Talanta 127 (2014) 181–190

Contents lists available at ScienceDirect

Talanta

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

Infrared biospectroscopy for a fast qualitative evaluation of sample preparation in metabolomics

Julia Kuligowski [a,1], David Pérez-Guaita [b,1], Javier Escobar [a], Isabel Lliso [a], Miguel de la Guardia [b], Bernhard Lendl [c], Máximo Vento [a,d], Guillermo Quintás [e,*]

[a] Neonatal Research Group, Health Research Institute La Fe, 46026 Valencia, Spain
[b] Department of Analytical Chemistry, University of Valencia, 46100 Burjassot, Spain
[c] Institute of Chemical Technologies and Analytics, Vienna University of Technology, A-1060 Vienna, Austria
[d] Division of Neonatology, University & Polytechnic Hospital La Fe, 46026 Valencia, Spain
[e] Leitat Technological Center, Bio in vitro Division, 46026 Valencia, Spain

ARTICLE INFO

Article history:
Received 3 February 2014
Received in revised form
1 April 2014
Accepted 4 April 2014
Available online 15 April 2014

Keywords:
Attenuated total reflectance-Fourier transform infrared (ATR-FTIR)
Liquid chromatography-time of flight mass spectrometry (LC-TOFMS)
Plasma
Sample preparation
Metabolomics

ABSTRACT

Liquid chromatography-mass spectrometry (LC-MS) has been increasingly used in biomedicine to study the dynamic metabolomic responses of biological systems under different physiological or pathological conditions. To obtain an integrated snapshot of the system, metabolomic methods in biomedicine typically analyze biofluids (e.g. plasma) that require clean-up before being injected into LC-MS systems. However, high resolution LC-MS is costly in terms of resources required for sample and data analysis and care must be taken to prevent chemical (e.g. ion suppression) or statistical artifacts. Because of that, the effect of sample preparation on the metabolomic profile during metabolomic method development is often overlooked. This work combines an Attenuated Total Reflectance-Fourier transform infrared (ATR-FTIR) and a multivariate exploratory data analysis for a cost-effective qualitative evaluation of major changes in sample composition during sample preparation. ATR-FTIR and LC-time of flight mass spectrometry (TOFMS) data from the analysis of a set of plasma samples precipitated using acetonitrile, methanol and acetone performed in parallel were used as a model example. Biochemical information obtained from the analysis of the ATR-FTIR and LC-TOFMS data was thoroughly compared to evaluate the strengths and shortcomings of FTIR biospectroscopy for assessing sample preparation in metabolomics studies. Results obtained show the feasibility of ATR-FTIR for the evaluation of major trends in the plasma composition changes among different sample pretreatments, providing information in terms of e.g., amino acids, proteins, lipids and carbohydrates overall contents comparable to those found by LC-TOFMS.

1. Introduction

Metabolomics, dealing with the global profiling of a whole set of small molecules or metabolites present in a biological system, is gaining broader recognition and it is increasingly being used to explore the dynamic responses of living systems under different physiological and pathological conditions. Nonetheless, the complexity of the metabolome is outstanding and, strictly speaking, no single analytical technique is suitable for its quantitative or qualitative profiling [1]. Thus, the subset of the metabolome covered is defined by a sample preparation method and an analytical technique is employed. Nuclear magnetic resonance (NMR) is widely used in metabolomic profiling. It provides a wealth of qualitative and quantitative information; limits of detection in the μM range, high reproducibility [2,3] and samples can be typically measured with minimal sample preparation. Nevertheless, instrumentation, especially NMR spectrometers working at high magnetic fields, is still expensive. Since the last few years, the use of liquid chromatography with mass spectrometric detection (LC-MS) is gaining interest in metabolomics [4]. High resolution LC-MS (e.g., UPLC-QqTOF, UPLC-Orbitrap) provides high throughput as well as outstanding levels of resolution, linearity, metabolite coverage and sensitivity with limits of detection (LODs) in the nM range.

Metabolomics in biomedicine relies on biofluids or tissue extracts, plasma being one of the most commonly used (ca. 65% of the LCMS metabolomic studies use this biofluid [5]) as it provides an integrated view of systems biology [6]. Plasma is obtained by collecting whole blood in anticoagulant-treated tubes

* Corresponding author. Tel.: +34 93 788 23 00.
E-mail address: guillermo.r.quintas@uv.es (G. Quintás).
[1] Both the authors have contributed equally in this work.

http://dx.doi.org/10.1016/j.talanta.2014.04.009
0039-9140/

(e.g., with heparin, citrate or ethylenediaminetetraacetic acid (EDTA)) to prevent the fibrinogen clotting cascade and then, removing cells and platelets by centrifugation and makes up to ca. 55% of the total blood volume. The use of plasma instead of serum presents some advantages as the clotting process increases the likelihood of degradation of labile metabolites and also facilitates possible losses and transformations (e.g. oxidation) of metabolites during clot formation and precipitation [5]. Besides, serum provides lower volumes of sample per volume of whole blood collected. However, with plasma protein contents in the range of 6–8 g dL^{-1}, protein removal is mandatory to prevent degradation of the LC column and contamination of the electrospray (ESI) interface.

Sample preparation is critical in MS-based metabolomic experiments. It affects the retrieved metabolomic profile, the data quality and the reproducibility and so, it might determine whether a study will be successful or not. Solvent precipitation is among the most widely used methods for sample preparation for metabolomic profiling of plasma due to its simplicity, high throughput and reproducibility [5]. However, an often overlooked aspect is that the selection of the parameters used for protein precipitation (e.g., solvent, plasma to precipitant ratio, pH) determines the type and amount of protein precipitated [7], the metabolite coverage and the precision levels [5,8,9]. High resolution LC-MS systems generate large multidimensional data sets that require time consuming statistical analysis to extract information [1,8,10] for exploratory or discriminant analysis. Difficulties in the LC-MS analysis of the solvent dependent metabolite changes in plasma or serum samples after protein precipitations are reflected by the number of approaches that can be found in the literature, including the comparison of the total number of detected features [8,9,12,13]; the use of the total ion chromatogram (TIC) traces comparing the number of signals observed and the general sensitivity and reproducibility [13]; the evaluation of the column lifetime prolongation [12]; the monitoring of the signal of isotopically labeled internal standards to monitor data quality and reproducibility [12]; or the comparison of the number of late eluting detected features [14] and the distribution of relative standard deviation (RSD) values calculated across quality control (QC) samples [14]. Moreover, metabolite identification must also be performed as the use of unidentified fingerprints that reduces the repeatability, reproducibility and interpretability of the comparison among sample pretreatments. Initial metabolite identification based on accurate *m/z* values (i.e., without fragmentation spectra) is carried out by matching the *m/z* within a user-defined *m/z* error to a theoretically accurate *m/z* related to a single or multiple molecular formulae [11]. In spite of recent advances in the automation of this process [15], this step is still very time consuming and often unfeasible. An additional difficulty of analyzing LCMS data is that the parameters selected for MS and chromatographic peak detection, de-convolution and alignment must be carefully selected as they have a great effect on the number of retrieved variables, which in turn affects the extracted information.

Therefore, there is an interest in the development of analytical tools for a straightforward evaluation of the main effects of sample pretreatment on the biochemical sample composition.

FTIR spectroscopy is recognized as a valuable tool for the simultaneous analysis of a wide range of biomolecules including polysaccharides, proteins, lipids, fatty acids, amino acids, nucleic acids and small inorganic ions [16]. Although clearly outperformed by LC-MS in terms of sensitivity and specificity, FTIR spectroscopy in the mid-IR region (4000–400 cm^{-1}) still yields an easily accessible fingerprint of complex biological samples [17,18] providing simultaneous multi-parameter information of the proteome, lipidome and metabolome [19]. Potential fields of applications of FTIR spectroscopy in biomedicine include cytology, histology and microbiology and it is expected to be used as an objective and robust tool for cancer diagnosis and screening [20,21,22].

FTIR spectra of liquid samples can be measured using transmission flow cells by adjusting the optical pathlength down to the low μm range or by using attenuated total reflection (ATR) flow cells to circumvent intense solvent absorption [23]. However, these approaches limit both sensitivity and accessible spectral range. Alternatively, a solvent elimination prior to the IR measurement of biofluids simultaneously enhances sensitivity, avoids interfering signals and reduces the sample volumes [19]. IR transmission spectroscopy of dried serum films provided good accuracy levels for the quantitative determination of albumin, total protein, urea, total cholesterol, triglycerides and glucose [24]. IR measurements of dry films of plasma samples have been used for the monitoring of 26 parameters providing a 'metabolic photography' of plasma that could be used for the classification of individuals according to their physiological condition [25]. A similar approach was used for the determination of plasma protein contents by FTIR [26]. Direct ATR-FTIR spectroscopy has been successfully applied for the quantification of major compounds including glucose, lactate, triglycerides, cholesterol, total protein, urea and albumin in serum and plasma through PLS modeling [27,28]. Recently, lipidic parameters were determined in serum by measuring the dry film formed in an ATR crystal after the deposition of the sample organic extracts obtained from a liquid–liquid extraction of serum samples [29].

In the aforementioned frame, this work combines ATR-FTIR spectroscopy and multivariate exploratory data analysis for the evaluation of the effect of different sample preparations in LC-MS metabolomic studies. As a model example, plasma samples precipitated using CH_3CN, CH_3OH and acetone were analyzed by ATR-FTIR and LC-TOFMS analysis in parallel. Data obtained was thoroughly compared to evaluate the strengths and shortcomings of using FTIR spectroscopy during the initial steps of metabolomic method development. Results obtained show the feasibility of ATR-FTIR for a fast and cost-effective evaluation of major trends in the plasma composition changes among different sample pretreatments, providing information in terms of e.g., amino acids, proteins, lipids and carbohydrates overall contents comparable to those found by LC-TOFMS.

2. Materials and methods

2.1. Chemicals and reagents

All solvents were of LC-MS grade and were purchased from Scharlau (Barcelona, Spain). Ultra-pure water was generated with a Milli-Q water purification system from Merck Millipore (Darmstadt, Germany). Deuterated internal standards DL-phenylalanine-D_5, L-methionine-D_3, DL-cystine-D_4 and DL-cysteine-D_2 with a purity of 98% were purchased from Cambridge Isotopes Laboratory Inc. (Andover, MA, USA). N-ethylmaleimide ($\geq$ 98%), formic acid ($\geq$ 95%) and other additives and standards were obtained from Sigma-Aldrich Quimica SA (Madrid, Spain). IsoFlo® (Isoflurane, USP) for anesthetizing animals during intervention was purchased from Laboratorios Dr. Esteve, S.A. (Barcelona, Spain). Sodium heparin (5% v/v) from Laboratorios LEO Pharma S.A. (Barcelona, Spain) was used for heparinizing syringes prior to blood extraction.

2.2. Sample collection and processing

Plasma samples were obtained from three healthy female mice (strain: C57BL6) that were maintained at constant conditions (23 $\pm$ 1 °C, 60% humidity and light/dark cycles of 12/12 h), receiving a standard laboratory diet and water ad libitum. Three

mice were sacrificed at an age of 6 months employing isoflurane at a concentration of 2.5% v/v and an oxygen flow of 0.5–1 L min^{-1} via inhalation as anesthesia during the whole intervention. Blood was extracted with a heparinized syringe from the cava vein, and centrifuged in Eppendorf tubes at 20,000 rpm during 15 min at 24 °C. Plasma samples were stored immediately in an ultra-low temperature freezer at −80 °C. Animal experiments were carried out at the University of Valencia (Spain) in compliance with the legal requirements for animal experiments (European Guidelines for use of Experimental Animals).

For processing, samples were kept on ice until their analysis to prevent degradation. After thawing, plasmas from three mice were mixed and homogenized on a vortex mixer during 10 s. Then, 150 μL of cold (4 °C) CH_3OH, CH_3CN or acetone was added to 50 μL of plasma. For each solvent, three replicate extracts were prepared keeping a constant sample concentration during protein precipitation. It has to be underlined that in this study, one single pooled plasma sample was used. This strategy is employed frequently during the development of metabolomic procedures, as it enables a straightforward comparison between sample pretreatments or analytical conditions excluding other possible sources of variance (e.g. biological variation among individuals). Samples were centrifuged at a speed of 10,000 rpm at 4 °C during 10 min. For LC-TOFMS measurements, 25 μL of supernatant was mixed with 100 μL of an internal standard mixture containing 5 μM Phenylalanine-D_5, 10 μM Cysteine-D_2-NEM and Methionine-D_3 and 50 μM cystine-D_4 in H_2O (0.1% v/v HCOOH). The remaining supernatant was employed for ATR-FTIR measurements without further processing.

Blank extracts were prepared for each solvent by adding 150 μL of cold (4 °C) CH_3OH, CH_3CN or acetone to 50 μL H_2O. Blanks underwent an identical treatment as plasma extracts prior to their analysis. In addition, a second set of blanks ('heparine blanks') was prepared by replacing blood by H_2O in the sample collection protocol.

2.3. ATR-FTIR sample analysis

Infrared spectra were obtained on a Bruker (Bremen, Germany) IFS 66/v FTIR spectrometer equipped with a liquid nitrogen-refrigerated mercury–cadmium–telluride detector, a vacuum system, and a dry air purged sample compartment. Measurements were made using an in-compartment ATR DuraSampleIR accessory with a nine reflections diamond/ZnSe DuraDisk from Smiths Detection Inc. (Warrington, UK). 1 μL of sample was deposited in the center of the ATR crystal. Samples were dried at room temperature for approximately 40 s. After that, the spectrum of each sample was collected co-adding 300 scans in the range between 4000 and 600 cm^{-1} with a resolution of 4 cm^{-1} and a zero filling factor of 2. Three plasma aliquots were used for protein precipitation using acetone, CH_3CN or CH_3OH. In order to evaluate the reproducibility of the measurement, for each precipitation solvent, three dried spots were prepared and for each dried spot, three spectra were acquired. A spectrum of the ATR crystal previously recorded using the same instrumental conditions was used as background. After each measurement, the ATR surface was thoroughly cleaned using H_2O, CH_3CN and the solvent used for protein precipitation. Spectral acquisition order was randomized to avoid biased results due to instrumental effects. ATR-FTIR spectra were baseline corrected using a polynomial function. The wavenumbers used as knots (i.e., points used for the estimation of the polynomial baseline) are depicted in Fig. S-1. PCA of ATR-FTIR data was performed on mean centered negative second order derivative spectra using Savitzky–Golay differentiation (number of points in filter: 5, order of the polynomial: 2; derivative: 2).

2.4. LC-TOFMS sample analysis

Chromatographic analysis of the plasma samples was performed on a 1200 RRLC Series Agilent instrument (Palo Alto, CA, USA.). Sample extracts and blanks were analyzed by triplicate using a Zorbax SB C_8 column (3 × 150 mm, 3.5 μm, Agilent). 100 μL aliquots of sample were transferred into 200 μL capped glass vials. Prior to the LC analysis, samples were diluted 1:5 with H_2O (0.1% v/v HCOOH) containing a mixture of deuterated internal standards. Autosampler and column temperatures were set to 6 °C and 55 °C, respectively and the injection volume was 5 μL. A gradient elution with a total run time of 15 min was performed at a flow rate of 350 μL min^{-1} as follows: initial conditions of 96% of solvent A (H_2O 0.05% v/v HCOOH) were kept for 1 min, followed by a linear gradient from 4% to 96% of mobile phase B (CH_3CN 0.05% v/v HCOOH) for 4 min; isocratic conditions of 96% B were held for 1.5 min and finally, a 0.5 min gradient was used to return to the initial conditions, which were held for 8 min. Mass spectrometry detection was performed using an ABSciex 5600-TripleTOF MS spectrometer (Framingham, MA, USA). The following electrospray ionization parameters were selected in the positive mode (ESI+): ion source gases 1 and 2 were set to 50 a.u., curtain gas flow to 25 L h^{-1}, temperature was set to 450 °C and the ion spray voltage floating and de-clustering potential to 5500 and 100 V, respectively. Full scan data were collected in the TOF MS mode from 70 to 950 mass to charge ratio (m/z) with a scan time of 0.08 s (cycle time: 0.27 s). Sample acquisition was randomized to avoid bias effects of instrument drifts during the LC batch. H_2O (0.1% v/v HCOOH) solution was analyzed for every 8 plasma samples and/or blank extracts for background correction and to monitor the lack of cross-contamination. For quality control, an aqueous standard mixture of endogenous metabolites at a concentration of 10 μM was injected after each blank (data not shown) to monitor instrument performance. A standard QC-sample, three plasma extracts and blank solution were analyzed at the beginning of the batch for column conditioning.

2.5. LC-TOFMS data processing

Continuous raw LC-TOFMS data (.wiff files) were converted in centroid mode to mzML format using MSConverter (ProteoWizard 3.0, http://proteowizard.sourceforge.net/). Then, .mzML files were imported into mzMine 2.10 (http://mzmine.sourceforge.net/). Initially, a mass list was generated by peak detection using a threshold of 200 a.u. in each MS spectrum within 1.5 and 12 min. For chromatogram construction, a minimum span time of 0.05 min and m/z tolerance of 2.5 mDa (or 5 ppm) were selected. Then, the obtained chromatograms were smoothed using the Savitzky–Golay filter (width: 9 data points). After that, chromatograms were de-convoluted to separate them into individual peaks using the baseline cut-off method included in mzMine 2.10, and the following parameters: minimum peak height: 250; peak duration range: 0.05–0.5 min, and baseline level: 150. Finally, peak alignment was performed to match features in the peak lists of the set of samples using the Joint Aligner tool included in mzMine 2.10. This tool aligns features through a match score based on m/z and retention time of each peak using the following parameters: m/z tolerance: 2.5 mDa; weight for m/z: 1; retention time tolerance: 0.5 min; weight for retention time: 1; same charge state among aligned peaks required. The obtained peak table was imported into MATLAB (Mathworks Inc., Natick, MA, USA) for data analysis providing a raw data matrix X_0 with samples in rows and variables in columns. The molecular formula of selected metabolites was estimated using PeakView (ABSciex) software, based on accurate m/z values of the molecular ion and its isotopic profile. After identification of the molecular formula, the Human Metabolome Database (HMDB, http://www.hmdb.ca), MassBank (http://www.

massbank.jp) and Lipid Maps (http://www.lipidmaps.org/) open databases were employed for the putative metabolite identification also using a mass tolerance of ±5 mDa or 5 ppm.

2.6. Software

For the acquisition and handling of ATR-FTIR spectra, the OPUS 6.5 software from Bruker was employed. LC-TOFMS data acquisition and instrument control employed software Analyst 4.1 (ABSciex). LC-TOFMS data was processed using open source ProteoWizard 3.0, mzMine 2.10, and in house written MATLAB 7.7.0 (Mathworks Inc.) functions. Multivariate analysis was performed using PLS Toolbox 7.0 from Eigenvector Research Inc. (Wenatchee, WA, USA) and the Statistics Toolbox (Mathworks). ATR-FTIR and LC-TOFMS raw data as well as the peak table used in this work are available from the authors.

3. Results and discussion

3.1. LC-TOFMS sample analysis

A total of 5293 features were detected in the 70–700 *m/z* range after peak detection, chromatographic de-convolution and peak alignment of the entire LC-ESI(+)-TOFMS chromatographic batch using the conditions described in the section LC-TOFMS data processing. Nonetheless, this number was inflated by source contaminants and other sample components originating from e.g. tubes, solvent impurities, etc. Thus, blank samples were used to identify a total of 2456 non-relevant features that were removed from the data set. Then, to further reduce the dimensionality of the data set, unreliable variables detected in <50% of the plasma samples of each class (i.e., CH_3CN, CH_3OH or acetone treated samples) were removed. The analysis yielded a similar number of detected features in plasma samples precipitated using CH_3CN (245), CH_3OH (242) and acetone (308), retaining a total of 415 reliable features. The feature distributions shown in Fig. 1, indicated that the main differences were found in the number and RSD values of late-eluting *m/z* ions (RT > 8 min). However, multivariate analysis was required to analyze trends, patterns and differences among samples.

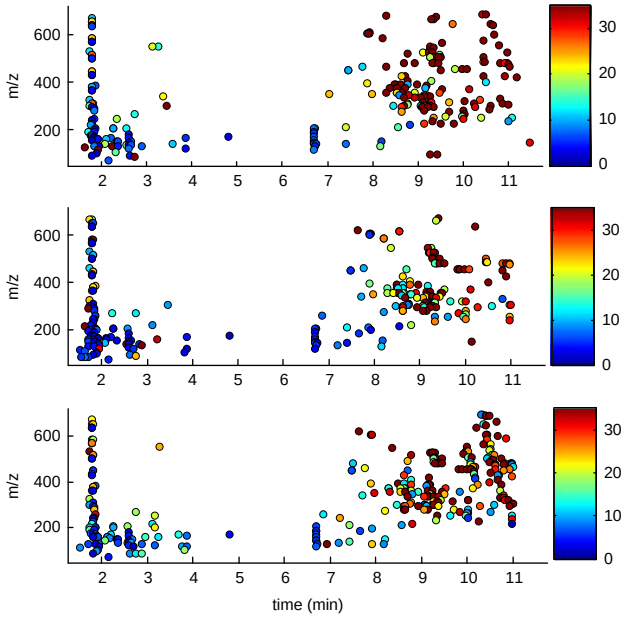

Fig. 1. Distribution of the detected features and precision levels, measured as % RSD (see colorbar scale), for the set of plasma samples after protein precipitation using CH_3CN (top), CH_3OH (middle) and acetone (bottom).

The mean intensities of the internal standards cystine-D_4 (RT: 1.84 min), methionine-D_3 (RT: 2.34 min), cysteine-D_2-NEM (RT: 3.38 min) and phenylalanine-D_5 (RT: 4.81 min) were calculated to measure the effect of ionic suppression during ESI. As shown in Fig. S-2, the intensity of phenylalanine-D_5 and methionine-D_3 was comparable among spiked blanks and plasma samples. On the contrary, the cysteine-D_2-NEM intensity was slightly higher (8%) in CH_3CN treated samples and blanks. The biggest difference was observed for cystine-D_4 whose intensity in plasma samples was intensively suppressed (96.4–97.9%) as compared to that found in blanks. Besides, cystine-D_4 peak intensities indicated higher ionic suppression at the beginning of the chromatogram for CH_3OH and acetone treated plasma samples.

3.2. ATR-FTIR sample analysis

Fig. 2 shows the average ATR-FTIR spectra of plasma samples after protein precipitation using CH_3CN, CH_3OH or acetone. Changes in intensity and shape of the ATR-FTIR bands indicated that the solvent used for protein precipitation had a remarkable effect on the sample composition. Plasma samples showed intense absorption bands in the 3100–2800 cm^{-1} and 1800–900 cm^{-1} regions. These intervals cover a high number of IR active modes of characteristic plasma metabolites including carbohydrates, proteins, creatinine, urea, triglycerides, amino acids or cholesterol [26]. Spectral bands in the 3100–2800 cm^{-1} range were assigned to ν(CH), $\nu_{as}(CH_3)$, $\nu_{as}(CH_2)$, $\nu_s(CH_3)$, and $\nu_s(CH_2)$ stretching vibrations characteristic of unsaturated fatty acids, cholesterol esters, triglycerides, long chain fatty acids, phospholipids and glycerol [25]. Likewise, bands in 1480–1430 cm^{-1} interval were assigned to $\delta_{as}(CH_3)$, $\delta_{as}(CH_2)$, $\delta_s(CH_3)$, and $\delta_s(CH_2)$ bending (scissoring) vibrations of fatty acids, phospholipids and triglycerides [25]. The region between 1740 and 1732 cm^{-1} included maxima of absorption bands of stretching vibrations of carbonyl groups of lipids, cholesterol esters and triglycerides [25]. Besides, 1720–1600 cm^{-1}, 1600–1480 cm^{-1} and 1400–1200 cm^{-1} regions were assigned to amide I, II and III bands, respectively, mainly associated with the protein backbone [30]. Protein characteristic bands were indicative of residual contents after protein precipitation. Bands in the regions between 1630 and 1560 cm^{-1} and between 1460 and 1360 cm^{-1} are assigned to C=O stretching and N–H bending vibrations of proteins and amino acids, respectively. Finally the

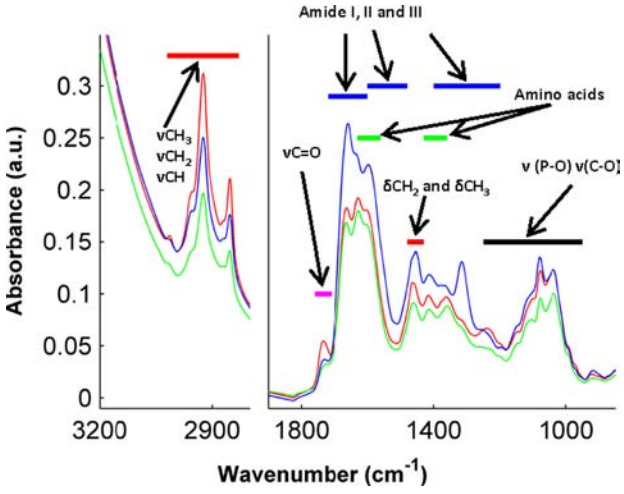

Fig. 2. Mean spectra of plasma samples after protein precipitation using acetone (red), CH_3CN (green) and CH_3OH (blue). Horizontal lines indicate the spectral regions associated with CH_2 and CH_3 (red), carbonyl (magenta), amides (blue) and amino acids (green) groups as well as to P–O and C–O bonds (black). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

region between 1250 and 950 cm^{-1} was assigned to active vibrations of C–O and P–O groups (e.g., saccharides and phospholipids) [31]. Heparin is an IR absorbing sulfated glycosaminoglycan that is present in tubes used for plasma collection and could be detected in 'heparin blanks' (see Fig. S-3). However, since spectra of plasma samples were dominated by the absorption of other sample components, the spectral differences observed among different types of plasma samples were not limited to the heparin absorbing regions.

In summary, Fig. 2 shows that although IR spectroscopy cannot provide an unambiguous identification of plasma metabolites due to high spectral overlapping, it yields a snapshot of the underlying global plasma composition.

3.2.1. *Repeatability of ATR-FTIR analysis of plasma samples*

Repeatability of ATR-FTIR measurements is a critical factor as it reduces the within-class data variation facilitating the identification of differences between protein precipitation methods. Automation of the ATR-FTIR sample handling improves manual error, repeatability, reproducibility and sample throughput [19]. However, this study was designed for laboratories with basic FTIR instrumentation and, because of that, it involved the manual deposition of plasma samples onto the ATR surface. Precision of ATR-FTIR measurements was affected by different sources of variation including the manual pipetting precision, changes in the positioning accuracy, shape and homogeneity of the dried droplet, drying time and the instrumental noise. Therefore, the intensity of IR spectra depends on the contact area and also on the skill of the technician to consistently form films of evenly distributed materials on a similar area upon drying the sample. Spectra normalization to e.g., vector unit, a reference band or using internal deuterated standards might reduce this unwanted data variation thus improving the precision levels. However, in this study spectra were not normalized to facilitate the evaluation of the variation due to manual sample measurement. Fig. 3 shows boxplots of areas of the ATR-FTIR spectra of dried spots of plasma samples measured in two regions: 2827–2985 cm^{-1} and 1522–1710 cm^{-1}, corresponding to C–H and amide I and II absorption, respectively. Intra-sample RSD% values were calculated from a set of spectra measured from three spots acquired by triplicate after protein precipitation. This value provided information on the effect of manual pipetting, positioning accuracy, shape and homogeneity of the dried droplet, drying time and the instrumental noise on the precision of the measurement. Inter-sample RSD% values were calculated using a set of 27 spectra acquired for each plasma sample. Acceptable intra- and inter-sample RSD% values in 2–16% (median = 6%) and 6–12% (median = 10%) range, respectively were obtained. Precision levels were compared among three solvents used for protein precipitation in both spectral regions (see Table S-1). From the obtained area values in the CH and amide I – II regions, ANOVA revealed significant differences among area values when employing acetone, CH_3OH and CH_3CN, with p-Values < 0.001.

3.4. *Principal component analysis (PCA) of LC-TOFMS and ATR-FTIR data*

PCA is among the most widely used methods for initial exploratory data analysis. It provides an unbiased overview of the data structure and is a common strategy for identifying outliers, trends or clusters present in multivariate data sets. PCA scores plot provides an overview of the latent patterns in the data sets as the distances among samples can be related to their similarity with respect to what pattern the model describes [32]. Fig. 4A shows the scores plot obtained from the PCA of the LC-TOFMS data set, calculated using autoscaling as data pretreatment and two PCs explaining 41.2% of the total variance. This plot reflected a clear separation among the three type of samples, supporting that the main source of variation was differenc in the metabolomic profiles due to the solvent used for protein precipitation. The analysis of the loadings shown in Fig. 4B revealed that variables responsible for the separation of plasma samples along PC1 and PC2 had characteristic distributions of m/z and RT values. The score of the first PC was lower for CH_3CN-precipitated samples than for the samples precipitated with other solvents, and the loadings plot indicated negative values on this PC for polar compounds of low molecular weight (e.g. amino acids and

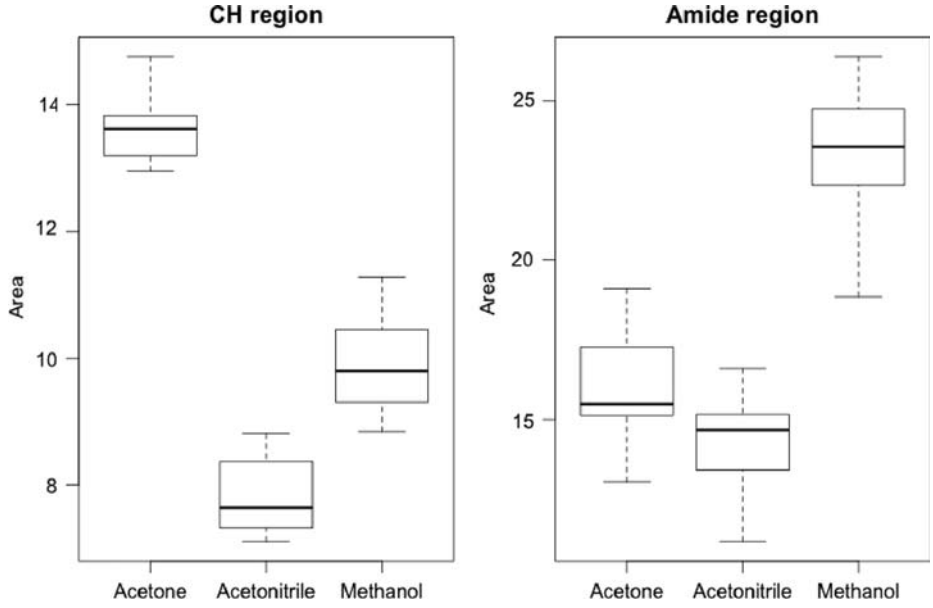

Fig. 3. Boxplots of ATR-FTIR area values in the CH and amide I and II regions measured in the dry spots of plasma samples after protein precipitation using acetone, CH_3CN and CH_3OH. Note: integration of the CH region between 2827 and 2895 cm^{-1} was performed using a baseline fitted between the 2748–2812 cm^{-1} and the 3039–3132 cm^{-1} regions (integration mode F in OPUS 6.5). For the integration of the amide I and amide II regions between 1522 and 1710 cm^{-1}, a two-point baseline at 1522 and 1710 cm^{-1} was used (integration mode B in OPUS 6.5).

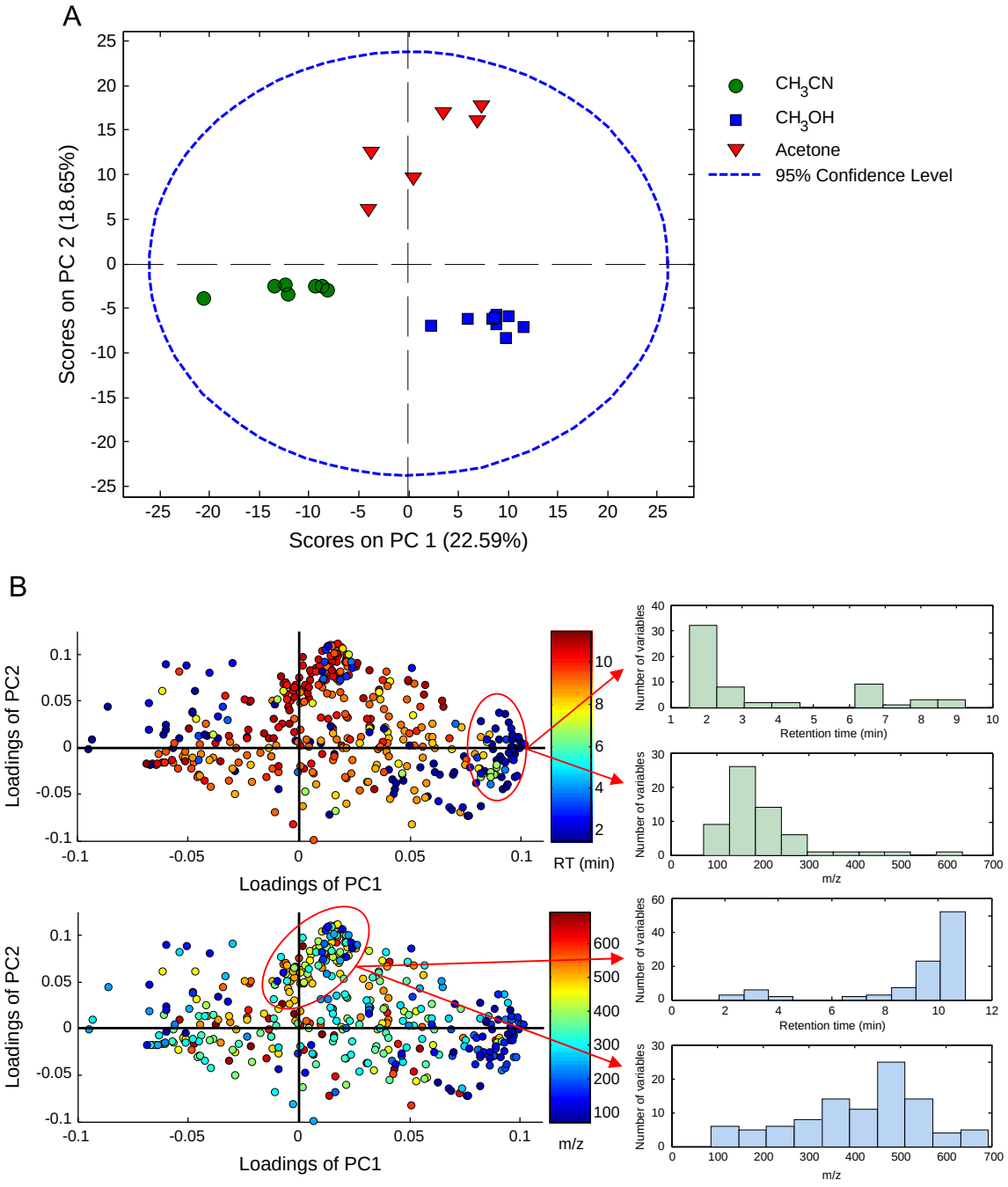

Fig. 4. PCA scores (A) and loadings (B) of PC1 and PC2 of LC-TOFMS data. Histograms of the RT ($_{min}$) and *m*/*z* values of the variables encircled in the figure are highlighted as responsible for the separation of plasma samples along PC1 and PC2. Note: color bars reflecting either the retention time (RT) (top) or the *m*/*z* (bottom) of each variable were included to facilitate interpretation of the plot.

derivatives) evidencing a more effective extraction of those compounds in CH_3CN. Besides, the positive values of the scores for acetone-treated samples are related in the loadings plot to positive weights for lipophilic compounds of high molecular weight (e.g. lipids, phospholipids) thus evidencing a higher content of lipids in acetone extracts.

Finally, to evaluate the effect of ionic suppression in the PCA scores plot, variables in the 0–2.0 min range were normalized using cystine-D_4 as internal standard and a second PCA model was calculated. Results obtained were comparable to those obtained before normalization (see Fig. S-4).

Differences in the FTIR spectra of plasmas between 850 and 1825 cm^{-1} and 2679 and 3200 cm^{-1} were evaluated by PCA, including FTIR spectra of 'heparin blanks' and plasma samples and using 2 PCs that described together 91.75% of the total variance. As expected, pattern recognition from the scores plot clearly differentiated plasma and blank samples. Besides, a major clustering related to the solvent used for protein precipitation was observed (see Fig. 5A). However, no clustering was observed among blank samples, thus supporting that the main source of variance among the spectra was the change in the sample composition due to protein precipitation and not the presence of heparin and/or solvent residues in the sample spectra.

Fig. 5B1 shows scores plot of a 2 PC model describing 90.06% of the total variance of the plasma sample set. The within-class clustering was indicative of the relative consistency among sample

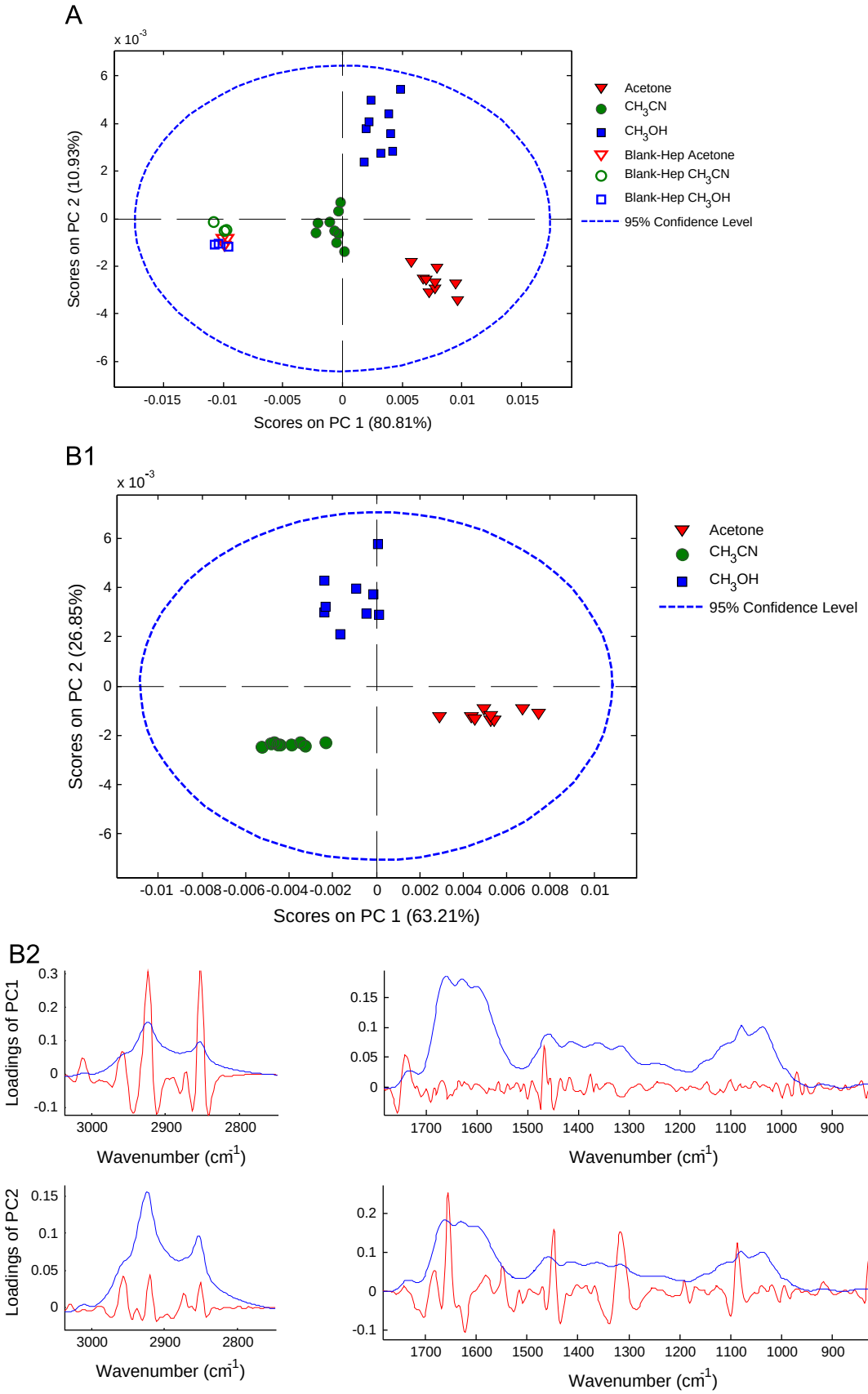

Fig. 5. PCA of the ATR-FTIR spectra of (A) plasma blanks with heparin and plasma samples after protein precipitation, and (B) plasma samples after protein precipitation. B1): Scores plot of PC1 vs. PC2; B2) Mean FTIR spectra of plasma samples. Color scale corresponds to the value of the loading of PC1 (top) and PC2 (bottom) for each variable.

replicates. The identification of spectral regions responsible for the difference among acetone, CH_3OH and CH_3CN-treated samples was carried out by considering simultaneously both, scores and loadings plots shown in Fig. 5B.2. The main conclusions obtained from the LCMS analysis are in good agreement with the results obtained from IR spectroscopy. PC1, explaining 63.21% of the data variance, discriminates plasma samples precipitated using acetone from CH_3CN and CH_3OH (see Fig. 5B.1). Thus, higher values of the scores of PC1 were associated with the use of acetone for protein precipitation. The loadings plot in Fig. 5B shows that the regions with a higher contribution to PC1 were centered on 2925, 2855, 1735 and 1460 cm^{-1}. Both, 2925 and 2855 cm^{-1} bands are located in the spectral interval that covers ν(CH), $\nu_{as}(CH_3)$, $\nu_{as}(CH_2)$, $\nu_s(CH_3)$, and $\nu_s(CH_2)$ stretching vibrations associated with methylene groups from long aliphatic chains and therefore indicate an effective extraction of lipids. In addition, absorption at 1735 cm^{-1} was attributed to ν(C=0) stretching vibrations also present in lipid spectra.

The higher scores values of CH_3OH-treated samples on the second principal component allowed their differentiation from CH_3CN-treated samples. Examination of the loadings plot suggests that discrimination was based on the absorption bands centered at 2955, 2920, 2852, 1655, 1448, 1315 and 1085 cm^{-1}. Absorption in 3200–2800 cm^{-1} range is mainly due to ν(CH), $\nu_{as}(CH_3)$, $\nu_{as}(CH_2)$, $\nu_s(CH_3)$, and $\nu_s(CH_2)$ stretching vibrations indicating a high relative concentration of lipids in CH_3OH treated samples. Proteins contribute to the Amide I band (C=O stretch) centered at 1650 cm^{-1}. Absorption bands in 1660–1300 cm^{-1} region were not assigned to specific functional groups due to strong spectral overlapping including $\delta(NH_2)$ bending and $\nu(COO^-)$ stretching of amino acids and amide II and III bands of proteins. The maximum found at 1448 cm^{-1}, however, was assigned to $\delta_{as}(CH_3)$, $\delta_{as}(CH_2)$, $\delta_s(CH_3)$, and $\delta_s(CH_2)$ bending (scissoring) vibrations of lipids. Absorption centered around 1085 cm^{-1} was assigned to ν(C–O) and ν(P–O) stretching vibrations of e.g., carbohydrates and phospholipids. The above mentioned bands do not correspond to heparin absorption bands, again supporting that spectral differences were not due to different extraction of this sulfated glycosaminoglycan. The residual Q and the Hotelling's T^2 statistics were used for outlier detection [33]. For the two previously described PCA models, calculated Q and Hotelling's T^2 statistics confirmed that the majority of samples fall within 95% confidence levels (see Fig. S-5).

3.5. Comparison of LC-TOFMS and ATR-FTIR patterns

In order to avoid flawed conclusions, trends in the biochemical composition observed by FTIR should agree with those observed by LC-TOFMS. In this work, the similarity between the ATR-FTIR and LC-TOFMS data was assessed by two multivariate analysis methods, namely the Mantel test [34] and procrustes analysis [35]. The Mantel test evaluates the significance of the correlation between two pairwise distance matrices. In this study, the standardized Euclidean multivariate distances between matched sample pairs analyzed by ATR-FTIR and LC-TOFMS were used. The significance of the multivariate correlation was determined by a Monte Carlo simulation in which a null distribution was obtained by randomly shuffling ATR-FTIR spectra and, for each permutation a correlation coefficient was calculated. Results after 10,000 random permutations formed an empirical null distribution that was used to estimate the p-Value as the proportion of permuted absolute correlation values that were equal to or greater than the actual value [36]. Results depicted in Fig. 6 shows a statistically significant correlation between ATR-FTIR and LC-TOFMS data. This finding was consistent with previous studies that showed the capability of FTIR to fingerprint complex biological samples. To

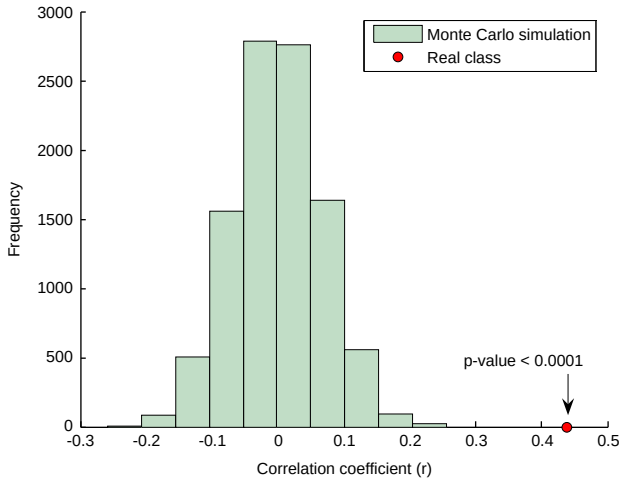

Fig. 6. Results of Mantel test on raw LC-TOFMS and ATR-FTIR data using the standardized Euclidean distance and 10,000 Monte Carlo permutations.

further confirm this, we performed a procrustes analysis on the principal components that described the LC-TOFMS and ATR-FTIR data. Procrustes analysis is a method for analyzing the relatedness between two data sets X and Y (here, the LC-TOFMS and ATR-FTIR data sets, respectively). It determines a linear transformation by translation, reflection, orthogonal rotation, and scaling of the points of the Y matrix so it resembles the points in matrix X. The employed goodness-of-fit criterion (d) was the sum of squared errors between the two transformed matrices. The smaller the d value, the more similar the two transformed matrices are (i.e. a perfect match would provide $d=0$) [37], and its significance was assessed empirically by using a Monte Carlo approach. As in the previous analysis, ATR-FTIR samples were randomly permuted and for each permutation the sum of squared errors between the transformed matrices after procrustes transformation was calculated. The match of the data sets and the associated residuals provide more information than the Mantel test. Results shown in Fig. 7 revealed again a strong similarity (empirical p-Value ≤ 0.0001) among MS and FTIR data sets with a clear clustering of the samples according to the solvent used for protein precipitation.

Overall, these results supported the hypothesis that major changes in the LC-TOFMS metabolic profiles can be anticipated by ATR-FTIR spectroscopy. Therefore, it could be anticipated that information extracted from FTIR could be useful to compare and optimize clean-up procedures in order to select the conditions providing, e.g. the highest protein, phospholipids or lipids removal from the sample matrix or to evaluate sample collection, handling and storage protocols. We therefore sought to gain further insight into the observed FTIR clustering among protein precipitation methods by comparing results from LC-TOFMS. Fig. 8 shows mean LC-TOF MS intensities of a set of metabolites classified as: amino acids and derivatives, fatty acid esters, monosaccharides, glycerophospholipids, imidazopyrimidines and indoles (see also Table S-2). Intensities of most of the identified amino acids, monosaccharides, imidazopyrimidines and indoles decreased in CH_3OH > acetone > CH_3CN series and mean intensities of glycerophospholipids and fatty acids were, as expected, higher in plasma samples precipitated using acetone. This is in accordance with the conclusions withdrawn from the PCA of ATR-FTIR spectra. Furthermore, the use of CH_3CN lead to a general decrease of band intensities in ATR-FTIR spectra and the number of features were detected in MS, which were also reflected in lower ion suppression during cystine-D_4 elution.

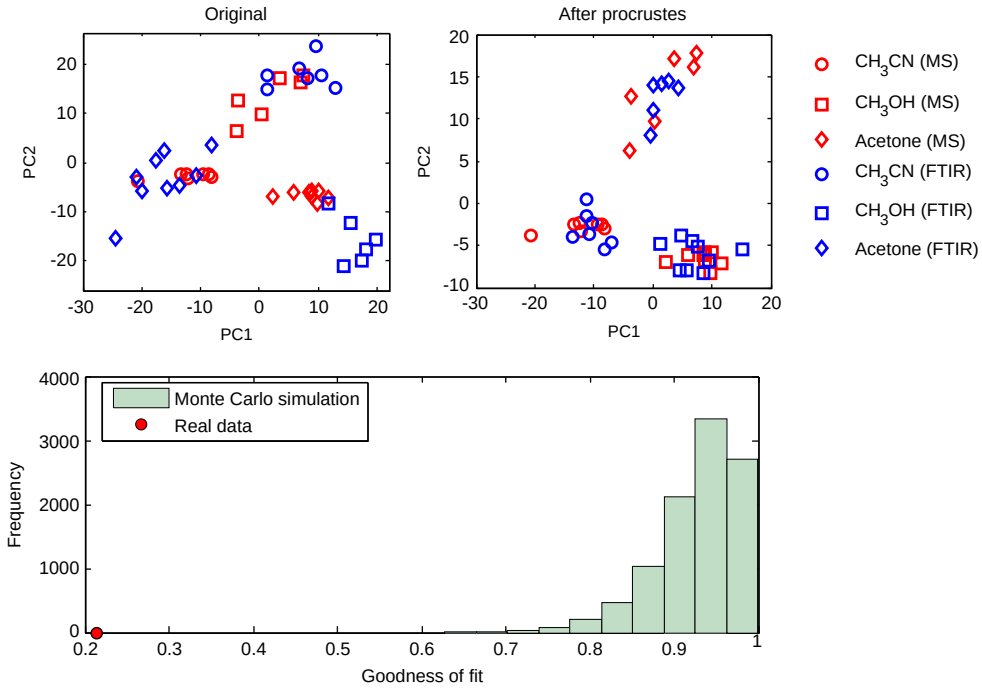

Fig. 7. Results from the procrustes analysis of LC-TOFMS and ATR-FTIR data. Top: Score plots before (left) and after (right) procruster analysis. Bottom: Results from the Monte Carlo assessment of the statistical significance of the sum of squared errors between the two transformed matrices.

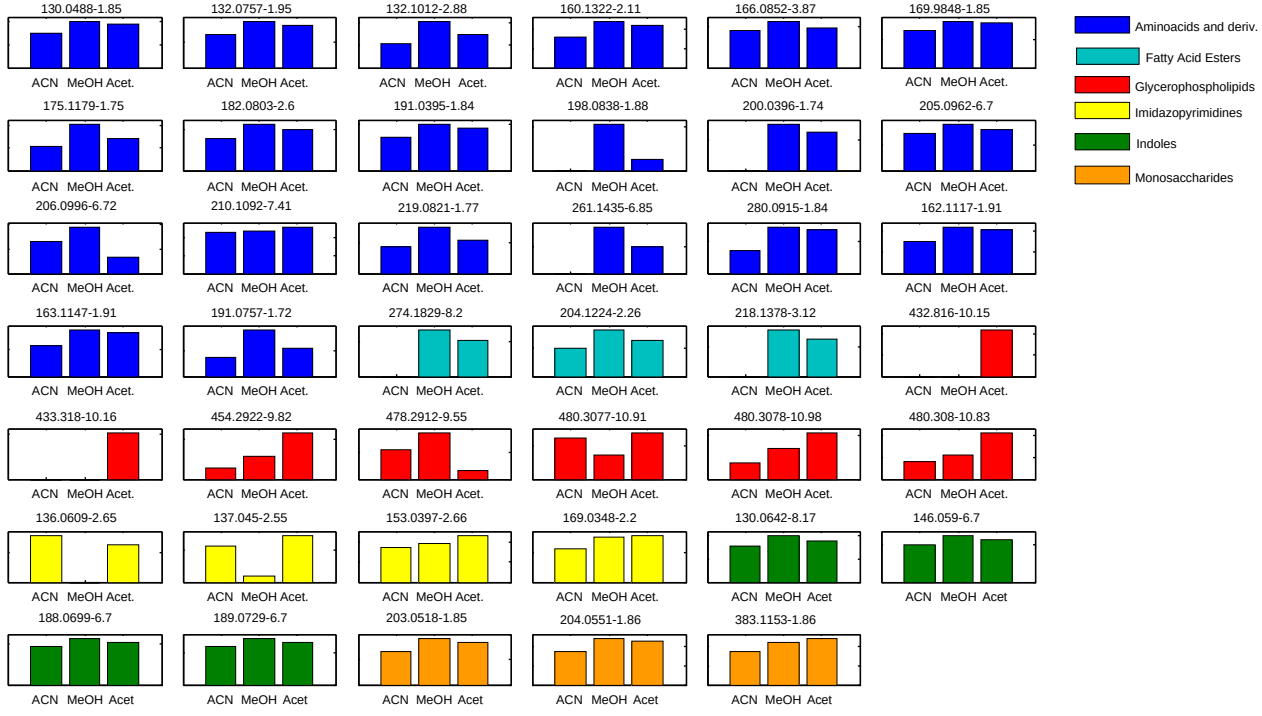

Fig. 8. Mean intensities of selected metabolites identified according to the *m*/*z* using a tolerance of 5 mDa or 5 ppm. From left to right and from top to bottom: Pyroglutamic acid; Creatine; Isoleucine; 2-Aminooctanoic acid; Phenylalanine; Iminoaspartic acid; Arginine; Tyrosine; Glutamine; Argininic acid; Histidine; Tryptophan; Tryptophan (isotope); N-Heptanoylglycine; Arginine; gamma-glutamyl-isoleucine; Pyro-L-glutaminyl-L-glutamine; Carnitine; Carnitine (isotope); Lysine; 3-Mercaptohexyl hexanoate; Acetylcarnitine; Propionylcarnitine; Phosphatidylcholine(18:3/18:1); Phosphatidylcholine(18:3/18:1) (isotope); LysoPC; LysoPE; LysoPE; LysoPE; LysoPE; Adenine; Hypoxanthine; Xanthine; Uric acid; 3-Methylene-indolenine; 1-H-Indole-3-carboxaldehyde; Indoleacrylic acid; Indoleacrylic acid (isotope); Glucose; Glucose; Glucose.

To be worthwhile, the evaluation of sample pretreatments by FTIR biospectroscopy should not only provide useful information, but it should also reduce the time required for data analysis, facilitate the interpretation of the results and it also has to be cost effective. Data analysis of large multidimensional LC-TOFMS metabolomic data sets is very time consuming. Raw data must be thoroughly processed and the efficiency, accuracy and peak detection performance of the processing depends on a set of user-selected parameters and algorithms including, e.g. MS and LC-peak detection, smoothing, de-convolution, integration and alignment

[10]. For example, previous results showed that only a partial overlap was obtained when comparing the metabolomic peak tables obtained using *centWave* [38], a matched filter implemented in XCMS [39] and mzMine [40]. Moreover, the optimum analytical conditions for LC-TOFMS metabolic profiling depend on the sample composition which in turn is affected by the sample pretreatment due to, e.g. ion suppression, adduct formation, ESI contamination or poor column performance that might bias the results. Results obtained using ATR-FTIR showed that no complex data pretreatment was needed and so, major trends in the plasma compositions following protein precipitation could be identified faster than by LC-TOFMS. In this study, protein precipitation, sample measurement and the comparison of the main differences among protein precipitation methods using ATR-FTIR could be carried out in less than 3 h at virtually no cost.

4. Conclusions

The use of LC-TOFMS data for the comparison among sample pretreatments is expensive, very time consuming and requires extensive data pre-processing and analysis. In this work we have shown the feasibility of ATR-FTIR biospectroscopy for the evaluation of major trends in the plasma composition after sample pretreatment, providing information in terms of e.g., amino acids, proteins, lipids and carbohydrates overall contents comparable to those found by LC-TOFMS. The partial trade-off for time saved during data acquisition and analysis is reduced specificity and sensitivity. The biochemical information extracted from ATR-FTIR data could be useful for the comparison and optimization of sample preparation in metabolomics studies and it might be used to tailor the clean-up steps to the objective of the study.

Acknowledgments

The authors thank Mari-Luz Moreno for providing mouse plasma samples. DPG acknowledges the “V Segles” grant provided by the University of Valencia and JE and JK the “Sara Borrell” grants (CD11/00154 and CD12/00667) from the *Instituto Carlos III* (Ministry of Economy and Competitiveness). MV acknowledges the FISPI11/0313 grant from the *Instituto Carlos III* (Ministry of Economy and Competitiveness). GQ acknowledges financial support from the Spanish Ministry of Economy and Competitivity (SAF2012-39948). Furthermore the technical support from the *Servicio Central de Soporte a la Investigación Experimental* (SCSIE) at the University of Valencia (Spain) for carrying out LC-TOFMS measurements is kindly acknowledged.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.talanta.2014.04.009.

References

[1] M. Sugimoto, K. Kawakami, M. Robert, T. Soga, M. Tomita, Curr. Bioinform. 7 (2012) 96–108.
[2] Z. Pan, D. Raftery, Anal. Bioanal. Chem. 387 (2007) 525–527.
[3] A. Smolinska, L. Blanchet, L.M. Buydens, S.S. Wijmenga, Anal. Chim. Acta 750 (2012) 82–97.
[4] T.Q. Metz, Q. Zhang, J.S. Page, Y. Shen, S.J. Callister, J.M. Jacob, R.D. Smith, Biomark. Med. 1 (2007) 159–185.
[5] D. Vuckovic, Anal. Bioanal. Chem. 403 (2012) 1523–1548.
[6] J.K. Nicholson, E. Holmes, J.C. Lindon, The Handbook of Metabonomics and Metabolomics, Elsevier, Oxford (UK), 2007.
[7] C. Polson, P. Sarkar, B. Incledon, V. Raguvaran, R. Grant, J. Chromatogr. B 785 (2003) 263–275.
[8] E.J. Want, G. O'Maille, C.A. Smith, T.R. Brandon, W. Uritboonthai, C. Qin, S. A. Trauger, G. Siuzdak, Anal. Chem. 78 (2006) 743–752.
[9] S. Naz, A. García, C. Barbas, Anal. Chem. 85 (2013) 10941–10948.
[10] Anon., in: M. Lämmerhofer, W. Weckwerth (Eds.), Metabolomics in Practice : Successful Strategies to Generate and Analyze Metabolic Data, Wiley-VCH & Co, Weinheim, Germany, 2013.
[11] W.B. Dun, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, M. Brown, J.D. Knowles, A. Halsall, J.N. Haselden, A.W. Nicholls, I.D. Wilson, D. B. Kell, R. Goodacre, Nat. Protoc. 6 (2011) 1060–1083.
[12] S.J. Bruce, I. Tavazzi, V. Parisod, S. Rezzi, S. Kochlar, P.A. Guy, Anal. Chem. 81 (2009) 3285–3296.
[13] S.J. Bruce, P. Jonsson, H. Antti, O. Cloarec, J. Trygg, SL. Marklund, T. Moritz, Anal. Biochem. 372 (2008) 237–249.
[14] F. Michopoulos, L. Lai, H. Gika, G. Theodoridis, I. Wilson, J. Proteome Res. 8 (2009) 2114–2121.
[15] M. Brown, D.C. Wedge, R. Goodacre, D.B. Kell, P.N. Baker, L.C. Kenny, M.A. Mamas, L. Neyses, W.B. Dunn, Bioinformatics 27 (2011) 1108–1112.
[16] JG. Kelly, J. Trevisan, A.D. Scott, P.L. Carmichael, H.M. Pollock, P.L. Martin-Hirsch, F.L. Martin, J. Proteome Res. 10 (2011) 1437–1448.
[17] D.I. Ellis, R. Goodacre, Analyst 131 (2006) 875–885.
[18] F. Severcan, P.I. Haris, Vibrational Spectroscopy in Diagnosis and Screening, IOS Press BV, Netherlands, 2012.
[19] J. Ollesch, S.L. Drees, H.M. Heise, T. Behrens, T. Brüning, K. Gerwert, Analyst 138 (2013) 4092–4102.
[20] J. Trevisan, P.P. Angelov, P.L. Carmichael, A.D. Scott, F.L. Martin, Analyst 137 (2012) 3202–3215.
[21] G. Bellisola, C. Sorio, Am. J. Cancer Res. 2 (2012) 1–21.
[22] M. Kolhed, B. Lendl, B. Karlberg, Analyst 128 (2003) 2–6.
[23] C. Kendall, M. Isabelle, F. Bazant-Hegemark, J. Hutchings, L. Orr, J. Babrah, R. Baker, N. Stone, Analyst 134 (2009) 1029–1045.
[24] R.A. Shaw, S. Kotowich, M. Leroux, H.H. Mantsch, Ann. Clin. Biochem. 35 (1998) 624–632.
[25] G. Deleris, C. Petitbois, Vib. Spectrosc. 32 (2003) 129–136.
[26] C. Petibois, G. Cazorla, A. Cassaigne, G. Deleris, Clin. Chem. 47 (2001) 730–738.
[27] M. Brandstetter, T. Sumalowitsch, A. Genner, A. Posch, C. Herwig, A. Drolz, V. Fuhrmann, T. Perkmann, B. Lendl, Analyst 138 (2013) 4022–4028.
[28] D. Pérez-Guaita, J. Kuligowski, G. Quintás, S. Garrigues, M. de la Guardia, Talanta 107 (2013) 368–375.
[29] D. Pérez-Guaita, A. Sanchez-Illana, J. Ventura-Gayete, S. Garrigues, M de la Guardia, Analyst 139 (2013) (2013) 170–178.
[30] A. Barth, BBA-Bioenergetics 9 (2007) 1073–1101.
[31] F.M. Goñi, J.L. Arrondo, Faraday Discuss. Chem. Soc. 81 (1986) 117.
[32] F. Savorani, M.A. Rasmussen, M.S. Mikkelsen, S.B. Engelsen, 2012 Food Res. Int. http://dx.doi.org/10.1016/j.foodres.2012.12.025.
[33] B.M. Wise, J.M. Shaver, N.B. Gallagher, W. Windig, R. Bro, R.S. Koch, Manual PLS_Toolbox, 4th ed., Eigenvector Research Inc., Wenatchee (USA), 2006.
[34] N. Mantel, Cancer Res. 27 (1967) 209–220.
[35] J.C. Gower, Psychometrika 40 (1975) 33–51.
[36] R.G. Brereton, Chemometrics for Pattern Recognition, John Wiley & Sons, Chichester, 2009.
[37] J.M. Andrade, M.P. Gómez-Carracedo, W. Krzanowski, M. Kubista, Chemom. Int. Lab. Syst. 72 (2004) 123–132.
[38] R. Tautenhahn, C. Bottcher, S Neumann, BMC Bioinformatics 9 (2008) 504.
[39] CA. Smith, EJ Want, G O'Maille, R Abagyan, G Siuzdak, Anal. Chem. 78 (2006) 779–787.
[40] M. Katajamaa, M. Orešič, J. Chromatogr. A 1158 (2007) 318–328.