Use of ion mobility-high resolution mass spectrometry in metabolomics studies to provide near MS/MS quality data in a single injection

Rubén Gil-Solsona [a#], Juan V. Sancho [a#], Anne-Laure Gassner [b], Céline Weyermann [b], Félix Hernández [a], Olivier Delémont [b], Lubertus Bijlsma [a*]

[a] Analytical Chemistry and Public Health, Research Institute for Pesticides and Water (IUPA). Avda. Sos Baynat, s/n. University Jaume I, 12071 Castellón, Spain.

[b] Ecole des Sciences Criminelles, Université de Lausanne, 1015 Lausanne, Switzerland

shared first authorship

*Corresponding author

Lubertus Bijlsma (ORCID: 0000-0001-7005-8775), Analytical Chemistry and Public Health, Research Institute for Pesticides and Water, University Jaume I, Avda Sos Baynat s/n, 12080 Castellón, Spain. E-mail address: bijlsma@uji.es

Abstract

The use of ion mobility separations (IMS) in metabolomics approaches has started to be deeply explored in the last years. In this work, the use of liquid chromatography (LC) coupled to IMS-quadrupole time-of-flight mass spectrometry (QTOF MS) has been evaluated in a metabolomics experiments using single injection of the samples. IMS has allowed obtaining cleaner fragmentation spectra, of nearly tandem MS quality, in data-independent acquisition mode. This is much useful in this research area as a second injection, generally applied in LC-QTOF MS workflows to obtain tandem mass spectra, is not necessary, saving time and evading possible compound degradation. As a case study, the smoke produced after combustion of herbal blends used to spray synthetic cannabinoids has been selected as study matrix. The smoke components were trapped in carbon cartridges, desorbed and analyzed by LC-IMS-QTOF MS using different separation mechanisms (reversed phase and HILIC) and acquiring in both positive and negative mode to widen the chemical domain. Partial Least Squares – Discriminant Analysis highlighted several compounds, and ratio between N-Isopropyl-3-(isoquinolinyl)-2-propen-1-amine and quinoline allowed differentiating between tobacco and herbal products. These two compounds were tentatively identified using the cleaner fragmentation spectra from a single injection in the IMS-QTOF MS, with additional confidence obtained by retention time (Rt) and collisional cross section (CCS) prediction using artificial neural networks. Data from this work show that LC-IMS-QTOF is an efficient technique in untargeted metabolomics, avoiding re-injection of the samples for elucidation purposes. In addition, the prediction models for Rt and CCS resulted of help in the elucidation process of potential biomarkers.

Keywords: Omics approaches, Ion Mobility, High Resolution Mass Spectrometry, herbal blends smoke

1. INTRODUCTION

Untargeted metabolomics has proven to be a powerful analytical approach in different research fields. Its workflow is based on discovering unexpected/unknown compounds that can be used as markers for differentiation of two or more groups using advanced statistical analysis. To this aim, a combination of powerful techniques with bioinformatics and multivariate statistics is used, initially developed for studying metabolite levels in the metabolic cascade of biological scenarios [1]. However, it has rapidly extended to other analytical research fields, such as food analysis [2] , drug metabolism [3], breath analysis [4] and environment [5], among others. It enables dealing with complex matrices, emphasising low concentrated substances (e.g. metabolites, xenobiotics) among a high number of components. The highlighted compounds are annotated based on information provided by the analytical techniques employed such as e.g. accurate mass and/or tandem mass spectra when using mass spectrometry.

In this workflow, the "Achilles heel" is probably the elucidation process of unknown compounds. The combination of separation techniques, such as liquid- and gas chromatography, with powerful high-resolution accurate-mass analysers (HRMS), has improved selectivity, and especially sensitivity, compared to more classical approaches, such as NMR. Despite the strong potential of this combination, the elucidation of highlighted compounds is still a challenging and time-consuming task. Using HRMS, different tools are available nowadays, such as mass spectra databases and in-silico fragmentation, which help assigning possible chemical structures to the candidates. Most databases, such as METLIN [6], contain spectra of biological compounds naturally occurring in animals or plants, and facilitate the tentative identification of the unknown compounds. Thus, just a small number of reference standards needs to be acquired by the laboratory to confirm the identification, as this process is limited to only those compounds that have been tentatively identified. However, current databases are far from being complete and therefore candidate compounds are often missing, especially when they are the result of transformation processes (e.g. degradation, combustion, oxidation, metabolization). In addition,

reference standards might not be commercially available, so one can only rely on tentative identifications based on well-defined criteria [7].

The recent introduction of ion-mobility separations (IMS) in the core of HRMS instruments [8] allows to achieve higher confidence in tentative identifications [9]. IMS separates ionized molecules by their drift time, providing an extra separation dimension to retention time (RT) and accurate mass, which is of great value for a reliable identification. One of the drawbacks of elucidation processes is the possible need to reinject the samples for obtaining accurate tandem mass spectra. So, extra work of re-analysing and additional data treatment is necessary. Besides, the low amount of sample available in some metabolomics experiments may limit the number of injections, and for long analysis batches, the possibility of compounds degradation must be also taken into account. The advantage of IMS-QTOF MS acquisitions is that reinjection can be avoided, as clean fragmentation spectra are obtained in the first injection with near MS/MS quality. Furthermore, the introduction of novel prediction tools, e.g. using artificial neural networks (ANN) for prediction of chromatographic retention time [10] and collisional cross section (CCS) values i.e. derived from IMS drift time [11,12], provides an extra power for reliable tentative identifications. The use of these machine-based prediction tools can reduce the number of possible candidates drastically. The potential of IMS for identification purposes has been illustrated for reported compounds, such as lipids or homemade explosives among others [13] [14] [15]. Hence, IMS appears as a promising tool to be further explored in omics approaches [16–18].

In this work, we show the additional value of combining ultra-high performance liquid chromatography (UHPLC) with IMS and HRMS in untargeted metabolomics studies. To this aim, smoke produced in the combustion of tobacco and other herbs has been selected as a case study. The herbs under study are known to be often used in spice products [19], hence representing the type of product/smoke to which spice consumers might be exposed. This study aims at highlighting and identifying unknown markers of herbs after combustion based on a single injection in a UHPLC-IMS-HRMS system. The identification of

pyrolytic compounds is of interest to understand possible related health effects and to be used as markers of synthetic cannabinoids (SCs) consumption [20]. New RT and CCS predictors have been also developed to reduce the number of possible candidates in the tentative identification of unknown compounds, thus yielding increased confidence in the annotation process.

2. MATERIALS AND METHODS

2.1. Chemicals and samples

HPLC-grade water was obtained by purifying demineralized water in a Milli-Q plus system from Millipore (Bedford, MA, USA). HPLC-grade acetonitrile (ACN), dichloromethane (DCM), methanol (MeOH) and ammonium acetate (NH_4Ac) were obtained from Scharlab (Barcelona, Spain). Leucine-enkephalin, formic acid (HCOOH, 98 - 100 %) and quinoline (98 % purity) were purchased from Sigma-Aldrich (Darmstadt, Germany).

Fourteen herbs mainly smoked in spice products: *Cannavalia Maritima, Nymphaea Alba, Scutellaria Lateriflora, Zornia Latifolia, Nelumbo Nucifera, Leonurus Sibiricus, Althaea Officinalis, Turnera Diffusa, Verbascum Thapsus, Trifolium Pratense, Claendula Officinalis, Leonotis Leonurus, Astragallus Root* and *Rosa Canina* were purchased from Worldherbals (Vlaardingen, The Netherlands). Tobacco from three different trademarks (*Domingo, Fortuna* and *Camel*) were purchased from a local tobacco shop.

2.2. Sample preparation and treatment

All the fourteen mainly employed herbs as well as three different tobacco brand samples (0.5 g of each one) were rolled in cigarettes and coupled to an SPE cartridge (ENVI-Carb®, Sigma-Aldrich), previously conditioned with 6 mL of MeOH and 6 mL of DCM. All cigarettes were rolled with the same cigarette paper and no filter was employed, in order to avoid the introduction of new variables into the experiment. Cigarettes were lighted and smoked through the SPE cartridge under vacuum. After, each

cartridge were eluted with 6 mL MeOH:DCM (20:80 v/v), it was brought nearer to dryness under vacuum using a MiVac Duo concentrator (Genevac, United Kingdom) at low temperature (40°C, 45 min) in order to minimize losses during this step, and reconstituted with 4 mL of MeOH. All the different herbs and tobacco extractions were carried out by triplicate, obtaining a total amount of 51 sample extracts.

A 0.2 mL aliquot was mixed with 1.8 mL Milli-Q water for Reversed Phase (RP) analysis and a second 0.2 mL aliquot with 1.8 mL ACN for HILIC analysis. Quality Control (QC) samples were also prepared by pooling all the extracts together creating an average one which allows to normalize sample signals in experiments where compounds of interest are not selected before the experiment.

2.3. Instrumentation.

A Waters Acquity UPLC system (Waters, Milford, MA, USA) was interfaced to an Ion mobility hybrid Quadrupole-Time of Flight (TOF) High Resolution Mass Spectrometer (UHPLC-IMS-HRMS, VION QTof, Waters, Manchester, UK) using an electrospray interface operating in both positive and negative ionization modes. Resolution of the TOF MS was approximately 40000 at full width half maximum (FWHM).

2.4. Instrumental conditions

2.4.1. IMS-QTOF MS analysis

Electrospray (ESI) was employed as interface, for which capillary voltage was set at 0.7 kV for ESI positive and 1.5 kV for ESI negative ionization modes respectively and 25 V were set as cone voltage. Source temperature was set at 130 °C. N_2 was employed as desolvation gas with a flow of 800 L h-1 heated at 550 °C. Argon was employed as collision gas (Purity 99.995 %, Carbagas, Lausanne, Switzerland). For IMS-QTOF experiments in high definition MS^E ($HDMS^E$) mode, with two acquisition

functions were configured, with different collision energies: Low energy function (LE), selecting 6 eV and high energy function (HE) with a ramp of collision energies from 15 to 40 eV. MS data were acquired over an *m/z* range of 50-1200 Da.

Equipment control and data acquisition were performed with UNIFI v1.8.2 software (Waters, UK). Finally, external calibrations of mass and drift time curves were conducted weekly with the "Major Mix IMS/Tof calibration kit" directly purchased from Waters, prepared and infused at a flow rate of 20 µL min-1 for both positive and negative mass axis calibrations as well as CCS calibration. For internal lock mass calibration, a Leucine-Enkephalin solution (50 ng mL-1) in ACN:H_2O (50:50 v/v) at 0.1 % HCOOH was pumped at 10 µL min-1 through the lock-spray needle and measured every 30 seconds, with a scan time of 0.4 seconds. Leucine-enkephalin, in positive and negative mode was used for recalibrating the mass axis during the injection and to ensure a robust accurate mass along time. Samples were injected in both positive and negative ionization modes.

First 10 samples injected were QC samples, employed to stabilize the column, and an extra QC sample was injected every 10 samples. These QC samples injected along the batch, helps to control that all the sequence have been correctly injected without signal failures, by observing all QCs grouped in the center of the PCA Scores-Plot.

2.4.2. UHPLC analysis

Two different UHPLC separations were performed in order to cover a wide range of compound polarities. Reversed Phase Liquid Chromatography (Phenomenex Kinetex 2.6 µm C_{18} 100Å, 2.1x100 mm fused core column) was used to separate semi-polar compounds while Hydrophilic Interaction Liquid Chromatography (HILIC) (CORTECS® HILIC 2.7 µm, 2.1x100 mm fused core column) was used for polar compounds analysis. Gradients and conditions are shown in **Table 1**.

So, four different datasets were obtained, for reversed phase (RP) and HILIC (HI) separations in both positive (RP+, HI+) and negative (RP-, HI-) ionization modes.

2.5. Data processing

Data was exported from UNIFI in *uep* (unifi export package) format into four different data tables. Progenesis QI, provided by Non-linear dynamics is, at this moment, the only data processing software able to interpret this file format. This program guides the user to import data, selecting a reference sample (in this case a QC sample) in order to correct retention time. This QC is equivalent to the use of external standards in target analysis, with the main benefit that represents all the samples in the set. The use of UHPLC-IMS-HRMS data provides extra separation information to the experiment, helping to better isolate all the compounds present in the samples. For this reason, four-dimensional data is then obtained (Retention time, CCS value, *m/z* and intensity) and data treatment software should be able to understand and work with this 4D data. Data were imported with automatic peak picking and selecting as reference for retention time alignment the last QC from the ten injected for stabilizing the column at the beginning of the sequence. Samples were divided into groups (*QC, Herb* and *Tobacco*) in the "*Experiment Design Setup*" step and finally data was exported to Excel format containing for each detected feature, its *m/z* ratio, RT, CCS and abundance.

After export process, feature labels were manually modified to "Mxxx.xTyyy.yCzzz.z_AAA", being xxx.x the nominal Mass, yyy.y the retention Time (in seconds), zzz.z the CCS value and AAA the chromatographic column and ionization mode (RP+, RP-, HI+ or HI-), obtaining four different datasets for all the four different chromatography/ionization mode. Data abundances were log2 transformed and Pareto scaling was applied, giving the same weight to all the ions [21].

2.6. Statistical analysis

Multivariate analysis was carried out with SIMCA 14 (Umetrics, Sweden). Data was first analysed by Principal Component Analysis (PCA) in order to ensure that QC samples, injected in the initial part of the batch (for column stabilization) and every 10 samples (to control the possible instrumental drift along the time) are joined together in the middle of the plot as well as to eliminate possible outliers. Then, PLS-DA was performed in order to extract a small group of markers to differentiate between herb and tobacco samples. CV-ANOVA was calculated to ensure that groups had strong differences between them (p-value < 0.05). Then, with the Bi-plot, ions placed nearer the herb and tobacco mean point (with VIP value higher than 1) were selected to elucidation.

2.7. Elucidation workflow

Accurate masses for the most significant ions from PLS-DA were retrieved from the feature table. Then, from HDMS^E low energy, the parent compound was assigned (observing different adducts formed with Na^+, NH_4^+ or K^+ for example). From high energy data, filtered with drift time to eliminate product ions generated from different coeluting precursor ions, fragment ions were retrieved. With this list of ions, the precursor ion was searched in different spectral databases (Metlin [6], Massbank [22]), in-silico fragmentation web resources applied (MetFrag [23]) and after additional searching in a chemical database (Chemspider [24]), the compounds were tentatively elucidated.

In order to evaluate the confidence in the identifications, tandem mass spectrometry experiments (MS/MS or MS^2) were also performed at different collision energies (10, 20 and 30 eV), in order to widen the product ions list and compare with the fragment ion list from HE HDMS^E spectra.

When reference standards were available, they were purchased and injected to confirm their identity, and thus their presence in the samples. When unavailable, RT and CCS values were predicted using the RT [10] and the CCS prediction tools [11] in order to provide extra confidence to the tentative elucidation of targeted compounds. The predictions of RT and CCS values were made using artificial

neural networks (ANN), i.e. software of Alyuda NeuroIntelligence 2.2 (Cupertino, CA). Both predictors were previously developed and optimized by using different algorithms and data of 544 (RT) and 205 (CCS) small molecules. Molecules were partitioned into training−verification−blind test sets in the ratio 68:16:16. The final network designs selected for RT and CCS were four-layer perceptrons, 16−19−9−1 and 8−2−8−1, respectively. More details of the predictors can be found elsewhere [10,11].

3. RESULTS AND DISCUSSION

3.1. Importance of separation techniques in non-targeted metabolomics approaches

In non-targeted approaches, the separation of the sample compounds is normally performed by means of chromatography, mass accuracy and fragmentation, but only few studies apply IMS. Commonly, Reversed Phase Liquid Chromatography (RPLC) (e.g. with C18-endcapped columns) coupled to HRMS is used for separation of the sample components. The use of chromatographic columns with orthogonal separations (C-18 for less-polar analysis and HILIC for polar compounds separation [25]), widens the polarity coverage of the analysis, increasing the chemical space and the amount of information obtained. The IMS benefits come from the use of drift time separations [17,26,27], which provides extra help in terms of compounds isolation.

In metabolomics approaches performed with LC-QTOF MS instruments, Data Independent Acquisition (DIA) mode is commonly used to obtain fragmentation information at the same time than the full scan acquisition. DIA allows acquiring spectral information at low (LE) and high (HE) collision energies in a single injection. With this strategy, all the ions that elute at a certain retention time and enter the mass analyser (i.e. Time of Flight) are separated regarding their accurate *m/z*, obtaining information of the non-fragmented compounds (protonated molecule, adducts with small ions) from the LE function and fragment ions from the HE function. The main drawback is that HE spectra often consist of fragment ions from multiple co-eluting compounds, which makes interpretation challenging and complex. Working with IMS-HRMS instruments, co-eluting compounds can be separated by their drift time before

fragmentation in the collision cell. Hence, all the fragment ions belong to the compound of interest as they will share the drift time with its precursor ion. This allows to obtain spectra similar to tandem MS data even for those compounds with lower abundance/intensity. Furthermore, four different descriptors can be used for each ion, i.e. mass accuracy (m/z), intensity, chromatographic retention time and the drift time in the ion mobility cell. The latter can be converted to CCS, which can be used as an additional and instrument independent parameter in the identification of a compound.

This information, allows selecting the most relevant compounds from our experiments (after statistical analysis) and turn back to the previously acquired high quality spectra for obtaining fragment ions, avoiding the re-injection of compounds as well as their possible degradation. Some examples will be shown in the next sections.

3.2. UHPLC-IMS-QTOF MS data treatment

With the aim of testing the capabilities and requirements that four-dimension data (accurate m/z, intensity, chromatographic retention time and drift time) have in metabolomics approaches, the smoke from herbs employed in spiced products was selected as a study case. Fourteen different herbs and three tobacco samples were individually rolled in 0.5 g cigarettes (by triplicate) and lighted (obtaining a total of 51 samples). The smoke trapped in an SPE cartridge was solvent-eluted and injected in the UHPLC-IMS-QTOF MS instrument. A total of 11318 and 4210 ions were obtained in RPLC under positive and negative ionization mode, respectively; while less ions (2212 and 931 in positive and negative ionization mode, respectively) were obtained in the HILIC column. This would confirm that low polarity compounds are majority in the smoke.

Data was exported from UNIFI in **.uep* extension. Progenesis QI was selected for data treatment purposes as only this program can be used for working with 4D data. Data was exported to and excel file and sample groups were introduced. Finally, data was exported for statistical analysis in SIMCA 14.

Finally, features were named "Mxxx.xTyyy.yCzzz.z_AAA", being xxx.x the nominal mass, yyy.y the retention time (in seconds), zzz.z the CCS value and AAA the chromatographic column and ionization mode (RP+, RP-, HI+ or HI-).

3.3. Statistical analysis

Principal Component Analysis (PCA) was performed for each dataset without normalization. A number of QCs, prepared as a pool of samples, were analysed, evaluating if such normalization step was actually required. The use of QCs allows controlling the correct injection of the sample batch, by observing them grouped in the Scores Plot of the PCA, and their composition correspond to an average of all. Thus, QC acts as an "external standard", indicating whether the normalization is necessary, and can be used as reference to align retention times in the rest of the samples. While a normalisation step is commonly applied in metabolomics approaches to compensate small differences in the injection process (e.g. changes in sensitivity of the instrument), the correct QCs grouping in the centre of the PCA plot (see example in **Figure 1**) indicated that normalisation was not required, thus simplifying data treatment.

Other important information obtained from PCA was the detection of possible outliers. As can be seen in **Figure 1**(right part of the scores plot), 6 samples had extreme values. However, all these points corresponded to the same herb (dog rose). Despite their behaviour strongly differed from the rest of the herbs, we decided to keep them in the model, as their differential behaviour did not seem to be related with instrumental variabilities, but with real differences in composition.

Partial Least Squares – Discriminant Analysis (PLS-DA) was carried out to highlight the most discriminative markers between tobacco and herb samples (considering all the herbs as a single group, named HERBS). P-values for CV-ANOVA were calculated for the PLS-DA model, obtaining values lower than 0.05 for RPLC+ ($1.89 \cdot 10^{-35}$) and HILIC+($7.45 \cdot 10^{-27}$) while for RPLC- and HILIC- p-value was 1. So, the

groups seemed to be correctly differentiated under positive ionization mode, while the differentiation was not achieved under negative mode.

From the total amount of 6422 ions in positive mode (from both stationary phases), only a small number were pointed out as different between both groups (HERB and TOBACCO), as shown in **Figure 2** (RPLC+). Those compounds with higher response in Herbs were selected and their spectra extracted from the IMS-QTOF MS ($HDMS^E$) data. These compounds showed a Variable Importance Parameter (VIP) greater than 1, which correlates with the importance of the ion in groups differentiation in PLS-DA model. The complete list of selected compounds with higher area in herbs than in tobacco is shown in **Table 2** and **Table 3**, corresponding to 24 compounds for RPLC+ and 17 for HILIC+.

In order to avoid false characterizations of tobacco samples as herb samples, the ion response was normalized with the area of a second ion. The second ion used for normalization was selected as having high abundance in tobacco smoke and low abundance in the herb extract. Using the ratio between both ion responses, it was possible to correct potential errors in the assignations of potential tobacco markers.

3.4. Benefits of IMS in elucidation

3.4.1. Cleaner spectra in a single injection

Tobacco and herbs smoke were selected as sample matrix within a research performed on consumption of New Psychoactive Substances in Europe. Injection of samples in the LC-QTOF MS system showed several co-eluting compounds, making the spectra interpretation complicated in some cases. So, we evaluated the potential of IMS to obtain tandem MS quality data from our $HDMS^E$ acquisitions. The ion M195.2T270.0C148.2 was selected as an illustrative example on the benefits of IMS to obtain clean DIA spectra. **Figure 3 a1**, shows the co-elution of *m/z* 243.1491, 217.1696, 195.1852 and 183.1851 ions, all with similar intensities. After applying drift time filtering (**Figure 3 b1**), only remained those ions that

shared the same drift time as the selected ion (*m/z* 195.1852). Moreover, the sodium adduct (*m/z* 217.1696, error -2.9 mDa) was maintained (**Figure 3 b1**), supporting that m/z 195.1852 was the protonated molecule. From the HE function, several abundant fragment ions disappeared after filtering (**Figure 3, b2**), while others were still observed (e.g. *m/z* 125.1069). The later seemed to correspond to a neutral loss of C_5H_{10} (70.0780 Da, error 0.35 mDa).

Despite the better quality of the drift time-filtered spectra, it was quite difficult to elucidate m/z 195.1852 due to the limited fragmentation observed which resulted that no good candidate was obtained from Metlin or Massbank. This drawback may occur when the selected potential markers are not included in available spectra databases, independently of whether IMS is used to obtain good quality spectra. Additionally, as HDMSE is only acquired at a single collision energy, the amount of information obtained from this kind of analysis is lower compared with the acquisition of MS/MS data at different collision energiees (e.g. 10, 20, 40 eV) which yield more information, comparing **Figure 4** (b2) (HDMSE spectra) with **Figure 5** (real MS/MS spectra at different collision energies). However, for biological samples (mostly explored in omics approaches experiments), the information provided by LC-IMS-QTOF MS could be enough for obtaining candidates from the available databases and performing a reliable identification.

3.4.2. Selection of candidate marker

After the selection of the most promising compounds for discrimination between herbs and tobacco samples, we tested the possibility to use HE spectra from $HDMS^E$ to elucidate these compounds. The criterion applied, based on the VIP (greater than 1), led to the selection of 24 ions in RPLC+ and 17 ions in HILIC+, which might be considered, in principle, as potential candidates to separate HERB and TOBACCO groups.

Despite that filtering with ion mobility helps obtaining cleaner spectra, the elucidation of unknowns still becomes the most challenging process in non-target metabolomics, as the wide majority of ions i.e.

potential candidates may not be found in databases, and/or their spectra did not give enough information for their complete identification. Thus, some selected ions useful for discriminating samples remained tentatively identified.

This was the case for the feature M227.2T426.2C157.9. After filtering its HE spectra with drift time (**Figure 4 b2**), several low-abundant ions (not related with the highlighted one) were eliminated. Some fragment ions remained with good intensity (*m/z* 183.0915), although other ions with lower response could also be considered as fragment ions (*m/z* 171.0903, 156.0804 or 130.0644). After searching in MetFrag (using Chemspider as chemical database), three candidates presented the highest scoring, with good fitting between the observed spectra and the predicted spectra. As they were not commercially available, we could not confirm which isomer was the marker. In order to assess the quality of the HE spectrum, the product ion spectrum was acquired as shown in **Figure 5**. As can be seen, all the ions observed in the drift time filtered HE spectra were also present in the true MS/MS, apart from new product ions obtained at higher collision energies. Despite MS/MS gives extra information, as different collision energies can be applied to the isolated precursor ion, the whole list of possibilities can also be reduced without extra injections using the $HDMS^E$ spectra after drift time filtering.

In order to improve the confidence in the tentative elucidation of marker *m/z* 227.1537 with a RT of 7.10 min and CCS value of 157.9 $Å^2$, RT and CCS were predicted. Predicted CCSs were 155.8 $Å^2$ (-1.3% error), 155.1 $Å^2$ (-1.8 % error) and 156.1 $Å^2$ (-1.1% error) for (2E)-N-Isopropyl-3-(6-quinolinyl)-2-propen-1-amine (Chemspider ID: 47250370), (2E)-N-Isopropyl-3-(4-isoquinolinyl)-2-propen-1-amine (Chemspider ID: 50494419) and (2E)-N-Isopropyl-3-(8-quinolinyl)-2-propen-1-amine (Chemspider ID: 52240184), respectively, really close to the experimental one. The predicted RT was 6.03 min for the three candidates, close to the experimental value of 7.10 min (-9% error).

Additionally, in certain cases the use of a single marker maybe is not enough to avoid false positive sample assignments. For example, in our case, when a tobacco extract is highly concentrated, the response for marker M227.2T426.2C157.9 could bring to a false positive assignation, as this compound is also present in tobacco but at lower concentrations. In order to solve this fact, the compound with better VIP value and higher intensity in tobacco (compared to herb extract) was selected. From the list of ions higher in tobacco than in herb extracts (see **Table 4**), the *m/z* 130.0646 was selected, being its best molecular formula $C_9H_8N^+$ (error 0.6 mDa). In this case, as can be seen in **Figure S1**, we obtained 2 fragment ions (m/z 117.0567 ($C_8H_7N^+$, error 0.7 mDa) and 103.0537 ($C_8H_7^+$, error 1.1 mDa). After searching these masses in MetFrag, it was tentatively identified as quinoline. In order to ensure the identity of this compound, we performed MS^2 experiments (**Figure S2**, a, b and c), and we only observed an extra product ion at m/z 77.0384 (error 0.5 mDa). Then, quinoline standard was purchased and the identity of the compound was fully confirmed with the MS/MS spectra and Retention Time (**Figure S2**, d, e and f).

Similar to feature M227.2T426.2C157.9, for M130.1T85.6C121.4 the predicted CCS values presented much lower errors than predicted RT, giving more confidence to the identification process. The higher error observed for RT could be explained by the use of a slightly different reversed phase column during ANN prediction model building.

From the whole list of significant compounds, we selected the ratio between ions at *m/z* 227.1537, RT 7.10min and CCS value of 157.9 $Å^2$ (in RP, **Table 2**) and *m/z* 130.065, RT 1.42 min and CCS value of 121.4 $Å^2$ (in RP, **Table 4**) for differentiating herb and tobacco smoke. As observed in **Figure S3**, ratio between both compounds allowed to differentiate both groups, with values for HERB group between 2.21 and 236.76 (78±53), while for TOBACCO group varied from 0.07 to 1.53 (0.39±0.36).

Despite the difficulties above mentioned to elucidate compounds in matrices less explored as the smoke from herb combustion, feature M130.1T85.6C121.4 was confirmed to be quinoline with a

reference standard, whereas the feature M227.2T426.2C157.9 was just tentatively identified due to the lack of reference standard available for this substance, obtaining a reduced list of 3 positional isomers. The identification of markers helps to ensure that selected compounds naturally occur in our samples and do not come, for example, from a contamination.

4. CONCLUSIONS

This study illustrates the capabilities of novel UHPLC-IMS-QToF-based omics approaches to obtain high quality MS/MS data in DIA mode after single injection of the samples. Tentatively elucidated biomarkers, not commercially available, have been confidently annotated using CCS predicted values, being an innovative way to highlight and elucidate unknown compounds in “poorly known matrices”. Compared with other biological samples, where the majority of the matrix compounds composing the matrix overlap, this is not observed in the smoke produced during combustion of different herbs. Samples generated a heterogeneous group with many intra-group differences, making it difficult to find out similarities between them. However, the extended sensitivity of IMS-QTOF MS instrument coupled to the high selectivity of UHPLC made it possible to treat sample sets without normalization, helping to obtain real differences. Two compounds have been discovered, whose ratio revealed as a good approach to differentiate tobacco and herbs used to prepare spice mixtures. The ratio between both markers, varied from an average of 78±53 for herbs to 0.39±0.36 for tobacco, and is suggested as a useful indication of herb smoke, linked to the potential consumption of spice products. This finding should be further studied on street samples or police seizures.

ACKNOWLEDGMENTS

Ruben Gil-Solsona acknowledge the financial support from SCORE-COST action ES1307 ”Sewage biomarker analysis for community health assessment” for their Short Term Scientific Mission (STSM). Lubertus Bijlsma acknowledges NPS-Euronet, (HOME/2014/JDRUG/AG/DRUG/7086) funded with support from the European Commission, for his post-doctoral fellowship. This communication reflects the views only of the authors, and the European Commission cannot be held responsible for any use that may be made of the information contained therein.

FUNDING

This work has been supported by Generalitat Valenciana (Group of Excellence Prometeo II 2019/040); Spanish Ministry of Economy and Competitiveness (Project CTQ2015-65603-P); University Jaume I (UJI-B2016-10); NPS-Euronet (HOME/2014/JDRUG/AG/DRUG/7086) funded with support from the European Commission; and SCORE-COST action ES1307 (COST-STSM-ES1307-150916-080342).

REFERENCES

[1] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics., Mass Spectrom. Rev. 26 (2007) 51–78. https://doi.org/10.1002/mas.20108.

[2] J.M. Cevallos-cevallos, E. Etxeberria, M.D. Danyluk, G.E. Rodrick, Metabolomic analysis in food science : a review, Trends Food Sci. Technol. 20 (2009) 557–566. https://doi.org/10.1016/j.tifs.2009.07.002.

[3] M. Raro, M. Ibáñez, R. Gil, A. Fabregat, E. Tudela, K. Deventer, R. Ventura, J. Segura, J. Marcos, A. Kotronoulas, J. Joglar, M. Farré, S. Yang, Y. Xing, P. Van Eenoo, E. Pitarch, F. Hernández, J.V. Sancho, Ó.J. Pozo, Untargeted Metabolomics in Doping Control: Detection of New Markers of Testosterone Misuse by Ultrahigh Performance Liquid Chromatography Coupled to High-Resolution Mass Spectrometry, Anal. Chem. 87 (2015) 8373–8380. https://doi.org/10.1021/acs.analchem.5b02254.

[4] M. Couto, C. Barbosa, D. Silva, A. Rudnitskaya, L. Delgado, A. Moreira, S.M. Rocha, Oxidative stress in asthmatic and non-asthmatic adolescent swimmers—A breathomics approach, Pediatr. Allergy Immunol. 28 (2017) 452–457. https://doi.org/10.1111/pai.12729.

[5] D.J. Beale, A. V. Karpe, J.D. McLeod, S. V. Gondalia, T.H. Muster, M.Z. Othman, E.A. Palombo, D. Joshi, An “omics” approach towards the characterisation of laboratory scale anaerobic digesters treating municipal sewage sludge, Water Res. 88 (2016) 346–357. https://doi.org/10.1016/j.watres.2015.10.029.

[6] C. Guijas, J.R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A.E. Aisporna, D.W. Wolan, M.E. Spilker, H.P. Benton, G. Siuzdak, METLIN: A Technology Platform for Identifying Knowns and Unknowns, Anal. Chem. 90 (2018) 3156–3164. https://doi.org/10.1021/acs.analchem.7b04424.

[7] E.L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H.P. Singer, J. Hollender, Identifying small

molecules via high resolution mass spectrometry: Communicating confidence, Environ. Sci. Technol. 48 (2014) 2097–2098. https://doi.org/10.1021/es5002105.

[8] V. D'Atri, T. Causon, O. Hernandez-Alba, A. Mutabazi, J.-L. Veuthey, S. Cianferani, D. Guillarme, Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry?, J. Sep. Sci. (2017). https://doi.org/10.1002/jssc.201700919.

[9] A. Kaufmann, S. Walker, Comparison of linear intrascan and interscan dynamic ranges of Orbitrap and ion-mobility time-of-flight mass spectrometers, Rapid Commun. Mass Spectrom. 31 (2017) 1915–1926. https://doi.org/10.1002/rcm.7981.

[10] R. Bade, L. Bijlsma, T.H. Miller, L.P. Barron, J.V. Sancho, F. Hernández, Suspect screening of large numbers of emerging contaminants in environmental waters using artificial neural networks for chromatographic retention time prediction and high resolution mass spectrometry data analysis, Sci. Total Environ. 538 (2015) 934–941. https://doi.org/10.1016/j.scitotenv.2015.08.078.

[11] L. Bijlsma, R. Bade, A. Celma, L. Mullin, G. Cleland, S. Stead, F. Hernandez, J. V. Sancho, Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis, Anal. Chem. 89 (2017) 6583–6589. https://doi.org/10.1021/acs.analchem.7b00741.

[12] Z. Zhou, J. Tu, X. Xiong, X. Shen, Z.J. Zhu, LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision to Support Ion Mobility-Mass Spectrometry-Based Lipidomics, Anal. Chem. 89 (2017) 9559–9566. https://doi.org/10.1021/acs.analchem.7b02625.

[13] G. Astarita, G. Astarita, Applications of ion mobility MS in metabolomics, Adv. LC-MS Appl. Metabolomics. (2009) 94–109. https://doi.org/10.4155/fseb2013.14.63.

[14] X. Zhang, K. Kew, R. Reisdorph, M. Sartain, R. Powell, M. Armstrong, K. Quinn, C. Cruickshank-Quinn, S. Walmsley, S. Bokatzian, E. Darland, M. Rain, K. Imatani, N. Reisdorph, Performance of

a High-Pressure Liquid Chromatography-Ion Mobility-Mass Spectrometry System for Metabolic Profiling, Anal. Chem. 89 (2017) 6384–6391. https://doi.org/10.1021/acs.analchem.6b04628.

[15] N. Hagan, I. Goldberg, A. Graichen, A. St. Jean, C. Wu, D. Lawrence, P. Demirev, Ion Mobility Spectrometry - High Resolution LTQ-Orbitrap Mass Spectrometry for Analysis of Homemade Explosives, J. Am. Soc. Mass Spectrom. 28 (2017) 1531–1539. https://doi.org/10.1007/s13361-017-1666-3.

[16] C.D. Chouinard, G. Nagy, R.D. Smith, E.S. Baker, Ion Mobility-Mass Spectrometry in Metabolomic, Lipidomic, and Proteomic Analyses, Compr. Anal. Chem. 83 (2019) 123–159. https://doi.org/10.1016/BS.COAC.2018.11.001.

[17] T. Mairinger, T.J. Causon, S. Hann, The potential of ion mobility–mass spectrometry for non-targeted metabolomics, Curr. Opin. Chem. Biol. 42 (2018) 9–15. https://doi.org/10.1016/J.CBPA.2017.10.015.

[18] A. Kaufmann, P. Butcher, K. Maden, S. Walker, M. Widmer, High-resolution mass spectrometry–based multi-residue method covering relevant steroids, stilbenes and resorcylic acid lactones in a variety of animal-based matrices, Anal. Chim. Acta. 1054 (2019) 59–73. https://doi.org/10.1016/j.aca.2018.12.012.

[19] EMCDDA. Understanding the 'Spice' phenomenon [Internet]. 2009, (n.d.). http://www.emcdda.europa.eu/publications/thematic-papers/understanding-spice-phenomenon_en (accessed February 18, 2019).

[20] L. Bijlsma, R. Gil-Solsona, F. Hernández, J.V. Sancho, What about the herb? A new metabolomics approach for synthetic cannabinoid drug testing, Anal. Bioanal. Chem. 410 (2018). https://doi.org/10.1007/s00216-018-1182-8.

[21] R. Di Guida, J. Engel, J.W. Allwood, R.J.M. Weber, M.R. Jones, U. Sommer, M.R. Viant, W.B. Dunn, Non-targeted UHPLC-MS metabolomic data processing methods: a comparative

investigation of normalisation, missing value imputation, transformation and scaling, Metabolomics. 12 (2016) 93. https://doi.org/10.1007/s11306-016-1030-9.

[22] MassBank | European MassBank (NORMAN MassBank) Mass Spectral DataBase, (n.d.). http://massbank.eu/MassBank/Index (accessed February 18, 2019).

[23] C. Ruttkies, E.L. Schymanski, S. Wolf, J. Hollender, S. Neumann, MetFrag relaunched: incorporating strategies beyond in silico fragmentation, J. Cheminform. 8 (2016) 3. https://doi.org/10.1186/s13321-016-0115-9.

[24] ChemSpider | Search and share chemistry, (n.d.). http://www.chemspider.com/ (accessed February 18, 2019).

[25] B. Buszewski, S. Noga, Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique, Anal. Bioanal. Chem. 402 (2012) 231–247. https://doi.org/10.1007/s00216-011-5308-5.

[26] T.J. Causon, V. Ivanova-Petropulos, D. Petrusheva, E. Bogeva, S. Hann, Fingerprinting of traditionally produced red wines using liquid chromatography combined with drift tube ion mobility-mass spectrometry, Anal. Chim. Acta. 1052 (2019) 179–189. https://doi.org/10.1016/j.aca.2018.11.040.

[27] C.M. Nichols, J.N. Dodds, B.S. Rose, J.A. Picache, C.B. Morris, S.G. Codreanu, J.C. May, S.D. Sherrod, J.A. McLean, Untargeted Molecular Discovery in Primary Metabolism: Collision Cross Section as a Molecular Descriptor in Ion Mobility-Mass Spectrometry, Anal. Chem. 90 (2018) 14484–14492. https://doi.org/10.1021/acs.analchem.8b04322.