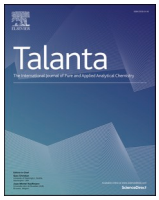
Talanta 257 (2023) 124333

Contents lists available at ScienceDirect

Talanta

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



Biological studies with comprehensive 2D-GC-HRMS screening: Exploring the human sweat volatilome

Matyas Ripszam [1,*], Tobias Bruderer [1], Denise Biagini, Silvia Ghimenti, Tommaso Lomonaco, Fabio Di Francesco

University of Pisa, Department of Chemistry and Industrial Chemistry, via Giuseppe Moruzzi 13, 56124 Pisa, Italy

ARTICLE INFO

Keywords:
Comprehensive gas chromatography
Large sample studies
Human sweat metabolites
Sweat volatilome
Dynamic headspace extraction

ABSTRACT

A key issue in GCxGC-HRMS data analysis is how to approach large-sample studies in an efficient and comprehensive way. We have developed a semi-automated data-driven workflow from identification to suspect screening, which allows highly selective monitoring of each identified chemical in a large-sample dataset. The example dataset used to illustrate the potential of the approach consisted of human sweat samples from 40 participants, including field blanks (80 samples). These samples have been collected in a Horizon 2020 project to investigate the capacity of body odour to communicate emotion and influence social behaviour. We used dynamic headspace extraction, which allows comprehensive extraction with high preconcentration capability, and has to date only been used for a few biological applications. We were able to detect a set of 326 compounds from a diverse range of chemical classes (278 identified compounds, 39 class unknowns, and 9 true unknowns). Unlike partitioning-based extraction methods, the developed method detects semi-polar ($log\ P < 2$) nitrogen and oxygen-containing compounds. However, it is unable to detect certain acids due to the pH conditions of unmodified sweat samples. We believe that our framework will open up the possibility of efficiently using GCxGC-HRMS for large-sample studies in a wide range of applications such as biological and environmental studies.

1. Introduction

There have been surprisingly few non-targeted biological studies with GCxGC-HRMS. One of the main reasons is that large-sample batch data processing of GCxGC-HRMS data is in its infancy. There are several levels of identification and verification steps needed for such a data-driven approach which were recently summarized by Stefanuto et al. [1] and Schymanski et al. [2] It is often concluded that the automation of non-target analysis and subsequent suspect-screening remains an elusive process.

This process can be intricate due to small deviations in retention times, suboptimal peak shapes, intensive background, and the diversity of chromatographic features among samples (e.g., large qualitative differences between samples), which makes the development of an efficient approach for use with large batches of samples challenging. There are automated packages available for the first step of peak alignment among a multitude of samples based on spectral similarities and retention time proximity [3–6]. Reichenbach et al. proposed cumulative chromatograms for comprehensive compound identification across multiple samples with chromatographic feature templates which can be geometrically adjusted to automatically account for these differences (e.g. a slight shift in first-dimension retention time) [7]. They showed the application of this approach to a set of 20 samples. However, batch data processing with automatic template generation above this number of samples became quickly very time-consuming, which might also explain why the suggested approach has never been used for comprehensive compound screening in biological studies. We therefore developed an approach which can handle batch sizes required for biological studies (i.e., with a minimum of 60 samples for a two-group comparison with 30 samples each).

Full automation of the whole process does not yet seem feasible due to several workflow steps which require expert knowledge and manual curation. For example, the compound identification of electron impact spectra is not expected to be replaced by machine learning in the near future due to distinct fragmentation mechanisms even for closely structurally related compounds, as well as ion rearrangements.

* Corresponding author. Present Addresses: University of Pisa, Department of Chemistry and Industrial Chemistry, via Giuseppe Moruzzi 13, 56124, Pisa, Italy.
E-mail address: matyas.ripszam@dcci.unipi.it (M. Ripszam).
[1] Shared first authors.

https://doi.org/10.1016/j.talanta.2023.124333
Received 20 October 2022; Received in revised form 25 January 2023; Accepted 6 February 2023
Available online 8 February 2023

Automated peak deconvolution and correct peak selection are limited for abundant low intensity peaks with co-elution, which do not yield high purity peak spectra. The assignment of automated peak formulae is challenging for low intensity peaks, especially when the molecular ions are not available [8]. In addition, the selection of characteristic ions across hundreds of compounds is difficult without knowing which ions can be relied upon for certain molecules.

Therefore, we have developed a method that simultaneously monitors a large set of chemicals (i.e., 326 compounds) across a large set of samples (i.e., 80 samples) with a semi-automated approach based on cumulative GCxGC chromatograms and what we call high-mass resolution extracted ion images (or in short: EIIs) and corresponding templates (i.e., 65 templates). Our framework enables new compounds to be easily added and allows to quickly adapt the data processing method to account for detrimental batch effects, or experimental artefacts (e.g., retention time shifts due to humidity issues). We believe this framework will open up the possibility to exploit GCxGC-HRMS for large-sample studies in a wide range of applications such as biological and environmental studies.

The example data set which we used to illustrate the potential of our approach contains human sweat volatiles and is part of the European Union Horizon 2020 POTION project aimed at studying the capacity of body odour to communicate emotion and influence social behaviour [9]. We used evaporative dynamic headspace extraction for this purpose, which enables comprehensive extraction with a high preconcentration capability, and has to date only been reported in three biological studies [10–12].

2. Materials and methods

2.1. Participants

We enrolled 40 healthy participants for this study, which was performed in accordance with the Declaration of Helsinki and approved by the local ethical review board at our university. Informed written consent was obtained from participants before the start of the study, whose requirements are described in Supplementary Table 1. There were 18 female and 22 male participants representing 45% and 55% of the total study population, respectively (Supplementary Table 2). The mean age of the participants was 27 ± 3 years. The study was performed in two batches of 20 participants, during the summer and following winter period.

2.2. Preparation of standards

For the method validation, a mixture of certified standard materials was used. The composition of the standard mixture is presented in Supplementary Table 3. For the internal standard spiking mixture, we used deuterium labelled standard materials (Cambridge Isotope Laboratories, Massachusetts, USA). All spiking mixtures were prepared in methanol and were added to artificial apocrine sweat (Pickering Laboratories, Mountain View, California, USA).

2.3. Sampling protocol

We used commercially available sweat pads called Dermatess (Pietrasanta Farmaceutica, Capannori, Italy), they were double layered, 10 × 10 cm made of viscose (67%) and polyester (33%). The pads and tools used for sampling were thermally pre-treated (150 °C for 20 min) and were applied to the armpits of the volunteers using head injury gauze for good flexibility. The samples were taken from the axilla of the non-dominant hand of each volunteer. The pad was applied using previously heat-treated tweezers. Sweat was collected for a 10-min period while the volunteers were presented with a virtual reality scenario of a relaxing environment which had been recently validated [13]. Afterwards, the pads were removed from the volunteers' armpits and transferred into glass containers, which were placed in a freezer at -80 °C until the analysis.

2.4. Dynamic headspace extraction

Thermal extraction was performed using a Markes μ-Chamber (Llantrisant, United Kingdom). Prior to extraction, the cuvettes of the chamber were thermally treated at 150 °C for 20 min with a nitrogen flow of 150 mL min^{-1}. Pure nitrogen (>4.0) was purified from compressed air using an LNI Swissgas NG EOLO 10 L nitrogen generator (Versoix, Switzerland). After thermal treatment, the μ-Chamber was cooled down to 35 °C, the purge flow was switched off and the sorbent tubes packed with 120 mg of Tenax-GR (poly-2,6-diphenylphenylene oxide with 23% carbon, 60/80 mesh, average particle size 0.5 μm) were placed into the outlet of the four cuvettes of the μ-Chamber. The sweat pads at this point were taken out of the freezer and immediately placed into the cuvettes of the μ-Chamber. Next, 3 μL of the internal standard mixture was spiked onto the sweat pads. The cuvettes were closed, and the flow was applied at this point for 30 s to test whether the system was leak tight. The temperature was then set at 60 °C, and after the setpoint was reached, the headspace was left to equilibrate for 15 min. Upon completion, the purge flow was applied (75 mL min^{-1}), and the extraction lasted for 20 min for a total purge volume of 1500 mL.

2.5. Thermal desorption

Thermal desorption of the sorbent tubes was performed on-line using the thermal desorption compartment of a Markes Centri multi-mode extraction and enrichment platform. The desorption process was carried out in two steps to refocus and concentrate the sample band and to interface the optimal flow rates both for thermal desorption (25–75 mL min^{-1}) and capillary chromatography (0.5–5 mL min^{-1}). This was achieved according to standard procedure by inserting a cold trap between the tube desorption and the chromatographic column. The trap material covered a range of C_4–C_{32}. A pre-purge of 2 min at 250 mL min^{-1} was used to mitigate the effect of humidity. Afterwards, the primary desorption flow was 35 mL min^{-1} at a starting temperature of 35 °C, then heated up to a final desorption temperature of 300 °C, which was held for 5 min. The in-line cold trap was kept at 2 °C throughout the tube desorption. The trap was desorbed afterwards using flash-heating to 300 °C and kept there for 7 min. The trap desorption flow was 4 mL min^{-1}, which consisted of a column flow of 0.5 mL min^{-1} and split flow of 3.5 mL min^{-1}. This split flow was passed through to a recollection tube for later use. After each sample, trap heating was performed through the split line in order to minimize the carryover (300 °C for 5 min at 50 mL min^{-1}).

2.6. GCxGC-TOF parameters

An overview of the method is presented in Fig. 1. The comprehensive two-dimensional gas chromatograph (GCxGC) was an Agilent 7890B equipped with a capillary flow technology flow modulator (Agilent Technologies, Santa Clara, California, USA). The first-dimension column configuration was a J&W DB-5MS (30 m * 0.25 mm * 0.25 μm), while the second-dimension column was a J&W DB-INNOWAX (5 m * 0.25 mm * 0.15 μm). The separation was performed in constant-flow mode. The first column volumetric flow rate was 0.5 mL min^{-1}, the second column flow rate was 18 mL min^{-1} with a modulation period of 2s. Both column flows were sub-optimal, the confining factor being the fixed volume loop of the modulator (19 μL) which limits the first column flow and therefore the sharpness of the first-dimension peaks. This resulted in at least five modulated sub-peaks for each 1D peak, which proved to be very useful in the recreation of first-dimension retention times (t_{R1}) by fitting a Gaussian-curve during data processing.

The outlet of the second-dimension column was connected to a passive splitter between two deactivated silica-capillaries. The first

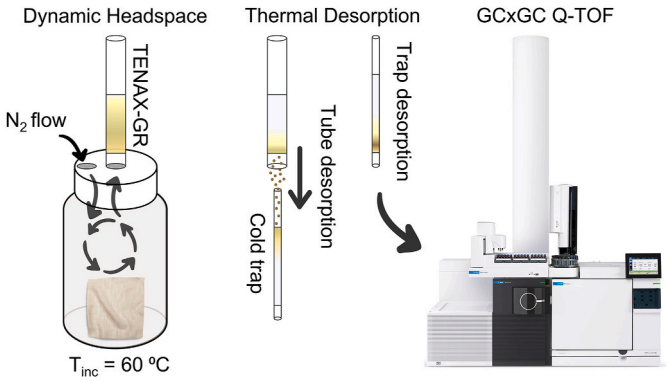


Fig. 1. The dynamic headspace, thermal desorption unit and GCxGC-QTOF MS system setup. The thermal desorption unit includes tube & trap desorption and re-collection. The GCxGC system Q-TOF uses the flow-modulation. The stationary phase of Column 1 is a DB-5MS UI, whereas Column 2 is coated with a DB-INNOWAX.

capillary (3 m * 0.18 mm) was connected to the quadrupole time-of-flight (7250 Q-TOF, Agilent Technologies) mass spectrometer, and the other capillary (0.9 m * 0.25 mm) was connected to an olfactory detection port. This resulted in a split ratio of 6:1, with less than 3 mL min^{-1} entering the mass spectrometer (maximal accepted flow for the Q-TOF).

2.7. GCxGC-HRMS data processing

The batch analysis of high resolution GCxGC-Q-TOF MS data currently poses the biggest challenge in GCxGC-Q-TOF MS data processing. We used GC Image (v.2.8. Lincoln, Nebraska, USA) for this purpose. First, a non-targeted qualitative approach was followed. The qualitative composition of a sample was examined with the high chromatographic resolution granted by the GCxGC system in combination with the high mass-resolving power of the Q-TOF. Using the power of this technique, both in terms of spectral similarity, linear retention index (LRI) comparability, and elemental formulae assignment through accurate mass measurement, the tentative identity of the chemicals can be inferred with good confidence.

3. GCxGC-HRMS data analysis for biological studies

3.1. Composite chromatograms for comprehensive compound identification

First, each chromatographic raw data file was pre-processed and saved as a total ion image (TII – a two-dimensional image based on the total ion current) in order to minimize the irrelevant information in the data. The application of a threshold filter practically eliminated the peaks on the mass spectra that originated from electronic noise or from the shift in mass peaks due to diverse effects such as detector saturation or space-charge effects in the mass analyzer. This filter (20 counts) also drastically reduced the data processing time. A further spectral filter was also applied to remove some of the peaks that were attributed to column bleed, carbon dioxide, argon, and their respective isotope peaks.

A phase shift of the second-dimension retention time (t_{R2}) was then applied to position the least retained class of organic compounds by the second column (*n*-alkanes) near the bottom of the 2D chromatogram. This is important for the easy interpretation of the physico-chemical information that is correlated with the first- and second-dimension retention times (Fig. 2). It is also important in aligning the chromatograms consistently between different batches of analyses. Differences in t_{R2} between batches may arise from different lengths of the uncoated capillary between the MS and the splitter which change the overall flow path between the MS and the modulator. Automatic peak detection was

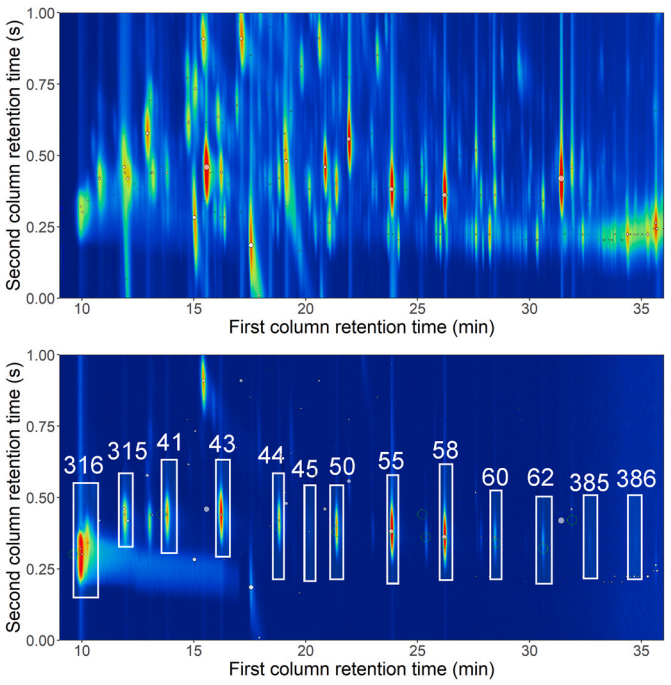


Fig. 2. The total ion image (TII - above) and the extracted ion image (EII) of *m/z* 44.0256, a characteristic ion for n-aldehydes (bottom). The rectangles for each EII template represent the retention time windows for integration of each individually monitored chemical.

performed after automatic baseline correction using a default algorithm. These TIIs were then used to automatically generate a composite chromatogram by batch reprocessing [7] (*Figure TOC*).

Automatic peak detection was performed on the composite chromatogram followed by a search in the EI spectral library of NIST (National Institute of Standards and Technology version 2017, Gaithersburg, Maryland, USA). LRI calibration was done using an *n*-alkane series (C_5–C_{20}). We used peak filters of at least 750 spectral similarity and maximum 50 LRI unit deviation from the library value (when applicable). The value of 750 may seem relatively low for a GC-MS library search. We loosened this criterion due to an inherent discrimination of TOFs for lighter ions (lower *m/z* values), which slightly decreases the spectral similarity values of the library hits. The discrimination happens because even though there is no voltage applied to the collision cell (which fragments the ions for multi-stage MS measurements), there is an automatic standby flow of 2 mL min^{-1} of nitrogen which causes ion-molecule collisions prior to the ions entering the flight tube.

Using the automated identification, only a quarter of the compounds were tentatively identified. This limited identification was due to coelutions, interferences of peak tailing, lack of linear retention index library data or having a feature with unknown compound identity. For this reason, some user supervised identification was necessary. Furthermore, elemental formulae assignment of molecular and characteristic ion peaks was not possible to perform automatically. In any case, automation reduced the time spent on identifying the major and pure peaks.

Each peak was examined using the mass spectra extraction function of the software, and these spectra were then matched with the NIST spectral library (spectral similarity >750 and ± 50 LRI units). To further strengthen the tentative identification, elemental formulae were assigned to the molecular ions when they were present. When molecular ions were not available, characteristic ions were used to classify the compound. The measure of the correct formula assignment is accuracy, which is the difference in the theoretical (m/z_{the}) and measured mass (m/z_{meas}) of the ion expressed in parts-per-million (ppm). A value below

10 ppm is considered to be an accurate match, although this threshold may depend on the intensity and the mass of the ion peak. Tentative identity was only assigned for the detected peaks once all the three key parameters matched well (spectral similarity, LRI proximity and molecular/characteristic ion formulae). The tentatively identified compounds are reported in Supplementary Table 4.

3.2. Characteristic ion templates for batch-wise data processing

After identifying and funnelling the suspected list of chemicals, we assigned characteristic ions to each one. The characteristic ions were chosen based on two criteria: a) the ion had to be selective for the individual chemical, b) the ion had to be an intensive peak on the spectrum, and at least 30% of the base peak. By selecting these characteristic ions with a narrow asymmetrical spectral window (m/z_{acc} - 0.003, +0.01), we extracted the ion images (EIIs) from the initial raw data files that were specific for either a class of chemicals (e.g., m/z 44.0256 – $C_2H_4O^{+*}$ for *n-aldehydes*) or in certain cases, for individual compounds (Fig. 2). The asymmetrical extraction window was chosen in this way due to the positive mass shift of the ions saturating the detector.

Based on the previous identification work and the prepared EIIs, a peak template was then created, which included the respective retention time windows for each compound which were drawn as rectangles or polygons on the EIIs (Fig. 2). This template was then applied to all EIIs prepared from the samples, thus resulting in the final feature matrix with all the chemicals as variables and all the samples as observations.

For all the tentatively identified unknown class and unknown chemicals, a set of 65 peak templates was generated, some including only one characteristic ion, others including up to 5 ions that did not co-elute with each other. We also successfully eliminated data processing difficulties caused by wraparound using the EIIs. Wraparound occurs when a highly retained chemical on the second column elutes in the successive modulation period [14]. This would normally cause issues with data processing but using the EIIs we were able to spectrally resolve the affected peaks. The templates were further adjusted for four different batches to account for minute retention time shifts. This means that from one raw data file, 65 SIC images were extracted. This procedure was then repeated for all the sweat samples and their respective blanks. This approach is useful because each compound can be very selectively monitored in terms of its retention times (first and second dimension) and characteristic ions. The templates and a guide for data processing are available in the Supplementary information (Supplementary Table 5). The templates can be relatively easily extended with additional compounds, while the semi-automated data processing steps could be performed by non-specialists.

3.3. Systematic compound origin assignment

Systematic origin assignment for hundreds of compounds is rarely carried out in biological studies with hundreds of chemicals due to the amount of work involved and the ambiguity with regards to labelling a compound as either as “endogenic” (known origin from the “human body”) or as “exogenic” (known “anthropogenic” compounds, i.e., created through human activity, without known origin of the human body). We developed the following workflow to address this issue. First, we investigated the compounds reported in the main human and biological databases, including the Human Metabolome Database (HMDB) [15], the Kyoto Encyclopaedia of Genes and Genomes database (KEGG) [16], and the Lipid Metabolites and Pathways Strategy database (LipidMaps) [17] (Supplementary Table 6). Matching was done by CAS identifier or compound name with the publicly available Chemical Translation Service tool (CTS) [18]. Compounds contained in these databases were individually checked and labelled with known origin from the “human body” including compounds originating from food or microbial origin. The remaining compounds were either classified as known “anthropogenic” compounds or as chemicals with “unknown” origin, by comparison with the EPA chemical dashboard from the U.S. Environmental Protection Agency (EPA), which is currently the most comprehensive public resource to label compounds as “anthropogenic” [19]. We narrowed down the most relevant databases to 18 concerning: personal care products, cleaning detergents, solvents, indoor dust, plastic packaging, and tobacco. The database search was done by CAS identifier, and if needed compound name. Of course, the compounds originating from the human body can have an anthropogenic contribution, and the currently known compounds in environmental databases are by no means comprehensive. In addition, it has to be carefully defined what “endogenic” and “exogenic” means depending on the type of study. Our main contribution here is the proposed workflow which allows origin assignment for a large number of compounds with available public databases in a timely manner. The resulting table with compound origin assignment for the identified sweat volatiles can be found in the Supplementary info (Supplementary Table 7).

Selecting the relevant environmental databases and organism identified clearly depends on the type of biological study. In our case the focus was on human sweat volatiles. A recent update regarding human volatiles was a good reference point [20]. This review contains a relatively large number of skin volatiles, including 581 compounds belonging to 26 compound classes.

3.4. Statistical analysis

Chemical data analysis was performed with R version 4.2.0 (R Foundation for Statistical Computing, Vienna, Austria) with the packages: tidyverse, ggplot2, and ggforce. Data analysis was based on the recommendations from Refs. [21,22] with the following main steps: 1.) replacement of artefacts and missing values with 1/5 of the minimum peak volume (V*s [2]) for each compound, 2.) sweat weight normalization (V*s [2]/mg), 3) log 10 transformation. Further information can be found in the data analysis info file, the data tables 1–13, and data analysis scripts 1–10.

4. Results and discussion

4.1. Detected compounds

Using a single analytical method, we were able to detect a set of 326 compounds from a diverse range of chemical classes (278 identified compounds, 39 class unknowns, and 9 true unknowns) (see Fig. 3). The developed GCxGC-TOF method with DHS at 60 °C allowed to identify compounds from most chemical classes. The highest number of compounds were either alcohols, aldehydes, hydrocarbons, esters, or ketones. We could detect a high abundance of nitrogen containing and heterocyclic oxygen containing compounds, which is probably because these generally more hydrophilic chemicals ($log\ P < 2$) are particularly

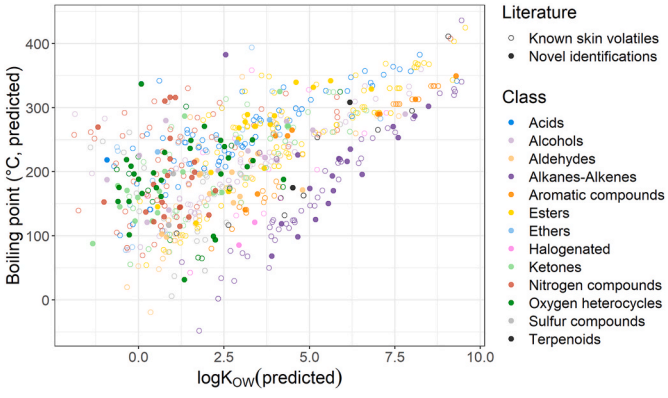


Fig. 3. Detected skin volatiles with the developed GCxGC-TOF method with DHS at 60 °C.

efficiently extracted using DHS. In contrast, most sampling methods reported in the literature have used partitioning-based sampling techniques such as liquid extraction, stir-bar sorptive extraction (SBSE) [23], sorptive tap extraction (STE) [24] or solid phase micro-extraction (SPME) [25] with less efficient extraction of nitrogen and heterocyclic compounds.

The retention of the polar chemicals has some variability, since we are using thermal desorption and despite purging the carrier gas in order to dry the sorbent tubes prior to desorption, there are still traces of humidity that enter the GC column. This results in the formation of water micro-droplets on the highly apolar inner surface of the dimethyl siloxane coated first column, due to its poor wettability by water. As a result, the hydrophilic compounds are trapped in these water droplets and are less efficiently focused on the head of the first column. However, with our strategy for data processing, we adjust the templates of these compounds individually. This increases the quality and the consistency of the resulting feature matrix before the statistical analysis is performed.

Our method's main drawback is the lack of detectability of some organic acids and of a few very volatile chemicals. In fact, the physiological pH of human skin ranges between 4.0 and 7.0, which shifts the equilibrium towards the deprotonated form of the short-chain fatty acids. This greatly reduces their ability to volatilize and thus the detectability with our method, which is proportional to their protonated fraction. This is a common issue for liquid-liquid extraction methods with decreased pH of the aqueous phase [26,27].

There is very limited literature about sweat volatiles, and even less about armpit sweat in particular. There is the review about human volatiles from healthy human breath and bodily fluids [20] and a recent review dedicated to the human skin volatilome [28]. This recent review highlights two studies which analysed armpit sweat samples with a 8 min cotton pad collection followed by static headspace TD and GC-MS analysis which could identify 23 sweat volatiles [29], respective a study with a 24 h sterile gauze collection followed by static headspace SPME and GC-MS analysis which resulted in 45 identified compounds [25].

4.2. Predicted physico-chemical properties for unknowns

The correlation between retention times and physico-chemical properties was evaluated for both the first- and second-dimension columns in relation to boiling point and log *P*, as a proxy measure to hydrophobicity. We used the OPERA predicted values for both properties. Since we used a linear oven temperature program, the boiling point – t_{R1} linear correlation proved to be very good both for the samples with the 278 tentatively identified chemicals ($R^2 = 0.90$) (Supplemental Fig. 1). For *log P* versus *log* t_{R2} relationship (Supplemental Fig. 2), there was a correlation of $R^2 = 0.54$. This effect can be explained by the selectivity of the DB-INNOWAX phase and the extended set of chemicals, where half of the dataset is not comprised of homologue series. The correlation ($R^2 = 0.90$) between the first-dimension retention times and boiling points allows us to make predictions about the boiling points of the class unknowns, unknown chemicals and tentatively identified chemicals with unknown boiling points (i.e., 65 features). This strictly indicative predicted information is included in Supplementary Table 4.

4.3. Dynamic headspace extraction

The estimated limits of detection (LOD) and quantification (LOQ) are shown in Supplementary Table 8. Our sample-amount normalized limits of detection were slightly higher than those reported for the analysis of environmental pollutants in natural water [30]. The usual dynamic range of our method was 2–8000 pg of reference standard chemicals in 1 μL of sweat (e.g., methyl salicylate – Supplementary Fig. 3). Repeatability was estimated to be 14 ± 6 RSD %. The robustness of the method was examined at 6, 20 and 60 ng of total spiked amounts to the tubes and pads, respectively. At the highest level, the average recovery ranges between 4 and 91% depending on the chemicals (Supplementary Table 9).

Note that the sweat pads and all applied sampling equipment (tweezers, sample bottles and aluminium foil surfaces) were treated at 150 °C for 20 min to minimize contamination (Supplementary Fig. 4).

The median sweat weight that we measured was 86 ± 91 mg, however including some extreme values of 33 and 408 mg. We tested for systematic errors in absolute peak volume as a function of the artificial sweat amount spiked to the pads. The general trend is a slight decrease in peak volumes as the amount of sweat increases. The set of chemicals that did not systematically increase in peak volume with the added artificial sweat amounts due to blank contribution show a systematic decrease in peak volumes by 56 ± 19% and 54 ± 23% at 300 μL and 600 μL sweat amounts, respectively (Supplementary Table 9).

The effect of humidity was initially a problem for high-temperature DHS of the sweat samples. The mean collected sweat weights were 121 mg for females (n = 18) and 133 mg for males (n = 22), therefore an adsorption of excessive amounts of water on the Tenax-GR is expected despite its low breakthrough volume. Our pre-purge step (750 mL at 250 mL min^{-1}) prevented the excessive introduction of water into the apolar 1st dimension GC column. In fact, water is very destructive during thermal re-focusing of the chemicals due to the poor wettability of the apolar stationary phase with water (Supplementary Figs. 5–8). This thus raises the question of the very volatile sweat chemicals breaking through during the extraction and pre-purging steps. We used selected-ion flow tube mass spectrometer (SIFT-MS) for real-time detection (SYFT, Christchurch, New Zealand) to evaluate the qualitative and quantitative extent of chemicals breaking through (SI Table). The results suggested that acetone and 1-propanol (C_3) are mostly broken through (Supplementary Figs. 9–10) and thus can only be considered for qualitative analysis. 2-Butanone and 1-Butanol showed significant breakthrough as well - about half of the spiked amounts were retained at the concentration level examined. Since the extent of the breakthrough depends on the amount of chemicals in the sample, their peak areas cannot be reliably used for comparative analysis. However, for 2-Pentanone and 1-Pentanol or higher homologues, we can safely assume the retention of the chemicals on Tenax-GR to be quantitative using this method (Supplementary Fig. 11). Therefore, only very few very volatile compounds could not be detected with the developed method. The areas on the chromatogram that are affected are shown in Supplementary Fig. 12.

The question of pH is especially relevant regarding the volatilization of organic acids since they lose their volatility in deprotonated form. The aim of this study was to characterize the volatilization process of chemicals from the skin in a similar way to real life conditions; we did not investigate the pH dependence of extraction dynamics. Since the physiological skin pH ranges between 4.1 and 5.8 or according to other sources 4.0–7.0, which is a complicated pH range with regards to the speciation of short-chain fatty acids, the efficient extraction of organic acids is a weakness of this method.

5. Conclusions

We have developed a semi-automated data-driven workflow from identification to suspect screening, which allows highly selective monitoring of the response of the identified chemicals in large-sample GCxGC data. In other words, we change the focus of a typical GCxGC analysis from: "What do we have in the sample?" to "How do the amounts of the identified compounds change across all samples?". We believe that our framework will open up the possibility to efficiently use GCxGC-HRMS for large-sample studies for a wide range of applications (e.g., biological or environmental studies). In addition, dynamic headspace extraction proved to be a powerful method to extract and preconcentrate sweat volatiles. We were able to extract more hydrophilic compounds than comparable partition-based extraction methods. In

total, we could detect 326 compounds from a diverse range of chemical classes (278 identified compounds, 39 class unknowns, and 9 true unknowns).

Credit author statement

Matyas Ripszam: Conceptualization, Methodology, Software, Validation, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Tobias Bruderer**: Conceptualization, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization. **Denise Biagini**: Conceptualization, Methodology. **Silvia Ghimenti**: Conceptualization, Methodology. **Tommaso Lomonaco**: Conceptualization, Methodology, Resources, Funding acquisition. **Fabio Di Francesco**: Conceptualization, Writing - Review & Editing, Resources, Funding acquisition, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported in part by the European Union Horizon 2020 Programme under Grant 824153 of the project "POTION-Promoting Social Interaction through Emotional Body Odours". The authors would like to acknowledge the technical assistance from Rasmus Bøg Hansen and Tore Vulpius from MSCi ApS (Copenhagen, Denmark). The authors also would like to thank Chiara Cordero and Simone Squara from the University of Torino for their valuable help with using GC Image.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.talanta.2023.124333.

References

[1] A. Rebryk, P. Haglund, Non-targeted screening workflows for gas chromatography-high-resolution mass spectrometry analysis and identification of biomagnifying contaminants in biota samples, 2020, https://doi.org/10.1007/s00216-020.
[2] P.H. Stefanuto, A. Smolinska, J.F. Focant, Advanced chemometric and data handling tools for GC×GC-TOF-MS: application of chemometrics and related advanced data handling in chemical separations, TrAC, Trends Anal. Chem. 139 (2021), 116251, https://doi.org/10.1016/J.TRAC.2021.116251.
[3] A. Lommen, H.J. van der Kamp, H.J. Kools, M.K. van der Lee, G. van der Weg, H.G. J. Mol, MetAlignID: A high-throughput software tool set for automated detection of trace level contaminants in comprehensive LECO two-dimensional gas chromatography time-of-flight mass spectrometry data, J. Chromatogr. A 1263 (2012) 169–178, https://doi.org/10.1016/J.CHROMA.2012.09.056.
[4] Z. Li, S. Kim, S. Zhong, Z. Zhong, I. Kato, X. Zhang, Coherent point drift peak alignment algorithms using distance and similarity measures for two-dimensional gas chromatography mass spectrometry data, J. Chemom. 34 (8) (2020) e3236, https://doi.org/10.1002/cem.3236.
[5] S. Castillo, I. Mattila, J. Miettinen, T. Hy€, O. Ainen, Downloaded via UNIV DI PISA on may 30, Anal. Chem. 83 (2011) 45, https://doi.org/10.1021/ac103308x.
[6] I.A. Titaley, O.M. Ogba, L. Chibwe, E. Hoh, P.H.Y. Cheong, S.L.M. Simonich, Automating data analysis for two-dimensional gas chromatography/time-of-flight mass spectrometry non-targeted analysis of comparative samples, J. Chromatogr. A 1541 (2018) 57–62, https://doi.org/10.1016/j.chroma.2018.02.016.
[7] S.E. Reichenbach, X. Tian, Q. Tao, E.B. Ledford, Z. Wu, O. Fiehn, Informatics for cross-sample analysis with comprehensive two-dimensional gas chromatography and high-resolution mass spectrometry (GCxGC–HRMS), Talanta 83 (4) (2011) 1279–1288, https://doi.org/10.1016/J.TALANTA.2010.09.057.
[8] T. Ieda, S. Hashimoto, T. Isobe, T. Kunisue, S. Tanabe, Evaluation of a data-processing method for target and non-target screening using comprehensive two-dimensional gas chromatography coupled with high-resolution time-of-flight mass spectrometry for environmental samples, Talanta 194 (2019) 461–468, https://doi.org/10.1016/J.TALANTA.2018.10.050.
[9] E. Commission, Promoting social interaction through emotional body odours. https://cordis.europa.eu/project/id/824153.
[10] F.A. Franchina, D. Zanella, E. Lazzari, P.H. Stefanuto, J.F. Focant, Investigating aroma diversity combining purge-and-trap, comprehensive two-dimensional gas chromatography, and mass spectrometry, J. Separ. Sci. 43 (9–10) (2020) 1790–1799, https://doi.org/10.1002/jssc.201900902.
[11] S. Herrera Lopez, M.J. Gomez Ramos, M. Hernando, A. Fernández-Alba, Automated dynamic headspace followed by a comprehensive two-dimensional gas chromatography full scan time-of-flight mass spectrometry method for screening of volatile organic compounds (VOCs) in water, Anal. Methods 5 (2013) 1165–1177, https://doi.org/10.1039/C2AY26234G.
[12] L.M. Dubois, P.H. Stefanuto, K.A. Perrault, G. Delporte, P. Delvenne, J.F. Focant, Comprehensive approach for monitoring human tissue degradation, Chromatographia 82 (5) (2019) 857–871, https://doi.org/10.1007/s10337-019-03710-3.
[13] A. Baldini, S. Frumento, D. Menicucci, A. Gemignani, E. P. Scilingo and A. Greco, "Subjective Fear in Virtual Reality: A Linear Mixed-Effects Analysis of Skin Conductance," in IEEE Transactions on Affective Computing, vol. 13, no. 4, pp. 2047-2057, 1 Oct.-Dec. 2022, doi: 10.1109/TAFFC.2022.3197842.Baldini, A. et Al. Preprint. 2022.
[14] H.Y.J. Chow, T. Górecki, Temperature programming of the second dimension in comprehensive two-dimensional gas chromatography, Anal. Chem. 89 (16) (2017) 8207–8211, https://doi.org/10.1021/acs.analchem.7b02134.
[15] T.M.I.C. (TMIC), The human Metabolome database v5.0. https://hmdb.ca/.
[16] Kanehisa Laboratories, Kyoto encyclopedia of Genes and Genomes. https://www.genome.jp/kegg/.
[17] Wellcome Trust, LIPID MAPS. https://www.lipidmaps.org/.
[18] CTS - the chemical translation Service. https://cts.fiehnlab.ucdavis.edu/.
[19] United States Environmental Protection Agency, CompTox chemicals dashboard. https://comptox.epa.gov/dashboard/.
[20] N. Drabińska, C. Flynn, N. Ratcliffe, I. Belluomo, A. Myridakis, O. Gould, M. Fois, A. Smart, T. Devine, B.D.L. Costello, A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome, J. Breath Res. 15 (3) (2021), https://doi.org/10.1088/1752-7163/abf1d0.
[21] Z. Pang, J. Chong, G. Zhou, D.A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li, J. Xia, MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights, Nucleic Acids Res. 49 (W1) (2021) W388–W396, https://doi.org/10.1093/NAR/GKAB382.
[22] R.A. Van Den Berg, H. Cj Hoefsloot, J.A. Westerhuis, A.K. Smilde, Van Der Werf, M. J. Centering, Scaling, and Transformations: Improving the Biological Information Content of Metabolomics Data, 2006, https://doi.org/10.1186/1471-2164-7-142.
[23] D.J. Penn, E. Oberzaucher, K. Grammer, G. Fischer, H.A. Soini, D. Wiesler, M. V. Novotny, S.J. Dixon, Y. Xu, R.G. Brereton, Individual and gender fingerprints in human body odour, J. R. Soc. Interface 4 (13) (2007) 331–340, https://doi.org/10.1098/rsif.2006.0182.
[24] B. Sgorbini, M.R. Ruosi, C. Cordero, E. Liberto, P. Rubiolo, C. Bicchi, Quantitative determination of some volatile suspected allergens in cosmetic creams spread on skin by direct contact sorptive tape extraction-gas chromatography-mass spectrometry, J. Chromatogr. A 1217 (16) (2010) 2599–2605, https://doi.org/10.1016/j.chroma.2009.12.052.
[25] A.M. Curran, S.I. Rabin, P.A. Prada, K.G. Furton, Comparison of the volatile organic compounds present in human odor using SPME-GC/MS, 2005, https://doi.org/10.1007/s10886-005-5801-4.
[26] X.N. Zeng, J.J. Leyden, A.I. Spielman, G. Preti, Analysis of characteristic human female axillary odors: qualitative comparison to males, J. Chem. Ecol. 22 (2) (1996) 237–257, https://doi.org/10.1007/BF02055096.
[27] X. nong Zeng, J.J. Leyden, H.J. Lawley, K. Sawano, I. Nohara, G. Preti, Analysis of characteristic odors from human male axillae, J. Chem. Ecol. 17 (7) (1991) 1469–1492, https://doi.org/10.1007/BF00983777.
[28] A. Mitra, S. Choi, P.R. Boshier, A. Razumovskaya-Hough, I. Belluomo, P. Spanel, G. B. Hanna, The human skin volatolome: a systematic review of untargeted mass spectrometry analysis, Metabolites 12 (9) (2022), https://doi.org/10.3390/metabo12090824.
[29] N.O. Verhulst, B.T. Weldegergis, D. Menger, W. Takken, Attractiveness of volatiles from different body parts to the malaria mosquito Anopheles coluzzii is affected by deodorant compounds, Sci. Rep. 6 (2016) 1–9, https://doi.org/10.1038/srep27141. May.
[30] S.H. López, M.J. Gómez, M.D. Hernando, A.R. Fernández-Alba, Automated dynamic headspace followed by a comprehensive two-dimensional gas chromatography full scan time-of-flight mass spectrometry method for screening of volatile organic compounds (VOCs) in water, Anal. Methods 5 (5) (2013) 1165–1177, https://doi.org/10.1039/c2ay26234g.