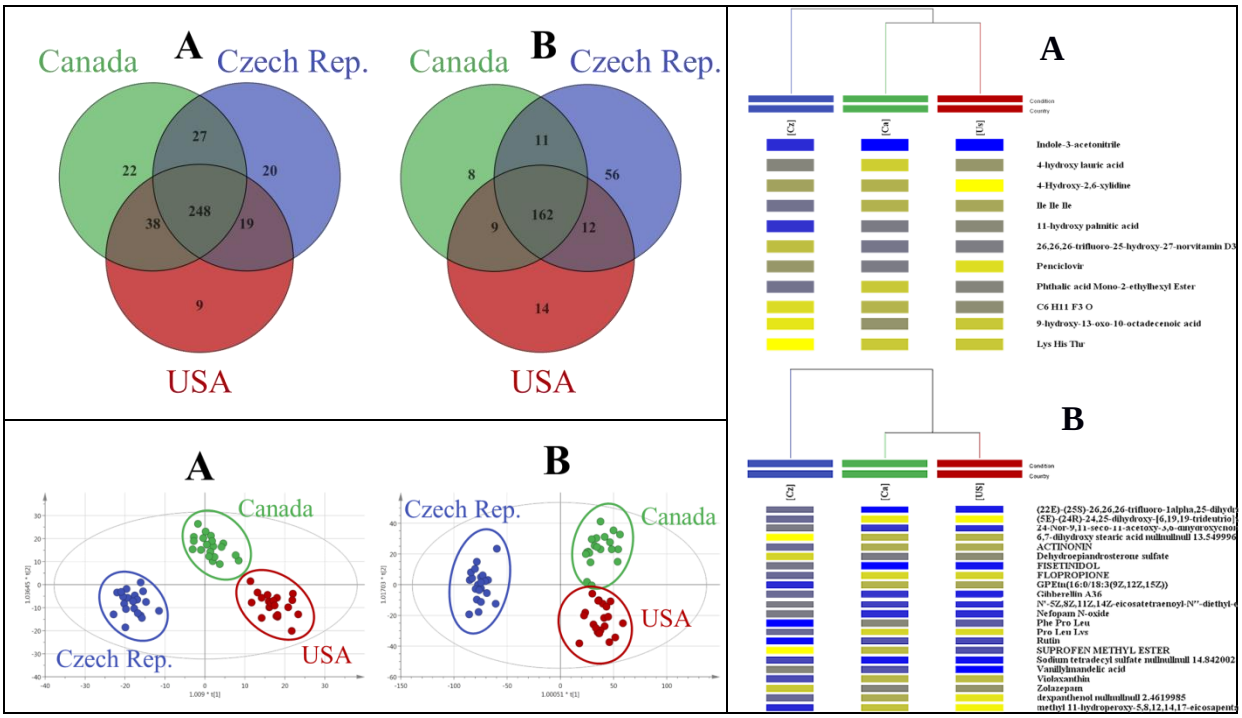

Figure 2: (top left) Venn diagrams of compounds detected in the samples from the 3 countries; **Figure 3**: (bottom left) OPLS-DA score scatter plots; **Figure 4**: Hierarchical clustering analysis (HCA) chart - ***Blue****: average abundance increase / **Yellow**: average abundance decrease / **Grey**: no change; **A**: positive/ **B**: negative.*