Published in final edited form as:

Anal Chim Acta. 2018 October 16; 1027: 158–167. doi:10.1016/j.aca.2018.03.037.

SPME-GC×GC-TOF MS FINGERPRINT OF VIRALLY-INFECTED CELL CULTURE: SAMPLE PREPARATION OPTIMIZATION AND DATA PROCESSING EVALUATION

Giorgia Purcaro[1,*,†], **Pierre-Hugues Stefanuto**[1,†], **Flavio A. Franchina**[1], **Marco Beccaria**[1], **Wendy F. Wieland-Alter**[2], **Peter F. Wright**[2,3], and **Jane E. Hill**[1,2]

[1]Thayer School of Engineering, Dartmouth College, Hanover, NH, 03755, United States

[2]Geisel School of Medicine, Dartmouth College, Hanover, NH, 03755, United States

[3]Dartmouth-Hitchcock Medical Center, Lebanon, NH, 03756, United States

Abstract

Untargeted metabolomics study of volatile organic compounds produced by different cell cultures is a field that has gained increasing attention over the years. Solid-phase microextraction has been the sampling technique of choice for most of the applications mainly due to its simplicity to implement. However, a careful optimization of the analytical conditions is necessary to obtain the best performances, which are highly matrix-dependent. In this work, five different solid-phase microextraction fibers were compared for the analysis of the volatiles produced by cell culture infected with the human respiratory syncytial virus. A central composite design was applied to determine the best time-temperature combination to maximize the extraction efficiency and the salting-out effect was evaluated as well. The linearity of the optimized method, along with limits of detection and quantification and repeatability was assessed. Finally, the effect of i) different normalization techniques (*i.e.* z-score and probabilistic quotient normalization), ii) data transformation (*i.e.* in logarithmic scale), and iii) different feature selection algorithms (*i.e.* Fisher ratio and random forest) on the capability of discriminating between infected and not-infected cell culture was evaluated.

Keywords

solid-phase microextraction (SPME); comprehensive two-dimensional gas chromatography (GC×GC); volatile organic compounds (VOCs); virus; data processing; chemometrics

1. Introduction

In the biomedical field, the analysis of the volatile organic compounds (VOCs) produced by human fluids (*e.g.* breath, urine, feces) and tissue (*e.g.* skin) has been proposed as a diagnostic tool for diseases since it does not require any invasive procedures [1–6]. However, preliminary *in vitro* studies are fundamental for a proof-of-concept of the diagnostic

*Corresponding author. Tel: 1 (603) 646-8656; Fax: 1 (603) 646-0743, Giorgia.purcaro@dartmouth.edu; giopurcaro@gmail.com.
†These authors contributed equally.

potentiality of VOCs. Many *in vitro* investigations have been carried out to define a bacterial core metabolome that can be critically translated into *in vivo* scenarios [7–9], while very few works have focused on the effect of viral infection on the volatile profile of cell culture [10–14]. The latter has been more extensively investigated for cancer diagnosis [3,15,16].

Solid-phase microextraction (SPME) has been used widely for the analysis of VOCs since its invention in the early 1990s [17–21]. SPME is a simple and effective sample preparation technique, which combines sampling, isolation, and concentration in a single step. Although very easy to use from a practical viewpoint, the optimization of extraction conditions needs to be tailored to the objective (targeted or untargeted) and the specific samples under investigation [22]. The performance of different SPME fibers and their behaviors for the analysis of VOCs in different samples have been extensively studied in food, environmental and flavor and fragrance studies [20,21]. While, in the bioanalytical and clinical applications, efforts have been mainly devoted to direct immersion SPME for small metabolite analysis, rather than for the optimization of VOCs analysis [19].

The aim of this work was to optimize and validate an SPME method for the analysis of VOCs produced in cell culture infected with respiratory syncytial virus (RSV), the most common cause of deadly lower respiratory tract infections in children younger than 2 years [23]. The extraction yields of different SPME fibers were compared prior to optimize the extraction temperature/time combination using a central composite design, and assessing the salting out effect. After pre-concentration, VOCs were analyzed by comprehensive two-dimensional gas chromatography (GC×GC) hyphenated with a time-of-flight mass spectrometer (ToF MS), which enhances the number of information available by increasing the sensitivity and the selectivity compared to a conventional GC system [24–26]. The final data matrix obtained from the analysis of cell culture infected with RSV was treated with different data processing techniques and the discriminatory capability of the selected volatiles was evaluated. This work is a preliminary step to establish a solid analytical basis for further investigations aimed to determine volatile biomarkers *in vitro*.

2. Materials and Methods

2.1. Reagents and standards

Hexane was HPLC grade (MilliporeSigma®, USA). A mixture of normal alkanes (C_6–C_{20}), and a mixture of 36 standards (containing 1-chlorodecane; 1-chlorododecane; 1-chloroundecane; 1-decanol; 1-dodecanol; 1-hexanol; 1-nonanol; 1-octanol; 1-undecanol; 2-decanol; 2-decanone; 2-heptanol; 2-heptanone; 2-nonanol; 2-nonanone; 2-octanol; 2-octanone; 2-undecanone; butyl-benzene; heptyl-benzene; hexyl-benzene; nonyl-benzene; octyl-benzene; decane; decanoic acid methyl ester; dodecane; dodecanoic acid methyl ester; heptanoic acid methyl ester; hexadecane; nonanoic acid methyl ester; octanoic acid methyl ester; pentadecane; pentanoic acid methyl ester; tridecane; undecane; undecanoic acid methyl ester) were from Supelco (Bellefonte, PA, USA). The mixture of alkanes was injected to calculate the linear retention index (LRI). The mixture of standards was used to optimize the chromatographic conditions, to test the linearity and the limit of detection (LOD) and quantification (LOQ) of the optimized method.

2.2 Respiratory syncytial virus (RSV) infection of human cell lines

A human laryngeal cancer cell line (HEp-2 cells) from the American Type Culture Collection (ATCC®, CL-23™) was seeded in a six-well microtiter plates (4×10^5 cells/well) to be 70–80 % confluent in 24 h. Human respiratory syncytial virus (ATCC® VR-1540™) was diluted to a multiplicity of infection (MOI) of 0.3 in phosphate-buffered saline (PBS) prior to addition to the HEp-2 cells. Viral infection was performed as described previously [27]. Briefly, Hep-2 cells were maintained in a growth media consisting of Minimum Essential Medium (MEM) containing penicillin-streptomycin and 2 % fetal bovine serum (Corning CellGro 15-010). The culture supernatant was removed just before inoculating the cells with the viral suspension. The plates were incubated at 37 °C with a 5 % CO_2 atmosphere for 1.5 h, then the viral suspension was aspirated and MEM was added to maintain the HEp-2 cells. Eight six-well plates were prepared, four were infected with RSV, while four were used as control. Supernatant from three out of four six-well plates of each group was collected 48h after infection and pooled together to form a quality control (QC) sample used for the SPME optimization. The QC sample was divided in 2.5 mL-aliquot into a 10 mL air-tight glass vials sealed with a PTFE/silicone cap (both from Sigma-Aldrich). The remaining two six-well plates (one RSV and one control), composed of six samples each, collected at 48 h after infections, and were analyzed using the final optimized SPME method. All samples were frozen at −20 °C and were analyzed within one week of collection.

2.3 Solid-phase microextraction optimization: Central Composite Design

The most performing SPME fiber was selected by analyzing the QC samples (described in Section 2.2) using all five fibers at 43°C for 30 min (corresponding to the central point of the experimental design described later). The fiber tested were: polydimethylsiloxane/carboxen/divinylbenzene (PDMS/Car/DVB) d_f50/30 μm, 2 cm length fiber, PDMS/DVB d_f65 μm, PDMS/CAR d_f85 μm, PDMS d_f100 μm, and over-coated (OC)-PDMS/DVB d_f65/10 μm (all from Supelco, Bellefonte, PA, USA). Three replicated extractions were performed for each fiber.

Then, a two-variable ($k=2$) inscribed rotatable ($\alpha = 1/\sqrt[2]{k}$) central composite experimental design (CCD) was used to optimize the sampling conditions, namely extraction temperature and time. The extraction temperature was tested between 37 °C and 50 °C, and the exposition time from 15 to 45 minutes. Nine different sampling conditions were included in the design, consisting of a central point, four axial and four factorial points (Table 1). To increase the precision of the model each condition was replicated three times and the central point was repeated six times to evaluate the repeatability of the method as well. The peak areas obtained by GC×GC time-of-flight (ToF) MS were used to evaluate the extraction efficiency using the response surface plot methodology [28].

The salting out effect was evaluated at the final optimized sampling condition by adding KCl at different concentrations [0%, 20%, or 40% (saturation)].

All samples were agitated at 250 rpm and incubated for 15 min before fiber exposure at the corresponding extraction temperature. Each extraction was replicated three times.

Final optimized conditions: fiber: PDMS/Car/DVB; sample volume: 2.5 mL of supernatant in a 10 mL vial; extraction temperature: 43 °C; equilibration time: 15 min; extraction time: 30 min; salt addition: 40% (w/v) of KCl.

In all the experiments, the fiber was thermally desorbed in the GC injector for 1 min at 250 °C in splitless mode.

2.4. Method validation

Linearity, limit of detection (LOD) and quantification (LOQ) were evaluated using the mixture of 36 standards listed in Section 2.1. The standards were added to the media used to growth the cell culture (MEM) at six different concentrations (0.1, 1.0, 5, 10, 50, 100 μg/L). Three measurements were performed at each concentration. MEM without any standard addition was analyzed to perform blank subtraction. The least squares method was applied to estimate the regression line. The significance of the intercept was established by running a *t*-test. The LOD and the LOQ were calculated for the entire method by considering the standard deviation (σ) at the lowest concentration used to build the calibration lines and applying the following formulas:

$$\text{LOD} : y_d = 3 \times \sigma/s$$

$$\text{LOQ} : y_q = 10 \times \sigma/s$$

where y_d and y_q are the signals at the LOD and LOQ, respectively, and s is the slope of each calibration curve.

Precision was evaluated as the coefficient of variation (CV %) at the central point of the CCD (n=6).

2.5 Analytical Instrumentation

A Pegasus 4D (LECO Corporation®, St. Joseph, MI) GC×GC time-of-flight (ToF) MS instrument with an Agilent® 7890 GC, and equipped with an MPS autosampler (Gerstel®, Linthicum Heights, MD, USA), was used. The primary column was an SLB-5MS (30 m × 250 μm × 0.25 μm) connected in series with a Carbowax secondary column (1 m × 250 μm × 0.5 μm) from Supelco (Bellefonte, PA, USA). The carrier gas was helium, at a flow rate of 0.7 mL/min, corresponding to about 28 cm/s and 140 cm/s in first and second dimension, respectively. The primary oven temperature program was 45 °C (hold 1 min) ramped to 200 °C at a rate of 5 °C/min and then to 220 °C at 15 °C/min. A modulation period of 3 s (alternating 0.85 s hot and 0.65 s cold) was used. The transfer line temperature was set at 250 °C. A mass range of *m/z* 30 to 500 was collected at a rate of 100 spectra/s following a 2 min acquisition delay. The ion source was maintained at 200 °C. Data acquisition and analysis were performed using ChromaTOF software, version 4.50 (LECO Corp.).

2.6 Processing and analysis of chromatographic data

Chromatographic data was processed and aligned using ChromaTOF®. For peak identification, a signal-to-noise (S/N) cutoff was set at 50:1 in at least one chromatogram and a minimum of 20:1 S/N ratio in all others. For the alignment of peaks across chromatograms, maximum first and second-dimension retention time deviations were set at

9 s and 0.2 s, respectively, and the inter-chromatogram spectral match threshold was set at 600.

Compounds with retention time prior to 4 min were removed. Artifacts such as siloxane, and phthalates, were identified with the support of the script tool available in ChromaTOF®. The scripts used to identify linear siloxanes (defined as “siloxane” in the script) and phthalates were adopted from Weggler *et al.* [29]. Additional scripts were also developed to define cyclosiloxanes and silanols. Each script corresponds to a particular class of chemicals (*e.g.* linear siloxane, cyclosiloxane, silanol, phthalate). Each script is based on the presence of a specific parent ion (*e.g.* 73, 207, 281), and on the relative abundance between different ions (*e.g.* 207/281). In any cases, flagged peaks using the aforementioned scripts were then manually checked and deleted from the peak table and not considered for further data analysis. The scripts used are reported in Supplementary data.

Compounds were tentatively identified based on mass spectral similarities to the NIST 2011 library, with a match score ≥ 850 (of 1000) and linear retention index (LRI) window (±10 units).

2.6 Data processing on virally-infected cell culture samples

Virally infected cell-culture samples obtained as reported in section 2.2 were analyzed with the optimized method and the resulting data matrix was treated with different data processing for comparison purposes. Normalization was performed using Probabilistic Quotient Normalization (PQN) [30] or z-score (then using the absolute z-score value) [31]. Originally, PQN was developed for NMR data. An integral normalization of each sample is first carried out, then a “reference” profile is created using the median, and finally all variables are divided by the median quotient [30]. The z-score normalization is based on the difference between a value and the mean, divided by the standard deviation (as a scaling factor) [40]. The effect of logarithmic transformation on the variables was also evaluated. Logarithmic transformation is a nonlinear conversion of data widely applied to stabilize the variance and making the distribution of the variables closer to normal [31]. All the results obtained were visualized using heatmaps and hierarchical clustering analysis. Data reduction was performed using Random Forest (RF) [32] or Fisher-ratio (F-ratio) [33]. RF is a machine learning algorithm that works building a multitude of de-correlated decision trees and merge them together to obtain a more accurate and stable prediction. The importance of each feature is evaluated by permuting the variables and measuring how much the permutation affects the accuracy of the model. Unimportant variables should not affect significantly the model. The decrease in accuracy is averaged over all trees (*i.e.* mean decrease accuracy) and it is used to rank the importance of the features. Features with a mean decrease accuracy value above 0.015 (which was the value at the elbow of the distribution of the mean decrease accuracy rank in all the cases tested as shown in the plots in Supplementary Figures S5C, S7–S10C) were selected.

F-ratio is a univariate ANOVA based data reduction tools. The F-ratio is calculated as the ration between the “between” and the “within” class variances. F-ratio values obtained were compared to the tabulated F-critical value defined regarding the number of classes, degrees of freedom and a confidence interval of 95% [31]. Features with a F-ratio value > F-critical

value were retained as the most significant. The separation performances of the different data analysis approaches were evaluated measuring the Euclidean distance intra- and inter-group [31].

All statistical analyses were performed using R v3.3.2 (R Foundation for Statistical Computing, Vienna, Austria) and Excel® (Microsoft Office, version 2013).

3. Results and Discussion

3.1. SPME method optimization

The aim of this work was to optimize SPME fiber-types and conditions for VOCs analysis in viral cell culture. Some basic parameters [*i.e.* agitation (250 rpm), sample volume-HS ratio (1:4), and desorption conditions (250 °C for 1 min)] were determined based on experience or previous evaluation with similar samples [34].

3.1.1. Fiber Selection—Prior to optimize the extraction condition using the CCD (Table 1), the extraction performance of five commercial fibers (PDMS, PDMS/DVD, OC-PDMS/DVB, PDMS/Car/DVB, and PDMS/Car) were compared by analyzing the QC sample at the central point of the CCD, namely 43 °C for 30 min. Desorption efficiency in the GC inlet was assessed for each fiber by verifying the absence of carry-over by desorbing again the fiber after the sample injection. No carry-over was observed for any of the tested fibers.

Twelve compounds were selected to evaluate the overall SPME performance of the fibers. The selection of the compounds was performed according to the following criteria: I) present in 75% of the samples and II) covering the most common chemical families. The selected compounds are reported in Table 2.

The higher extraction yield for all the considered compounds was obtained using the PDMS/Car/DVB fiber, followed by the PDMS/DVB (about 60 % on average) except for the lighter compounds [#1 (Cyclohexane), #2 (2-Pentanone) and #3 (2-Pentanone, 4-methyl-)] which were none or barely extracted with the latter fiber (Figure 1). The PDMS/Car fiber extracted only 5 compounds out of 12 and, as expected, the most volatile ones since the micropores of the Carboxen sorbent are particularly suitable for retaining smaller analytes [18]. The extraction yields of the 12 compounds using the OC-PDMS/DVB fiber were about half of the intensity obtained using the conventional PDMS/DVB fiber, although with a very similar profile, except for cyclohexane (#1) which was extracted in the larger amount using the OC-PDMS/DVB fiber (Figure 1). The OC-PDMS/DVB fiber was first developed by Souza Silva *et al.* by adding an extra 10 μm PDMS layer to a commercial PDMS/DVB to enhance the robustness of the latter when performing direct-immersion extraction, without altering the extraction performance [35].It has been shown in previous studies that the additional PDMS layer increases the extraction rate of highly hydrophobic compounds, while it does not affect the uptake of polar and midpolar analytes when compared to the conventional PDMS/DVB [36,37]. Neverthless the extra layer significantly affect the kinetics of the sorption for more polar analytes, requiring longer extraction time to obtain sensitivity comparable to the conventional PDMS/DVB [36,37]. It is interesting to notice that, in general, the repeatability (as given by CV %) of OC-PDMS/DVB is better than

PDMS/DVB, which resulted in an average 20% and 35%, respectively. A longer extraction time might give higher extraction yields with an overall better repeatability, but it would have extended the entire analysis time reducing the sample throughput. The fiber coated with PDMS provided poor results extracting few compounds, therefore it was excluded and it will not be discussed any further. The PDMS/Car/DVB fiber was selected for further optimization.

3.1.2. Temperature and time optimization: Central Composite Design—A two-variable inscribed rotatable CCD was used to optimize the sampling conditions related to temperature and time of exposition for PDMS/Car/DVB [38]. The extraction temperature was evaluated between 37 °C (physiological temperature) and 50 °C (maximum temperature to avoid the formation of artifacts due to thermal degradation). The extraction time was tested between 15 and 45 minutes. Longer time was not considered to avoid instrument idle time and maximize the throughput. Since the optimal point within the region of investigation was not known *a priori*, a rotatable CCD was selected thus any tested points, equally distant from the central point, had the same magnitude of prediction error. Nine different sampling points (1 central, 4 axial, and 4 factorial) were tested (Table 1). The peak areas of the previously selected compounds were used to calculate the response surfaces. The maximum of the response surface represented the best combination of extraction time and temperature. A clear maximum in the response surfaces, corresponding to the maximum extraction yield, was observed for five compounds: 2,4-dimethyl-heptane (#4), styrene (#5), 2,6-dimethyl-nonane (#6), 5-ethyl-2-methyl-octane (#8), and 1,3-bis(1,1-dimethylethyl)-benzene (#11) at around the central point (43 °C for 30 min) (Figure 2). Nonanal (#9) showed a very similar trend, but a plateau was reached at about the same temperature and time pair, but no decrease was observed within the tested range (Figure 2). A maximum was not observed for the remaining six compounds [cyclohexane (#1), 2-pentanone (#2), 4-methyl 2-pentanone (#3), 2-ethyl 1-hexanol (#7), dodecane (#10), and phenol,2,4-bis(1,1-dimethylethyl)-(#12)], although a direct correlation between time and temperature was clearly observed (Supplementary Figure 1). Peculiar trends were observed for 2-pentanone (#2), whose uptake was independent of the extraction temperature but required long extraction time, and for cyclohexane (#1), whose extraction increased with the temperature irrespectively of the time (Supplementary Figure 1).

The central point conditions (43 °C for 30 min) were selected as the best compromise for the following analysis.

3.1.3. Salting out effect—The PDMS/Car/DVB fiber was exposed to the HS of QC samples containing different amounts of KCl (*i.e.* 0%, 20%, or 40% w/v) at 43 °C for 30 min. Salt addition to liquid samples can alter the solubility of the solutes, thus changing their concentration in the HS. The specific effect depends on the compound characteristics and salt concentration. The salting out effect was evaluated on the 12 compounds selected previously. The extraction yields for the selected compounds are shown in Figure 3. A significantly higher uptake ($p<0.05$) was obtained for all the compounds adding 20 and 40 % of salt, except for cyclohexane (#1), styrene (#5) and dodecane (#10) for which no significant differences ($p>0.05$) were observed among the three conditions tested. No

significant differences were observed between the addition of 20% and 40% of salt, except for 2,6-dimethyl nonane (#6) for which a slightly higher uptake was obtained adding 40% of salt. The overall precision of the measurements, in terms of standard deviation of the areas obtained adding 0%, 20% and 40% of salt were compared using a *t*-test on the average of the 12 compounds. No significant difference was observed ($p>0.05$). It can be concluded that the salting-out effect affect only the absolute uptake without altering the precision performance of the method. Therefore, the addition of 40% of salt (saturating condition) was selected to maximize the extraction and minimize the ionic strength variability, which can occur when biological samples are analyzed.

3.2. Method validation

Although an absolute quantification is out of the scope of the present paper, linearity, LOD, and LOQ were evaluated using a mixture of 36 standards belonging to different chemical classes to cover a broader range of possible volatile compounds detectable in microbiological specimens. In fact, unsupervised and supervised algorithms, used to perform pattern recognition analysis, rely on the absolute areas obtained from the instrument. Therefore, it was fundamental to assess that the working concentration range was within the linear dynamic range of the analytical method used.

The media (MEM) was spiked with the mixture of standards at six different concentrations in the 0.1–100 μg/L range. The solutions were analyzed using the optimized method, three replicates at each concentration level. Ten compounds reached saturation after 50 μg/mL and 3 (1-chloroundecane, 1-undecanol, and 1-chlorododecane) were not detectable at all at the lowest concentration tested (0.1 μg/L). The least squares method was applied to estimate the regression lines, obtaining regression coefficients (R^2) ranging between 0.906 and 0.991, with an average of 0.944 (Table 3).

The intercept resulted in a value significantly different from 0 only for decane ($p=0.0048$).

The LOD and the LOQ were calculated for the entire method obtaining 0.6 and 2.0 μg/L on average as LOD and LOQ, respectively, not considering the higher values obtained for 1-chlorododecane (5.3 and 17.7 μg/L for LOD and LOQ, respectively.

Precision was verified by assessing the repeatability at the central point of the CCD (n=6). The precision was calculated in terms of coefficient of variation (CV%) and was below 20% for all the compounds considered (Table 2).

3.2. Sample analyses and statistical evaluation

The optimized SPME method was used to generate a data matrix from the analysis of six samples of cell culture (defined as “control”) and six samples of cell culture infected with RSV (defined as “RSV”). After assessing that the chromatographic signal for each chemical class was within the linear range, the raw dataset was elaborated using different data treatment and data reduction approaches: I) no normalization nor transformation steps, II) normalization (PQN or z-score), III) transformation (data were logarithmically transformed), and IV) data reduction applying different features selection algorithms (F-ratio or RF). The overall scheme of the comparison is reported in Figure 4.

Data mining of complex dataset usually requires different steps of pre-processing (*e.g.* normalization and transformation). Normalization is usually applied to minimize unwanted systematic variations between samples, however, it could hinder difference between samples if not properly used. Although, from a strictly analytical viewpoint, the most efficient technique to minimize such a bias is the use of multiple internal standards, this is not always possible for untargeted profiling of complex samples and in exploratory experiments. Different normalization techniques have been proposed in the literature but there is no standard rule to select which is the most suitable [39]. For this work two completely different normalization approach were selected, namely z-score and PQN. The PQN accounts for dilution of the biological samples. This method assumes that biological changes only influence parts of the entire profile, while dilution effect affects the entire profile [30]. The use of median values for normalization insures a stability towards outliers and sampling variability, which can occur in metabolomics. The z-score (or auto-scaling) preserves the intensity range and accounts for the dispersion of the variable over all the samples. It is more sensitive to outliers, but it provides a better conservation of small peak area variations and considers all the features as equally important [40]. The effect of logarithmic data transformation was tested as well. Logarithmic transformation is a nonlinear conversion of data widely applied to stabilize the variance and making the distribution of the variables closer to normal [31]. Finally, data reduction was carried out to minimize noise and redundancy, and to extract the most useful information. Among the many features reduction/selection algorithms a univariate and a multivariate approach were selected to be evaluated in the present paper, namely F-ratio and RF, since widely applied in the field. F-ratio is a univariate data reduction technique that measures the discrimination of two sets of real numbers. Larger is the F-ratio value more likely the feature is discriminatory. A disadvantage of F-score is that it is not capable to reveal mutual information among features. Differently, RF is a non-parametric approach that can deal with highly collinear data and it is resistant to different type of outliers. Some examples of the use of these normalization and data reduction techniques in the GC×GC literature are the following [9,34,41,42]).

All the results were visualized using heatmaps with hierarchical clustering analysis (Supplementary Figures 2–10), while for a more straightforward and objective comparison the Euclidean distances intra- and inter- groups were measured (Table 4). The intra-group Euclidean distance is a value of the dispersion of the samples and the inter-group gives a measure of the separation between the two classes. In order to compare the results, the intra-group distances were reported as the ratio between the intra- and the inter-group ones, while to evaluate and compare the inter-group separation the Euclidean distances within the group (Control and RSV) were summed and divided by the intra-groups distance. The result obtained can roughly be compared to the chromatographic resolution, where, in this case, 1.0 represented the minimum value to have a separation between the two groups. If the resolution was < 1.0 the groups were partially overlapped, while if it was >1.0 they were better separated (Table 4).

The raw data matrix after alignment and artifacts removal (siloxane and phthalates) was composed of 260 features. Hierarchical cluster analysis was performed on this matrix without any data treatment and failed in a proper discrimination of the two classes (Supplementary Figure 2A), with a resolution of 0.66.

Normalization, variable transformation, and data reduction were then employed.

Prior to any feature selection (Features selection “None” in Table 4), the PQN normalization slightly improved the resolution (0.71) when no log transformation was applied (Supplementary Figure S3), while the resolution was 0.66 without normalization or applying the z-score normalization (Supplementary Figure S2A and S4A). Logarithmic transformation negatively affected the separation between groups (regardless the normalization technique applied (Supplementary Figure S2B, S3B, and S4B). On the contrary, when data reduction is applied, the logarithmic transformation improved the discrimination between groups (Supplementary Figure S5–S10), except when F-ratio is used without any normalization (Supplementary Figure S6). As expected, the logarithmic transformation reduced the intra-group dispersion (lower Euclidean distance). The highest resolution was obtained applying the z-score normalization, logarithmic transforming and selecting the features using F-ratio (resolution = 2.54), followed by the combination of PQN, logarithmic transformation, and RF (resolution=2.41).

Feature selection was proved to be a fundamental step to increase the discriminatory capability. The most significant features in the RF were selected based on the mean decrease accuracy values. Plotting such values an elbow at 0.015 was clearly present and features above such a value were retained as the most discriminatory (Supplementary Figures 3C, 4C, 6C, 7C, 9C, 10C). When the feature selection is performed using the F-ratio value, the cutoff is the tabulated F-critical value defined according to the number of classes, degrees of freedom and a confidence interval of 95% [31]. Only the features with an F-ratio above the critical value were retained. Features without an F-value (because sparse or present only in one group) were retained when present in at least 50% of samples within a class (indicated with a * in Table 5). The identification of the selected features is reported in Table 5, along with the combination of data treatments that determined their selection. A name was assigned when MS library match was over 850 (out of 1000) or when the similarity was >750 and the LRI within ±10 (except for 2 oxo-propanoic acid, which was ±14). The RF algorithm was proved to be very robust and consistent, in fact very similar resolutions were obtained regardless of the normalization technique applied and the same 7 features were always selected. The compounds constantly selected using RF overlapped with the features selected applying the F-ratio, except for cyclohexanol, which was picked up by F-ratio only when z-score with no logarithmic transformation were applied. Interestingly, in the latter case, the feature selected by F-ratio completely overlap with RF.

Conclusions

Great attention is currently devoted to the untargeted study of VOCs produced by different cell cultures; however, a proper systematic optimization of the analytical conditions has never been reported before. Furthermore, although most of the investigations are qualitative and exploratory, assessment of minimal validation parameters, such as linearity, LOD, and LOQ of the applied method are fundamental for a reliable further data mining, which relies on both quali- and quantitative differences in VOCs profile. In this work, a careful optimization procedure was performed to choose the most suitable fiber for the analysis of VOCs in the HS of human cell cultures. Among the 5 different tested fibers, the

PDMS/Car/DVB gave the best extraction yields for all the compounds considered. A CCD was then employed to establish the best time-temperature combination to maximize the extraction efficiency (43 °C, 30 min). The addition of salt (40 %) was also proved to be beneficial for increasing the overall extraction yield.

The optimized method was then used for the analysis of cell cultures infected with RSV as well as uninfected cell cultures. Different data mining methods were applied to this dataset to evaluate the effect of normalization, transformation, and data reduction on the final discrimination capability. It is interesting to notice that feature selection is fundamental to remove noise and redundancy, increasing the level of information extractable from a complex dataset. The feature selection step was effective for increasing the pattern recognition capability using both F-ratio and RF, but it is worthy to stress that RF gave very consistent results irrespective of prior data treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support for this work was provided by Hitchcock Foundation and the National Institute of Health (NIH/NIAID, Project # 1R21AI12107601). P-H. Stefanuto is a Marie-Curie COFUND postdoctoral fellow co-funded by the European Union and the University of Liège.

The authors gratefully acknowledge Supelco for providing the SPME fibers.

References

1. Boots AW, Bos LD, van der Schee MP, van Schooten FJ, Sterk PJ. Exhaled Molecular Fingerprinting in Diagnosis and Monitoring: Validating Volatile Promises. Trends Mol. Med. 2015; 21:633–644. DOI: 10.1016/j.molmed.2015.08.001 [PubMed: 26432020]
2. Amann A, Costello B de L, Miekisch W, Schubert J, Buszewski B, Pleil J, Ratcliffe N, Risby T. The human volatilome: volatile organic compounds (VOCs) in exhaled breath, skin emanations, urine, feces and saliva. J. Breath Res. 2014; 8:34001.doi: 10.1088/1752-7155/8/3/034001
3. Filipiak W, Mochalski P, Filipiak A, Ager C, Cumeras R, Davis CE, Agapiou A, Unterkofler K, Troppmair J. A Compendium of Volatile Organic Compounds (VOCs) Released By Human Cell Lines. Curr. Med. Chem. 2016; 23:2112–31. DOI: 10.2174/0929867323666160510122913 [PubMed: 27160536]
4. Capelli L, Taverna G, Bellini A, Eusebio L, Buffi N, Lazzeri M, Guazzoni G, Bozzini G, Seveso M, Mandressi A, Tidu L, Grizzi F, Sardella P, Latorre G, Hurle R, Lughezzani G, Casale P, Meregali S, Sironi S. Application and uses of electronic noses for clinical diagnosis on urine samples: A review. Sensors (Switzerland). 2016; 16:1–23. DOI: 10.3390/s16101708
5. Sethi S, Nanda R, Chakraborty T. Clinical application of volatile organic compound analysis for detecting infectious diseases. Clin. Microbiol. Rev. 2013; 26:462–475. DOI: 10.1128/CMR.00020-13 [PubMed: 23824368]
6. Gisbert JP, Pajares JM. Review article: 13C-urea breath test in the diagnosis of Helicobacter pylori infection - A critical review. Aliment. Pharmacol. Ther. 2004; 20:1001–1017. DOI: 10.1111/j.1365-2036.2004.02203.x [PubMed: 15569102]
7. Mellors TR, Rees CA, Wieland-Alter WF, von Reyn CF, Bean H, Hill JE. The volatile molecule signature of four mycobacteria species. J. Breath Reserch. 2017; 11:31002.

8. Ratiu I-A, Ligor T, Bocos-Bintintan V, Buszewski B. Mass spectrometric techniques for the analysis of volatile organic compounds emitted from bacteria. Bioanalysis. 2017; 9:1069–1092. DOI: 10.4155/bio-2017-0051 [PubMed: 28737423]
9. Rees CA, Franchina FA, Nordick KV, Kim PJ, Hill JE. Expanding the Klebsiella pneumoniae volatile metabolome using advanced analytical instrumentation for the detection of novel metabolites. J. Appl. Microbiol. 2017; 122:785–795. DOI: 10.1111/jam.13372 [PubMed: 27930839]
10. Schivo M, Aksenov Aa, Linderholm AL, McCartney MM, Simmons J, Harper RW, Davis CE. Volatile emanations from in vitro airway cells infected with human rhinovirus. J. Breath Res. 2014; 8:37110.doi: 10.1088/1752-7155/8/3/037110
11. Aksenov AA, Sandrock CE, Zhao W, Sankaran S, Schivo M, Harper R, Cardona CJ, Xing Z, Davis CE. Cellular scent of influenza virus infection. ChemBioChem. 2014; 15:1040–1048. DOI: 10.1002/cbic.201300695 [PubMed: 24719290]
12. Phillips M, Cataneo RN, Chaturvedi A, Danaher PJ, Devadiga A, Legendre Da, Nail KL, Schmitt P, Wai J. Effect of influenza vaccination on oxidative stress products in breath. J. Breath Res. 2010; 4:26001.doi: 10.1088/1752-7155/4/2/026001
13. Rochford K, Chen F, Waguespack Y, Figliozzi RW, Kharel MK, Zhang Q, Martin-Caraballo M, Hsia SV. Volatile Organic Compound Gamma-Butyrolactone Released upon Herpes Simplex Virus Type-1 Acute Infection Modulated Membrane Potential and Repressed Viral Infection in Human Neuron-Like Cells. PLoS One. 2016; 11:e0161119.doi: 10.1371/journal.pone.0161119 [PubMed: 27537375]
14. Abd El Qader A, Lieberman D, Shemer Avni Y, Svobodin N, Lazarovitch T, Sagi O, Zeiri Y. Volatile organic compounds generated by cultures of bacteria and viruses associated with respiratory infections. Biomed. Chromatogr. 2015; 29:1783–1790. DOI: 10.1002/bmc.3494 [PubMed: 26033043]
15. Buszewski B, Ulanowska A, Ligor T, Jackowski M, Kłodzińska E, Szeliga J. Identification of volatile organic compounds secreted from cancer tissues and bacterial cultures. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 2008; 868:88–94. DOI: 10.1016/j.jchromb.2008.04.038
16. Filipiak W, Sponring A, Mikoviny T, Ager C, Schubert J, Miekisch W, Amann A, Troppmair J. Release of volatile organic compounds (VOCs) from the lung cancer cell line CALU-1 in vitro. Cancer Cell Int. 2008; 8:17.doi: 10.1186/1475-2867-8-17 [PubMed: 19025629]
17. Arthur CL, Pawliszyn J. Solid phase microextraction with thermal desorption using fused silica optical fibers. Anal. Chem. 1990; 62:2145–2148. DOI: 10.1021/ac00218a019
18. Bojko B, Reyes-Garce´s N, Bessonneau V, Goryn´ski K, Mousavi F, Souza Silva EA, Pawliszyn J. Solid-phase microextraction in metabolomics. TrAC - Trends Anal. Chem. 2014; 61:168–180. DOI: 10.1016/j.trac.2014.07.005
19. Souza-Silva ÉA, Reyes-Garcés N, Gómez-Ríos GA, Boyaci E, Bojko B, Pawliszyn J. A critical review of the state of the art of solid-phase microextraction of complex matrices III. Bioanalytical and clinical applications. TrAC - Trends Anal. Chem. 2015; 71:249–264. DOI: 10.1016/j.trac.2015.04.017
20. Souza-Silva ÉA, Jiang R, Rodríguez-Lafuente A, Gionfriddo E, Pawliszyn J. A critical review of the state of the art of solid-phase microextraction of complex matrices I. Environmental analysis. TrAC - Trends Anal. Chem. 2015; 71:224–235. DOI: 10.1016/j.trac.2015.04.016
21. Souza-Silva ÉA, Gionfriddo E, Pawliszyn J. A critical review of the state of the art of solid-phase microextraction of complex matrices II. Food analysis. TrAC - Trends Anal. Chem. 2015; 71:236–248. DOI: 10.1016/j.trac.2015.04.018
22. Risticevic S, Lord H, Górecki T, Arthur CL, Pawliszyn J. Protocol for solid-phase microextraction method development. Nat. Protoc. 2010; 5:122–139. DOI: 10.1038/nprot.2009.179 [PubMed: 20057384]
23. Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA, Auinger P, Griffin MR, Poehling KA, Erdman D, Grijalva CG, Zhu Y, Szilagyi P. The Burden of Respiratory Syncytial Virus Infection in Young Children. N. Engl. J. Med. 2009; 360:588–598. DOI: 10.1056/NEJMoa0804877 [PubMed: 19196675]

24. Tranchida PQ, Purcaro G, Dugo P, Mondello L. Modulators for comprehensive two-dimensional gas chromatography. Trac-Trends Anal. Chem. 2011; 30:1437–1461. DOI: 10.1016/j.trac.2011.06.010
25. Tranchida PQ, Maimone M, Purcaro G, Dugo P, Mondello L. The penetration of green sample-preparation techniques in comprehensive two-dimensional gas chromatography. TrAC - Trends Anal. Chem. 2015; 71:74–84. DOI: 10.1016/j.trac.2015.03.011
26. Tranchida PQ, Purcaro G, Conte L, Dugo P, Dugo G, Mondello L. Enhanced resolution comprehensive two-dimensional gas chromatography applied to the analysis of roasted coffee volatiles. J. Chromatogr. A. 2009; 1216:7301–7306. DOI: 10.1016/j.chroma.2009.06.056 [PubMed: 19596333]
27. Purcaro G, Rees CA, Wieland-alter WF, Schneider MJ, Wang X, Stefanuto P, Wright PF, Enelow RI, Hill JE, States U, States U. Volatile fingerprinting of human respiratory viruses. J. Breath Res. 2017 in press h.
28. Hanrahan G, Lu K. Application of factorial and response surface methodology in modern experimental design and optimization. Crit. Rev. Anal. Chem. 2006; 36:141–151. DOI: 10.1080/10408340600969478
29. Weggler BA, Gröger T, Zimmermann R. Advanced scripting for the automated profiling of two-dimensional gas chromatography-time-of-flight mass spectrometry data from combustion aerosol. J. Chromatogr. A. 2014; 1364:241–248. DOI: 10.1016/j.chroma.2014.08.091 [PubMed: 25234498]
30. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics. Anal. Chem. 2006; 78:4281–4290. DOI: 10.1021/ac051632c [PubMed: 16808434]
31. Massart, D. Chemometrics: a textbook. Elsevier Science Ltd; New York: 1988.
32. Breiman L. Random forests. Mach. Learn. 2001; 45:5–32. DOI: 10.1023/A:1010933404324
33. Johnson KJ, Synovec RE. Pattern recognition of jet fuels: comprehensive GC×GC with ANOVA-based feature selection and principal component analysis. Chemom. Intell. Lab. Syst. 2002; 60:225–237. DOI: 10.1016/S0169-7439(01)00198-8
34. Rees CA, Nordick KV, Franchina FA, Lewis AE, Hirsch EB, Hill JE. Volatile metabolic diversity of Klebsiella pneumoniae in nutrient-replete conditions. Metabolomics. 2017; 13:1–11. DOI: 10.1007/s11306-016-1161-z [PubMed: 27980501]
35. Souza Silva EA. Paw, Optimization of Fiber Coating Structure Enables Direct Immersion Solid Phase Microextraction and High-Throughput Determination of Complex Samples. Anal. Chem. 2012; 84:6933–6938. [PubMed: 22834917]
36. Souza-Silva EA, Gionfriddo E, Nazmul A Md, Pawliszyn J. Insights into the E ff ect of the PDMS-Layer on the Kinetics and Thermodynamics of Analyte Sorption onto the Matrix-Compatible Solid Phase Microextraction Coating. Anal. Chem. 2017; 89:2978–2985. DOI: 10.1021/acs.analchem.6b04442 [PubMed: 28192963]
37. Gionfriddo E, Souza-Silva EA, Pawliszyn J. Headspace versus Direct Immersion Solid Phase Microextraction in Complex Matrixes: Investigation of Analyte Behavior in Multicomponent Mixtures. Anal. Chem. 2015; 87:8448–8456. DOI: 10.1021/acs.analchem.5b01850 [PubMed: 26196654]
38. Lenth RV. Response-surface methods in R, using rsm. J. Stat. Softw. 2012; 32:1–17. doi:http://dx.doi.org/10.18637/jss.v032.i07.
39. Bylesjö M, Cloarec O, Rantalainen M. Normalization and Closure. Compr. Chemom. 2010; 2:109–127. DOI: 10.1016/B978-044452701-1.00109-5
40. Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics. 2012; 8:S146–S160. DOI: 10.1007/s11306-011-0350-z
41. Stefanuto P-H, Perrault K, Grabherr S, Varlet V, Focant J-F. Postmortem Internal Gas Reservoir Monitoring Using GC×GC-HRTOF-MS. Separations. 2016; 3:24.doi: 10.3390/separations3030024
42. Stefanuto PH, Perrault KA, Dubois LM, L'Homme B, Allen C, Loughnane C, Ochiai N, Focant JF. Advanced method optimization for volatile aroma profiling of beer using two-dimensional gas chromatography time-of-flight mass spectrometry. J. Chromatogr. A. 2017; 1507:45–52. DOI: 10.1016/j.chroma.2017.05.064 [PubMed: 28587778]

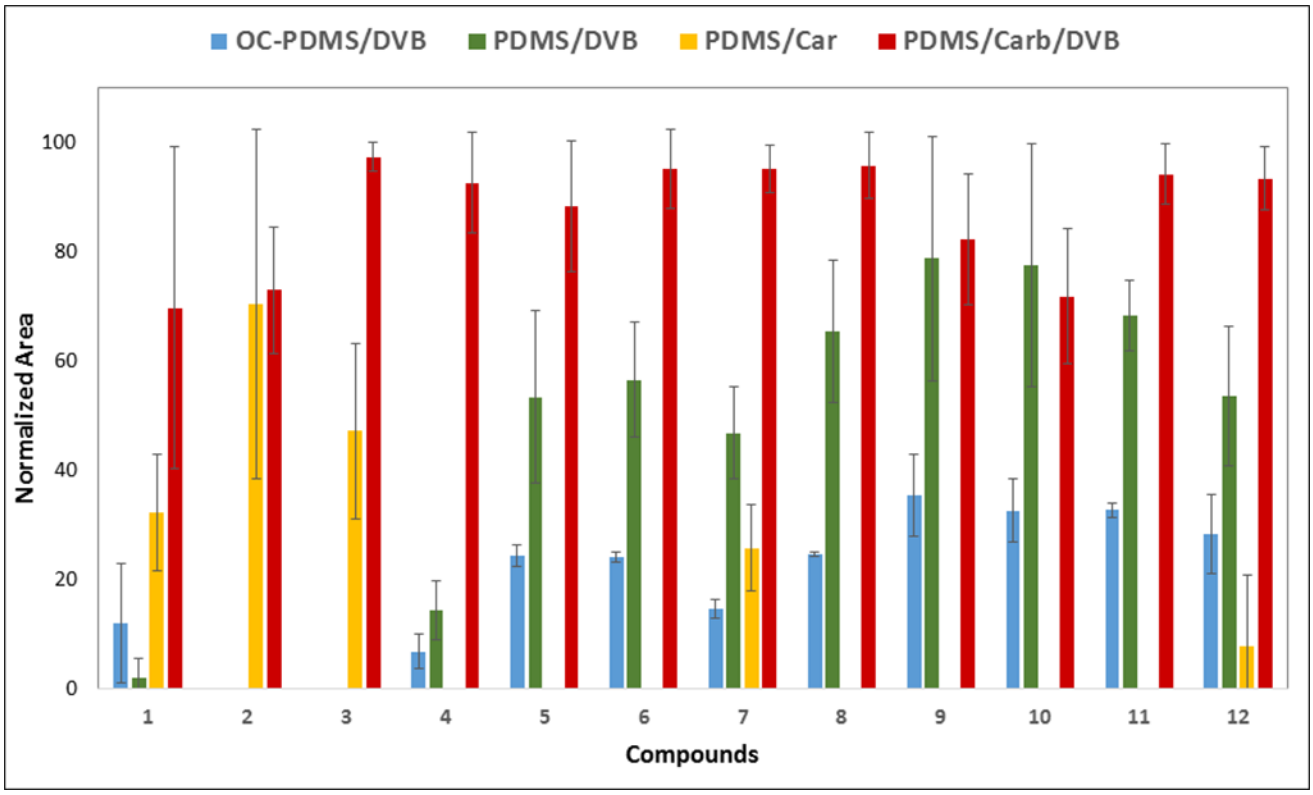

Figure 1.

Comparison of the extraction yields obtained using four different SPME fiber evaluated on 12 compounds. Data normalized against the highest value. Extraction conditions: 43 °C for 30 min (equilibration time: 15 min, stirring: 250 rpm). Compounds identification as in Table 2.

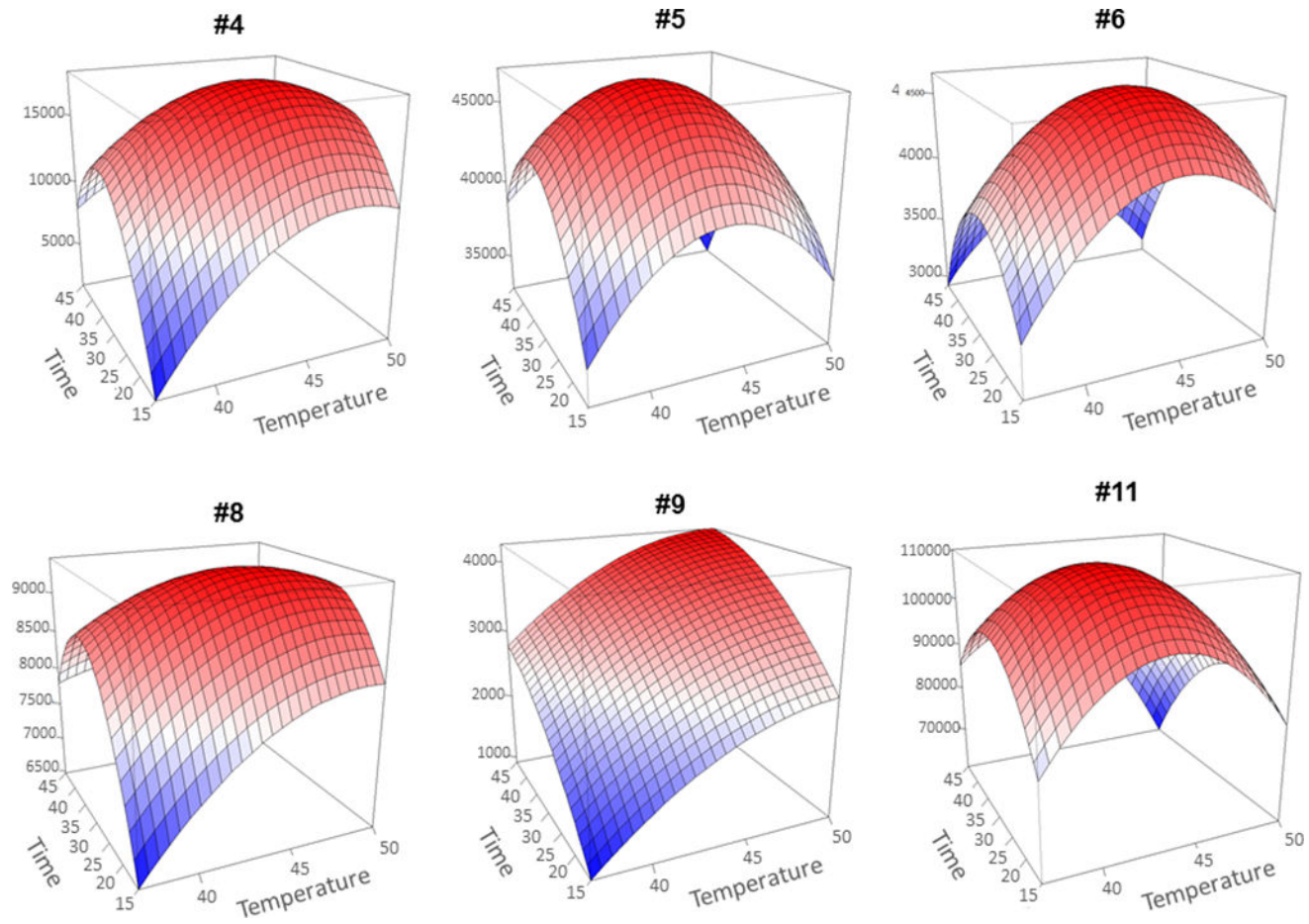

Figure 2.

Response surface obtained from the SPME analysis using the conditions determined with the central composite design (Table 1). Data normalized against the highest value. Fiber used: PDMS/Car/DVB, equilibration time: 15 min, stirring 250 rpm. Compounds identification as in Table 2.

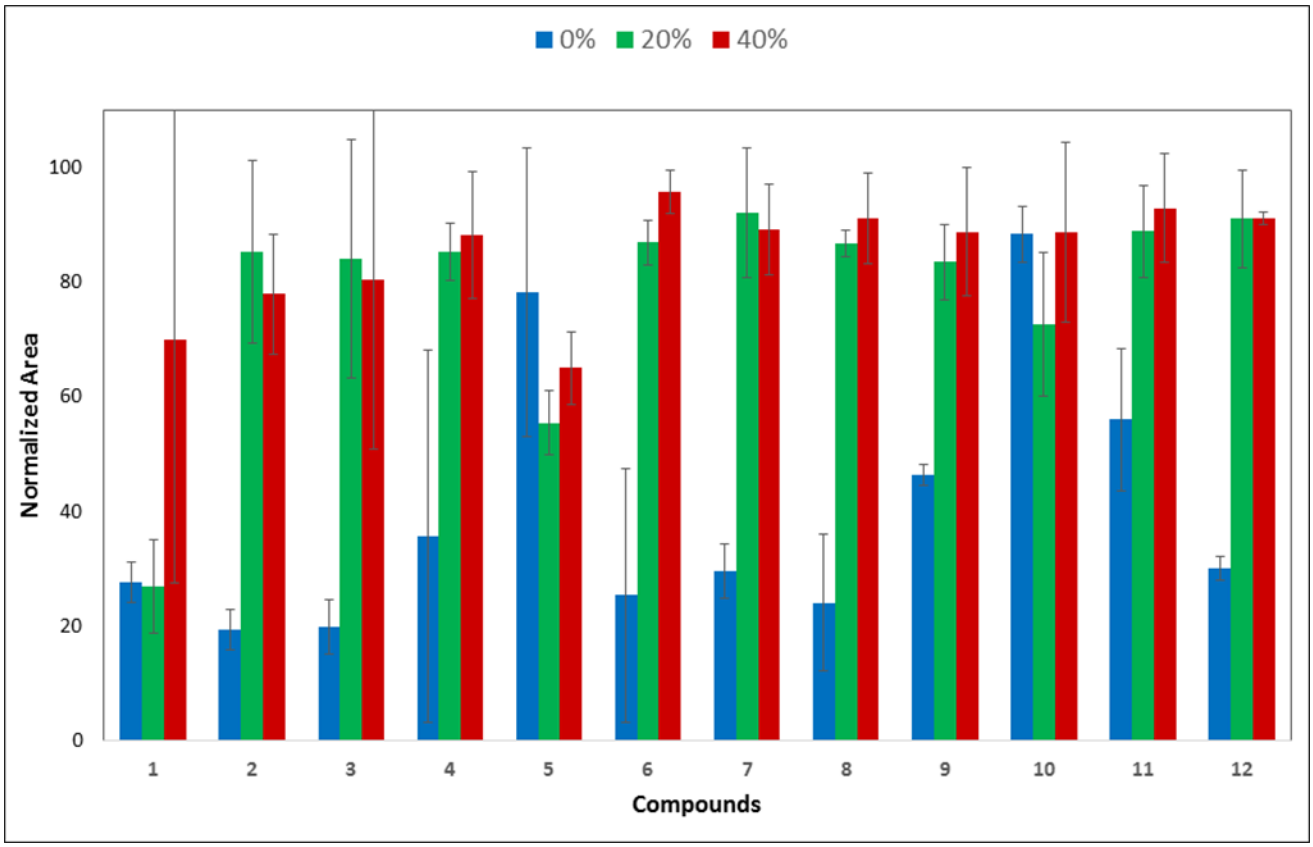

Figure 3.

Comparison of the extraction yields adding different amount of salt (i.e. KCl) on the 12 compounds as identified in Table 2. Data normalized against the highest value. Extraction conditions: 43 °C for 30 min (equilibration time: 15 min, stirring: 250 rpm).

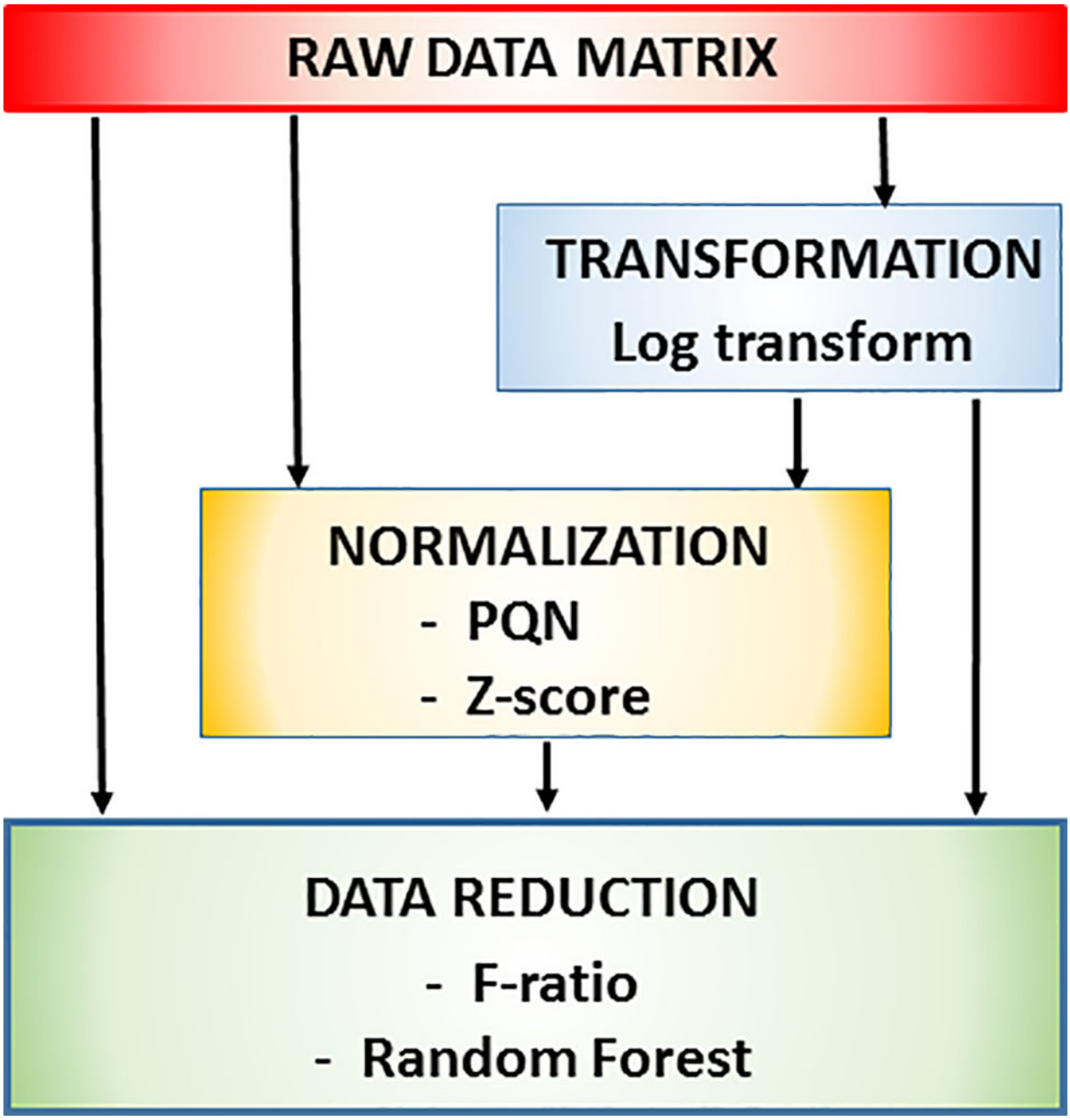

Figure 4.
Scheme of data elaboration comparison.

Table 1

Temperature and time conditions tested for the central composite design (CCD).

#	T (°C)	Time (min)
1	39	19
2	48	19
3	39	41
4	48	41
5*	43	30
6	37	30
7	50	30
8	43	15
9	43	45

* central point

Table 2

List of compounds considered for optimization of the SPME; CAS numbers;experimental (LRI_{Exp}) and the library-derived (LRI_{Lib}) linear retention index; absolute first (^1D) and second (^2D) dimension retention time and their standard deviation.

ID#	Compound	CAS n.	LRI_{Exp}	LRI_{Lib}	^1D (s)		^2D (s)		CV %
					Mean	SD	Mean	SD	
1	Cyclohexane	110-82-7	665	658	231	0.8	1.079	0.025	13.0
2	2-Pentanone	107-87-9	695	689	249	1.8	1.520	0.100	16.0
3	4-methyl- 2-pentanone	108-10-1	739	732	295	1.5	1.588	0.022	2.8
4	2,4-dimethyl-heptane	2213-23-2	819	822	393	0.9	1.131	0.007	9.9
5	Styrene	100-42-5	895	891	508	1.4	2.182	0.038	13.5
6	2,6-dimethyl-nonane	17302-28-2	1012	1022	713	3.1	1.187	0.012	7.6
7	2-ethyl-1-hexanol	104-76-7	1030	1030	749	1.6	2.488	0.027	4.5
8	5-ethyl-2-methyl-octane	62016-18-6	1054	-	791	2.1	1.165	0.013	6.3
9	Nonanal	124-19-6	1107	1107	889	1.3	1.858	0.027	14.6
10	Dodecane	112-40-3	1200	1200	1056	1.3	1.179	0.003	17.1
11	1,3-bis(1,1-dimethylethyl)-benzene	1014-60-4	1252	1249	1144	1.3	1.450	0.007	5.9
12	2,4-bis(1,1-dimethylethyl)-phenol	96-76-4	1511	1513	1552	2.2	1.303	0.051	6.1

Table 3

Linearity range, limit of detection (LOD) and quantification (LOQ), along with the quantifier mass used and the experimental linear retention index (LRI) and the LRI reported in the literature for 36 standards.

Compounds	Quant Mass	LRI_{Exp}	LRI_{Lit}	Linearity range (μg/mL)	LOD	LOQ
Pentanoic acid, methyl ester	74	822	821	0.9–100	0.9	3.1
1-Hexanol	56	869	867	0.6–100	0.6	2.0
2-Heptanone	58	891	898	0.4–50	0.4	1.2
2-Heptanol	45	903	913	0.1–50	0.1	0.4
2-Octanone	58	991	990	1.9–50	1.9	6.3
Decane	57	1000	1000	1.2–100	1.2	4.0
2-Octanol	45	1004	1004	1.0–100	1.0	3.3
Heptanoic acid, methyl ester	74	1024	1025	0.2–100	0.2	0.6
Butyl-benzene	92	1058	1058	0.2–100	0.2	0.7
1-Octanol	56	1074	1076	1.7–50	1.7	5.6
2-Nonanone	58	1092	1093	0.1–100	0.1	0.2
Undecane	57	1100	1100	1.8–100	1.8	6.1
2-Nonanol	45	1104	1105	0.4–100	0.4	1.5
Octanoic acid, methyl ester	74	1123	1125	0.1–100	0.1	0.3
1-Nonanol	56	1174	1176	0.1–100	0.1	0.3
2-Decanone	58	1194	1196	0.1–100	0.1	0.4
Dodecane	57	1200	1200	1.1–50	1.1	3.7
2-Decanol	45	1203	1211	0.1–50	0.1	0.5
Nonanoic acid, methyl ester	74	1223	1224	0.1–100	0.1	0.3
Hexyl-benzene	92	1263	1260	0.7–100	0.7	2.3
1-Chlorodecane	91	1269	1264	0.3–100	0.3	1.1
1-Decanol	56	1276	1278	0.1–100	0.1	0.3
2-Undecanone	58	1294	1294	0.1–100	0.1	0.5
Tridecane	57	1300	1300	0.6–100	0.6	2.0
Decanoic acid, methyl ester	74	1323	1327	0.1–100	0.1	0.4
Heptyl-benzene	92	1368	-	0.6–50	0.6	1.9

Compounds	Quant Mass	LRI_{Exp}	LRI_{Lit}	Linearity range (μg/mL)	LOD	LOQ
1-Chloroundecane	91	1371	-	1.8–100	1.8	5.9
1-Undecanol	56	1377	1379	0.5–100	0.5	1.8
Undecanoic acid, methyl ester	74	1423	1423	0.1–100	0.1	0.2
Octyl-benzene	92	1472	1468	1.1–50	1.1	3.8
1-Chlorododecane	91	1474	-	5.3–100	5.3	17.7
1-Dodecanol	56	1477	1476	0.4–100	0.4	1.5
Pentadecane	57	1500	1500	0.7–100	0.7	2.4
Dodecanoic acid, methyl ester	74	1524	1527	0.5–50	0.5	1.6
Nonyl-benzene	92	1576	1586	0.9–50	0.9	2.9
Hexadecane	57	1600	1600	0.4–100	0.4	1.4

Table 4

Normalized Euclidean distances obtained with the different normalization techniques (None, PQN, z-score), with and without log transformation, and features selection (F-ratio and Random Forest). Resolution is calculated as the ratio between the sum of the intra-group Euclidean distance (Control and RSV) and the inter-group Euclidean distance.

Normalization	Transformation	Features Selection										
		None				F-ratio				Random Forest		
		Euclidean distance			Features selected	Euclidean distance			Features selected	Euclidean distance		
		Control	*RSV*	*Resolution*		*Control*	*RSV*	*Resolution*		*Control*	*RSV*	*Resolution*
None	NO	0.86	0.67	*0.66*	13	0.35	0.38	*1.37*	7	0.43	0.38	*1.24*
	LOG	1.08	0.91	*0.50*	13	0.70	0.62	*0.76*	7	0.10	0.35	*2.22*
PQN	NO	0.84	0.57	*0.71*	16	0.19	0.41	*1.66*	7	0.31	0.43	*1.35*
	LOG	1.09	0.91	*0.50*	5	0.19	0.29	*2.10*	7	0.07	0.35	*2.41*
z-score	NO	0.86	0.66	*0.66*	7	0.60	0.52	*0.90*	7	0.43	0.38	*1.23*
	LOG	1.05	0.92	*0.51*	4	0.24	0.16	*2.54*	7	0.27	0.17	*2.24*

Table 5

Tentatively identification of the discriminatory features selected with F-ratio or Random Forest applying or not normalization (PQN or z-score) and log transformation.

ID#	Compound name	CAS	LRI_{Exp}	LRI_{Lit}	MS%	F-ratio						Random Forest					
						No Norm		PQN		z-score		No Norm		PQN		z-score	
						No LOG	LOG	No LOG	LOG	No LOG	LOG	No LOG	LOG	No LOG	LOG	No LOG	LOG
1	3-methyl-2-butanone	563-80-4	668	673	863	x*	x*	x*									
2	2-oxo-propanoic acid	127-17-3	682	696	825	x*	x*	x*									
3	dimethyl-disulfide	624-92-0	742	747	962	x	x	x	x	x	x	x	x	x	x	x	x
4	2-methyl-3-pentanone	565-69-5	749	742	830	x*	x*	x*									
5	butyl ester-acetic acid	123-86-4	816	813	882	x*	x*	x*									
6	Unknown		836	-	-	x	x	x									
7	1-propoxy-2-propanol	1569-01-3	847	-	879	x	x	x	x	x	x	x	x	x	x	x	x
8	Cyclohexanol	108-93-0	895	891	948					x		x	x	x	x	x	x
9	butyl ester-propanoic acid	590-01-2	910	910	804	x*	x*	x*									
10	Benzaldehyde	100-52-7	970	960	928			x	x	x		x	x	x	x	x	x
11	Unknown		983					x									
12	á-Methylstyrene	98-83-9	986	994	925	x	x	x	x	x	x	x	x	x	x	x	x
13	1-ethyl-3,5-dimethyl-benzene	934-74-7	1062	1065	857	x*	x*	x*									
14	á,á-dimethyl-benzenemethanol	617-94-7	1093	1089	965	x	x	x		x	x	x	x	x	x	x	x
15	Unknown		1095			x*	x*	x*									
16	4-methyl-4-penten-2-one	3744-02-3	1112	1110	750			x	x	x		x	x	x	x	x	x

*feature without a experimental F-value