Universidad de Córdoba
FACULTAD DE CIENCIAS
Departamento de Química Analítica

Analytical strategies to improve qualitative and quantitative determination in metabolomics by mass spectrometry

Estrategias analíticas para la mejora de la determinación cualitativa y cuantitativa en metabolómica por espectrometría de masas

María Asunción López Bascón

Tesis Doctoral
Córdoba, 2019

TITULO: *Analytical strategies to improve qualitative and quantitative determination in metabolomics by mass spectrometry*

AUTOR: *María Asunción López Bascón*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

UNIVERSIDAD DE CÓRDOBA

FACULTAD DE CIENCIAS
DEPARTAMENTO DE QUÍMICA ANALÍTICA

Estrategias analíticas para la mejora de la determinación cualitativa y cuantitativa en metabolómica por espectrometría de masas

Analytical strategies to improve qualitative and quantitative determination in metabolomics by mass spectrometry

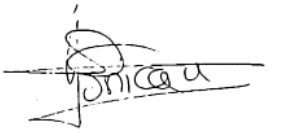
María Asunción López Bascón

Córdoba, 2019

Estrategias analíticas para la mejora de la determinación cualitativa y cuantitativa en metabolómica por espectrometría de masas

Los Directores,

Fdo. Feliciano Priego Capote
Profesor Titular del Departamento
de Química Analítica de la
Universidad de Córdoba



Fdo. Mónica Calderón Santiago
Profesora Ayudante Doctora del
Departamento de Didáctica de las
Ciencias Sociales y Experimentales de la
Universidad de Córdoba

Trabajo presentado para optar al grado de
Doctora en Ciencias, Sección Químicas

Fdo. María Asunción López Bascón

Feliciano Priego Capote, Profesor Titular del Departamento de Química Analítica de la Facultad de Ciencias, y **Mónica Calderón Santiago**, Profesora Ayudante Doctora del Departamento de Didáctica de las Ciencias Sociales y Experimentales de la Facultad de Ciencias de la Educación, ambos de la Universidad de Córdoba, en calidad de Directores de la Tesis Doctoral presentada por María Asunción López Bascón, con el título "Estrategias analíticas para la mejora de la determinación cualitativa y cuantitativa en metabolómica por espectrometría de masas",

CERTIFICAN:

Que la citada Tesis Doctoral se ha realizado en los laboratorios del Departamento de Química Analítica, Facultad de Ciencias, Universidad de Córdoba y que, a su juicio, reúne los requisitos necesarios exigidos en este tipo de trabajo.

Y para que conste y surta los efectos pertinentes, expiden el presente certificado en Córdoba, a 21 de octubre de 2019.

Fdo. Feliciano Priego Capote

Fdo. Mónica Calderón Santiago

Mediante la defensa de esta Memoria se pretende optar a la mención de **Doctorado Internacional**, habida cuenta de que la doctoranda reúne los requisitos exigidos para tal mención, a saber:

1. Informes favorables de dos doctores pertenecientes a Instituciones de Enseñanza Superior de otros países:
    - Samira Salihovic, School of Medical Sciences, University of Örebro, Örebro, Sweden.
    - Stephan Hann, Division of Analytical Chemistry, University of Natural Resources and Life Sciences Vienna, Austria.
2. Uno de los miembros del tribunal que ha de evaluar la Tesis pertenece a un centro de Enseñanza Superior de otro país:
    - Jean-Michel Kauffmann, Faculty of Pharmacy, Free University of Brussels, Belgium.
3. La exposición y la defensa de parte de esta Tesis se realizarán en una lengua diferente a la materna: inglés.
4. Estancia de tres meses en un centro de investigación de otro país:

School of Science and Technology of the University of Örebro (Sweden), bajo la supervisión de los profesores Tuulia Hyötyläinen y Matej Öresic.

TÍTULO DE LA TESIS: Estrategias analíticas para la mejora de la determinación cualitativa y cuantitativa en metabolómica por espectrometría de masas

DOCTORANDO: María Asunción López Bascón

INFORME RAZONADO DE LOS DIRECTORES DE LA TESIS

La doctoranda María Asunción López Bascón, a lo largo de toda su formación, ha mostrado ser muy trabajadora y constante. Tanto es así que consiguió terminar la Licenciatura de Ciencias Ambientales en la Universidad de Córdoba en cuatro años (2009-2013), en lugar de los cinco cursos en los que se dividía el plan de estudios en ese momento. A continuación, obtuvo la Beca de Colaboración del Ministerio y durante su disfrute cursó el Máster Interuniversitario de Química en la misma universidad. Desde su incorporación al grupo de investigación, en 2013, ha mostrado gran interés por la metabolómica, tanto por la instrumentación necesaria para obtener la información requerida en cada caso en estudio, como por el proceso de tratamiento de datos, sin importar la complejidad de éstos. Tras finalizar el Máster, obtuvo una Beca-Contrato de Formación de Profesorado Universitario de la que disfruta desde 2016 (convocatoria 2015).

Durante el tiempo de disfrute de la Beca-Contrato, la doctoranda ha realizado investigación principalmente en el área de la metabolómica clínico-nutricional y, en menor medida, en el área de la metabolómica agroalimentaria. Además, ha colaborado en la investigación realizada por otros miembros de grupo, así como con otros grupos de investigación, ha escrito un capítulo de libro y ha presentado comunicaciones a congresos. Toda esta información se desglosa a continuación:

Los avances conseguidos con el desarrollo de la investigación recogida en esta Memoria se deducen del título de la Tesis y son los siguientes:

1. El desarrollo de metodologías innovadoras para la determinación de distintos grupos de metabolitos mediante análisis metabolómico orientado, que ha dado lugar a la publicación de 2 artículos publicados

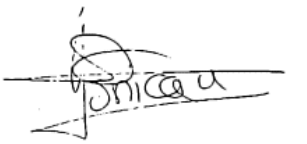
en revistas internacionales de alto índice de impacto, una del primer decil y otra del primer cuartil.

2. El desarrollo de estrategias innovadoras para contribuir a superar algunos puntos débiles del análisis en metabolómica, que ha dado lugar a 4 publicaciones en revistas internacionales de alto índice de impacto, todas ellas pertenecientes al primer cuartil.
3. La identificación de metabolitos con carácter predictivo de enfermedades (diabetes tipos I y II) o con aplicación en el área agroalimentaria (régimen alimentario en cerdos), que ha resultado en 4 artículos, de los cuales 2 ya han sido publicados en revistas internacionales de alto índice de impacto, ambas del primer decil.
4. Colaboración en la investigación realizada por otros miembros del grupo o con otros grupos de investigación, que ha dado lugar a 2 artículos publicados en revistas internacionales de alto índice de impacto.
5. La redacción de un capítulo de libro multiautor publicado por una editorial internacional.
6. La presentación oral o en cartel de 20 comunicaciones en congresos nacionales e internacionales.
7. La realización de una estancia de 4 meses en la Universidad de Örebro, de Suecia, como requisito para la mención de Doctora Internacional.

Por todo ello, consideramos que la investigación desarrollada y recogida en esta Memoria reúne los requisitos de originalidad, innovación y calidad, y autorizamos la presentación de la Tesis Doctoral de Doña Asunción López Bascón.

Córdoba, octubre 2019

Firma de los directores



Fdo.: Feliciano Priego Capote

Fdo.: Mónica Calderón Santiago

INFORME SOBRE EL FACTOR DE IMPACTO DE LAS PUBLICACIONES DE LA TESIS

TÍTULO DE LA TESIS: Estrategias analíticas para la mejora de la determinación cualitativa y cuantitativa en metabolómica por espectrometría de masas

DOCTORANDA: MARÍA ASUNCIÓN LÓPEZ BASCÓN

PUBLICACIÓN	FI*	DECIL/CUARTIL
Determination of fatty acids and stable carbon isotopic ratio in subcutaneous fat to identify the feeding regime of Iberian pigs. **J. Agric. Food Chem.** 63 (2015) 692–699	2.857	D1 3/57 Agriculture, multidisciplinary
Influence of the collection tube on metabolomic changes in serum and plasma. **Talanta** 150 (2016) 681–689	4.162	Q1 9/76 Analytical Chemistry
Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry. **Anal. Chim. Acta** 943 (2016) 82–88	4.950	D1 7/76 Analytical Chemistry
MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls. **Talanta** 174 (2017) 29–37	4.244	Q1 10/81 Analytical Chemistry
Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue. **Talanta** 177 (2018) 86–93	4.916	Q1 11/84 Analytical Chemistry

Determination of primary fatty acid amides in different biological fluids by LC–MS/MS in MRM mode with synthetic deuterated standards: Influence of biofluid matrix on sample preparation. **Talanta** 193 (2018) 29–37	4.916	Q1 11/84 Analytical Chemistry
Early Salmonella Typhimurium infection in pigs disrupts Microbiome composition and functionality principally at the ileum mucosa. **Sci. Rep.** 8 (2018) 7788	4.011	Q1 15/69 Multidisciplinary sciences
Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage. **Talanta** 199 (2019*) 303–309	4.916	Q1 11/84 Analytical Chemistry
Development of a qualitative/ quantitative strategy for comprehensive determination of polar lipids by LC–MS/MS in human plasma Accepted in **Anal. Bioanal. Chem.**	3.286	Q1 18/84 Analytical Chemistry
Metabolic Alterations in Immune Cells Associate with Progression to Type 1 Diabetes Accepted in **Diabetologia**	7.113	D1 12/145 Endocrinology & metabolism
Multimetabolite panels predict the occurrence of type 2 diabetes mellitus by examining the postprandial response Sent to **Mol. Cell. Endocrinol.**	3.693	Q2 48/145 Endocrinology & metabolism
Postprandial metabolic response to the oral fat tolerance test (OFTT) by plasma metabolomics analysis Sent to **Metabolomics**	3.167	Q2 71/145 Endocrinology & metabolism

*Factor de impacto obtenido de Journal Citation Report (JCR). Para los datos de 2019 se usaron los correspondientes de 2018.

ABSTRACT

1. Introduction

The primary objective of metabolomics is to analyze the set of low molecular weight compounds present in a biological fluid, cell, tissue or organism, in specific physiological conditions or in response to different disturbances or stimuli. The main limitations of metabolomic analysis are related to the identification of metabolites, since the existing databases are not as up-to-date as those of the other omic disciplines. On the other hand, there are no universal protocols to address the metabolomic analysis of certain samples, since there are still aspects to be standardized. Among them, it is worth noting the acquisition of information that allows the unequivocal identification of metabolites, even for mass spectrometry, the technique of greater detection power, since there is no protocol that guarantees a good identification level of compounds without an exhaustive formation of the personnel involved in this task. Another scarcely established aspect is the application of validation protocols that ensure the analytical quality of the generated results.

Almost all the limitations of metabolomics are a consequence of the great complexity of the biological samples, which include a wide variety of metabolites of diverse chemical nature and that cover a range of concentrations of several orders of magnitude. For this reason, it is technically unfeasible the development of an analytical method that in a single analysis provides results of metabolites such as, for example, glucose, a polar metabolite present at millimolar concentration, and calcitriol, a non-polar metabolite present at picomolar level. In this sense, the presence of major compounds complicates the detection of the metabolites present at low concentration. It is evident that there is a demand for more sensitive, selective and precise methodologies, which provide greater capacity for metabolite detection (measured by what is called the metabolomic coverage).

With these premises, the current major challenge of metabolomics is to maximize the detection capacity of analytical methods in order to achieve the unambiguous identification of thousands of metabolites in an organism. This challenge is currently utopian when working with complex organisms because there is no single methodology that allows reaching the level of detection of some metabolites in higher organisms.

Therefore, the main motivation of the Doctoral Thesis was to find solutions for the different challenges currently faced by metabolomics.

2. Research content

The *basic objective* of the research in this Thesis Book was to develop new analytical strategies based on the use of low- and high-resolution mass spectrometry to improve the detection and identification coverage in metabolomic analysis. These new strategies were applied throughout the main steps of the analytical process—sampling, sample preparation, determination and data analysis—and allowed improving basic analytical features such as sensitivity, selectivity and precision of metabolomic analysis methods (targeted and untargeted) and their detection capacity. The achievement of this basic objective has leaded to metabolomic analysis methods capable of providing a higher level of information, which is a key milestone for the resolution of biological problems. This objective was divided into three general objectives according to the different topics in this research: (i) to take benefit from the versatility of the triple quadrupole mass spectrometer to improve the identification/quantification of certain families of metabolites; (ii) to develop approaches to improve the detection and identification of metabolites by chromatographic techniques coupled to mass spectrometry in high resolution mode; and (iii) to create strategies for searching potential biomarkers in clinical and agro-food studies. From each general objective, we defined several concrete objectives:

- To develop an automated qualitative and quantitative method based on on-line coupling of solid phase extraction (SPE) and liquid chromatography–tandem mass spectrometry (LC–MS/MS) to maximize sensitivity for determination of fatty acid esters of hydroxy fatty acids (FAHFAs) in serum (Chapter I). The method was further applied to a cohort of individuals to evaluate the influence of glycaemia on FAHFA levels.
- To propose a qualitative/quantitative strategy for determination of polar lipids in human plasma by LC–MS/MS. Two MS/MS acquisition methods were combined to identify and confirm the presence of polar lipids in plasma (Chapter II). Thus, the process was carried out in two steps: (i) identification of lipids through the characteristic fragmentation pattern for each family; and b) confirmation of detected lipids by monitoring product ions corresponding to the fatty acids (FAs) conforming them or other characteristic product ions.
- To study the differences at metabolite level between serum and plasma obtained with conventional tubes (heparin tube for plasma) and polymeric gel tubes by application of an untargeted approach based on gas chromatography coupled to time-of-flight mass spectrometry (GC–TOF/MS) (Chapter III). A cohort of volunteers was selected

for blood sampling using four different tubes (plasma, plasma-gel, serum and serum-gel).

- To evaluate the influence of sample preparation on the determination of polar lipids in visceral adipose tissue (Chapter IV). Two different extractants were tested to compare their efficiency for the extraction of polar lipids, but also their inefficiency for extraction of acylglycerides (the main interference in the detection of polar lipids). Additionally, the implementation of an SPE step with a selective sorbent for retention of glycerophospholipids was assessed to check its influence on the subsequent detection of this family of lipids.
- To maximize the identification coverage of metabolites found in pig fecal samples through the study of sample preparation (Chapter V). For this purpose, two analytical platforms such as LC–QTOF MS/MS and GC–TOF/MS were combined to evaluate their additivity in terms of identification. Concerning sample preparation, six solvents with different polarity were tested to evaluate the extraction performance and, in case of GC–MS, two derivatization protocols were compared.
- To develop a new statistical package, called MetaboQC, to study and filtrate experimental variabililty in data sets generated by MS analysis of sequences processed for several days (Chapter VI). This new tool uses quality controls (QCs) to individually correct any tendency on quantitative signals of metabolites that can be associated to instrumental variability.
- To study, by untargeted metabolomics analysis, the postprandial response to the oral fat tolerance test (OFTT) on plasma metabolomic profile (Chapter VII). Collected plasma samples were analyzed by LC–QTOF MS/MS and GC–TOF/MS. This test can open possibilities for the application of OFTT to the diagnostic of a wide range of pathologies.
- To evaluate the predictive capacity of type 2 diabetes mellitus (T2DM) occurrence by examining the postprandial response (after OFTT) (Chapter VIII). With this aim, plasma samples were collected from 215 patients (CORDIOPREV project) at baseline and four hours after the OFTT. 107 individuals developed diabetes after five years. Collected plasma samples were analyzed by LC–QTOF MS/MS and GC–TOF/MS.
- To elucidate the early events preceding the onset of islet autoimmunity and overt type 1 diabetes mellitus (T1DM). Metabolomics was used to determine levels of molecular lipids and polar metabolites in human peripheral blood mononuclear cells (PBMCs)

isolated from prospective samples collected in the Type 1 Diabetes Prediction and Prevention (DIPP) study (Chapter IX).

- To develop discrimination models and search for panels of markers with capability to classify slaughtered pigs by their feeding regime (Chapter X). 80 samples of subcutaneous adipose tissue from Iberian pigs subjected to four different feedings were used. Data were obtained from the classical method for the determination of FAs based on GC–FID and from a method for determination of carbon isotopic abundances by isotope ratio mass spectrometry (IRMS).

3. Conclusions

The most outstanding conclusions of the Doctoral Thesis according to the initially proposed objectives are the following:

1. Benefits derived from the versatility of the triple quadrupole analyzer to improve the identification/quantification of certain families of metabolites.

   (i) An automated method based on SPE on-line coupled to LC–MS/MS has been developed for determination of FAHFAs in serum. Eleven FAHFAs have been identified and quantified in relative terms in serum by application of a confirmatory test. PAHSA and PAHOA were the most concentrated FAHFAs in serum. PAHPA and POHPO reported significant differences between glycaemic states, while only POHPA was found significantly different considering BMI.

   (ii) A dual analysis strategy for massive quantitative determination of polar lipids by LC–MS/MS with a QqQ analyzer was optimized. A combination of two MRM methods enabled to monitor 398 polar lipids in 64 minutes (32 minutes per run) after application of a simple protocol for sample preparation. The proposed method can be used in lipidomic analysis of biological samples to provide qualitative and semiquantitative information of lipid polar families.

2. Improvement of the analytical process (sampling, sample preparation and data analysis) through methodological development in untargeted metabolomics analysis.

   Sampling

(iii) It is essential to take into account that metabolic alterations occur in applying experimental protocols for metabolomics analysis in blood sampled. Serum and plasma collected in polymeric gel tubes were compared with serum and plasma obtained in conventional tubes using a GC–TOF/MS untargeted approach. Significant changes attributable to the polymeric gel were only detected in serum, while no differences were observed in plasma, which in overall terms provided a metabolite profile similar to that of plasma collected in conventional tubes. An additional issue was to evaluate the metabolite differences between serum and plasma collected from the same group of individuals in conventional tubes. These differences affected to critical pathways such as the citric acid cycle, the metabolism of amino acids, the fructose and mannose metabolism and that of glycerolipids, and pento-se/glucuronate interconversions.

Sample preparation

(iv) The influence of sample preparation for lipidomics analysis of polar lipids in adipose tissue was studied. According to these results, the recommended sample preparation for analysis of polar lipids in adipose tissue would be LLE combined with an SPE step to enhance detection of glycerophospholipids. Concerning the extractant, MTBE favored the detection of less abundant lipids such as ceramides and unsaturated fatty acids and, therefore, it would be recommended for untargeted analysis of polar lipids.

(v) The influence of sample preparation on the identification coverage for pig fecal samples analysis by LC–MS/MS and GC–MS has been evaluated. A total number of 303 compounds by combination of all the extractants and analytical platforms were tentatively identified. According to the results obtained, it should be recommended the utilization of MeOH/water as extractant for GC–MS analysis, but for LC–MS/MS analysis the combined analysis of extracts obtained with MeOH or MeOH/water and ethyl acetate can lead to a significant increase of identified compounds. Concerning the derivatization step, the implement-tation of methoximation previous silylation provided the identification of three α-keto acids that are not detected by the other tested strategies. Regarding the complementarity of the

two analytical platforms, it is obvious that both approaches should be combined to obtain a comprehensive view of the pig feces metabolome.

Data analysis

(vi) A tool for correction of experimental variability associated to the instrumental quantitative response has been developed for implementation in metabolomics workflows based on MS detection. The proposed package is based on functions that can be used to correct variability on data sets obtained in metabolomics studies with large set of samples. The strategy included in this package involves that each metabolite is corrected according to the function that best fits its variability trend. Therefore, correction is independently applied to each metabolite. The only requirement for its application is the implementation of QCs, preferentially prepared with the same samples of the cohort, following a given planning in the sequence of analysis.

3. Development of strategies for searching potential biomarkers in clinical and agro-food applications.

Clinical applications

(vii) Postprandial alterations in the level of plasma metabolites after the OFTT were studied by combination of LC–QTOF MS/MS and GC–TOF/MS. The most important metabolic alterations affected inflammatory and oxidative processes, synthesis of primary and secondary bile acids and cortisol production. This study revealed that OFTT can be used to interpret deviations associated to metabolic diseases, increasing its usefulness.

(viii) The capability of metabolic changes occurring in the OFTT postprandial to predict the development of T2DM has been assessed. Taking into account the complexity of T2DM pathogenesis, two multimetabolite panels were configured to identify future T2DM patients. The combination of the two panels led to a model with sensitivity of 86.6% and specificity of 71.6%. The HR obtained for both panels, 5.4 (3.0–9.6) and 6.5 (3.7–11.4), revealed the predictive power, which reflected metabolic alterations associated to oxidation, insulin secretion and mitochondrial and peroxisomes activity.

(ix) Metabolic patterns occurring in PBMCs of children developing pancreatic β-cell autoimmunity or overt T1DM have been studied. Pathway

analysis suggested that alanine, aspartate, glutamate, glycerol phospholipid and sphingolipid metabolism were overrepresented in PT1D. Genome-scale metabolic models of PBMCs in T1DM progression were developed using available transcriptomics data and constrained with metabolomics results. Metabolic modeling confirmed altered ceramide pathways as specifically associated with T1DM progression.

Agro-food application

(x) The statistical combination of the results obtained by different analytical methods (concretely fatty acids determination by GC–FID and $\delta^{13}C$ by IRMS analysis) could be the key for the correct discrimination of feeding regimes in the sector of Iberian pig. Taking into account the demand for methods to discriminate feeding regimes, this approach could be implemented in routine laboratories since the analysis of fatty acids is already carried out at this level, while IRMS analysis is frequently used in reference food laboratories.

4. References

1. M.A. López-Bascón, M. Calderón-Santiago, F. Priego-Capote. Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry. Anal. Chim. Acta 943 (2016) 82–88.

2. M.A. López-Bascón, F. Priego-Capote, A. Peralbo-Molina, M. Calderón-Santiago, M.D. Luque de Castro. Influence of the collection tube on metabolomic changes in serum and plasma. Talanta 150 (2016) 681–689.

3. M.A. López-Bascón, M. Calderón-Santiago, J. Sánchez-Ceinos, A. Fernández-Vega, R. Guzmán-Ruiz, J. López-Miranda, M.M. Malagon, F. Priego-Capote. Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue. Talanta 177 (2018) 86–93.

4 M.A. López-Bascón, M. Calderón-Santiago, H. Argüello, J.J Garrido, F. Priego-Capote. Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage. Talanta 199 (2019) 303–309.

5. M. Calderón-Santiago, M.A. López-Bascón, A. Peralbo-Molina, F. Priego-Capote. MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls. Talanta 174 (2017) 29–37.

6. P. Sen, A.M. Dickens, M.A. López-Bascón, T. Lindeman, E. Kemppainen, S. Lamichhane, T. Rönkkö, J. Ilonen, J. Toppari, R. Veijola, H. Hyöty, T. Hyötyläinen, M. Knip, M. Orešič. Metabolic Alterations in Human Peripheral Blood Mononuclear Cells Associate with Progression to Islet Autoimmunity and Type 1 Diabetes. BioRxiv 658500; doi: https://doi.org/10.1101/658500.

7. M.A. López-Bascón, F. Priego-Capote, M. Calderón-Santiago, V. Sánchez de Medina, J.M. Moreno-Rojas, J.M. García-Casco, M.D. Luque de Castro. Determination of fatty acids and stable carbon isotopic ratio in subcutaneous fat to identify the feeding regime of Iberian pigs. J. Agric. Food Chem. 63 (2015) 692–699.

RESUMEN

1. Introducción

El objetivo primordial de la metabolómica consiste en analizar el conjunto de compuestos de bajo peso molecular presentes en un fluido biológico, célula, tejido u organismo, en unas condiciones fisiológicas específicas o en respuesta a diferentes perturbaciones o estímulos. Las principales limitaciones del análisis metabolómico están relacionadas con la identificación de los metabolitos, ya que las bases de datos existentes no están tan actualizadas como las del resto de disciplinas ómicas. Por otro lado, no existen protocolos universales para abordar el análisis metabolómico de determinadas muestras, ya que aún quedan aspectos por estandarizar. Entre ellos cabe destacar la obtención de información que permita la identificación inequívoca de metabolitos, ya que incluso para la espectrometría de masas, la técnica de mayor poder de detección, no hay un protocolo que garantice un buen nivel de identificación de compuestos sin una formación exhaustiva del personal implicado en los análisis. Otro aspecto poco establecido es la aplicación de protocolos de validación que aseguren la calidad analítica de los resultados generados.

Casi todas las limitaciones de la metabolómica son consecuencia de la gran complejidad de las muestras biológicas, que incluyen una amplia variedad de metabolitos de diversa naturaleza química y que cubren un rango de concentraciones de varios órdenes de magnitud. Por este motivo, es técnicamente inviable el desarrollo de un método analítico que en un único análisis proporcione resultados de metabolitos como, por ejemplo, la glucosa, de tipo polar presente a concentración milimolar, y el calcitriol, metabolito no polar presente a nivel picomolar. En este sentido la presencia de compuestos mayoritarios complica la detección de los metabolitos presentes a baja concentración.

Resulta evidente que existe una demanda de metodologías más sensibles, selectivas y precisas, que proporcionen mayor capacidad de detección de metabolitos (medida por lo que se denomina cobertura metabolómica del término inglés “metabolomics coverage”).

Con estas premisas, se puede decir que el mayor reto actual de la metabolómica es maximizar la capacidad de detección de los métodos analíticos con el fin de conseguir la identificación inequívoca de los miles de metabolitos existentes en un organismo. Este reto es actualmente utópico cuando se trabaja con organismos complejos pues no existe una única metodología que permita alcanzar el nivel de detección de algunos metabolitos en organismos superiores.

Por ello, la principal motivación de la Tesis Doctoral fue buscar soluciones para los distintos retos a los que se enfrenta actualmente la metabolómica.

2. Contenido de la investigación

El *objetivo básico* de la investigación en esta Tesis fue desarrollar nuevas estrategias analíticas basadas en el uso de espectrometría de masas de baja y alta resolución para mejorar la detección y la cobertura de identificación. Este objetivo se dividió en tres *objetivos*

generales de acuerdo con los diferentes temas de esta investigación: (i) aprovechar la versatilidad del analizador de triple cuadrupolo (QqQ) para mejorar la identificación/cuantificación de ciertas familias de metabolitos; (b) desarrollar nuevas herramientas para mejorar la detección e identificación de metabolitos mediante técnicas cromatográficas acopladas a espectrometría de masas de alta resolución; y (c) crear estrategias para buscar biomarcadores potenciales en estudios clínicos y agroalimentarios. En este contexto, la Tesis ha dado lugar a los siguientes resultados:

- Desarrollo de un método cualitativo y cuantitativo automatizado basado en el acoplamiento en línea de la extracción en fase sólida (SPE) y la cromatografía líquida con detección por espectrometría de masas en tándem (LC–MS/MS) para maximizar la sensibilidad en la determinación de ésteres de ácidos grasos y ácidos grasos hidroxilados (FAHFAs) en suero. El método se aplicó a una cohorte de individuos para evaluar la influencia de la glicemia en los niveles de FAHFAs [1].

- Propuesta de una estrategia cualitativa/cuantitativa para la determinación de lípidos polares en plasma humano por LC–MS/MS. Se combinaron dos métodos de adquisición de datos MS/MS para identificar y confirmar la presencia de lípidos polares en plasma. La propuesta se llevó a cabo en dos pasos: a) identificación de lípidos a través del patrón de fragmentación característico para cada familia; y b) confirmación de los lípidos detectados mediante la monitorización de iones producto correspondientes a los ácidos grasos (FAs) que los conforman u otros iones característicos.

- Estudio sobre las diferencias a nivel de metabolitos entre suero y plasma obtenidos con tubos convencionales (tubo de heparina para plasma) y tubos con gel polimérico mediante la aplicación de un enfoque no dirigido basado en cromatografía de gases con detección por espectrometría de masas con analizador de tiempo de vuelo (GC–TOF/MS). Se seleccionó una cohorte de voluntarios para el muestreo de sangre utilizando cuatro tipos de tubos (plasma, plasma-gel, suero y suero-gel) [2].

- Evaluación de la influencia de la preparación de la muestra en la determinación de lípidos polares en tejido adiposo visceral. Se probaron dos disolventes para comparar su eficiencia en la extracción de lípidos polares, pero también su ineficiencia para la extracción de acilglicéridos (las principales interferencias en la detección de lípidos polares). Además, se evaluó la implementación de una etapa SPE con un sorbente selectivo para la retención de glicerofosfolípidos con el fin de verificar su influencia en la detección posterior de esta familia de lípidos [3].

- Desarrollo de una estrategia para maximizar la cobertura de metabolitos identificados en muestras fecales de cerdo a través del estudio de la preparación de la muestra. Con este propósito se combinaron dos técnicas de detección, LC–QTOF MS/MS y GC–TOF/MS, para evaluar su complementariedad en términos de identificación. Con respecto a la preparación de la muestra, se probaron seis disolventes con diferente polaridad para evaluar su rendimiento de extracción y, en el caso de GC–MS, se compararon dos protocolos de derivatización [4].

- Diseño y desarrollo de un nuevo paquete estadístico, llamado MetaboQC, para estudiar y filtrar la variabilidad instrumental en conjuntos de datos generados mediante análisis por espectrometría de masas en secuencias desarrolladas durante varios días. Esta nueva herramienta utiliza controles de calidad (QCs) para corregir individualmente cualquier tendencia en las señales cuantitativas de metabolitos que puedan estar asociadas a la variabilidad instrumental [5].
- Estudio, mediante un análisis metabolómico no dirigido, de la respuesta posprandial a la prueba oral de tolerancia a la grasa (OFTT) en el perfil metabólico plasmático. Las muestras de plasma recolectadas se analizaron por LC–QTOF MS/MS y GC–TOF/MS. Los resultados de este estudio abren una vía al uso de este test para el diagnóstico de un amplio abanico de patologías.
- Evaluación de la capacidad predictiva de la aparición de diabetes mellitus tipo 2 (T2DM) mediante el examen de la respuesta posprandial (después de la OFTT). En este estudio se recogieron muestras de plasma de 215 pacientes (proyecto CORDIOPREV) justo antes y cuatro horas después de la prueba OFTT. 107 personas desarrollaron diabetes después de cinco años. Las muestras de plasma se analizaron por LC–QTOF MS/MS y GC–TOF/MS.
- Dilucidación de los eventos que preceden al inicio de la autoinmunidad de los islotes y la diabetes mellitus tipo 1 (T1DM). Se utilizó la metabolómica para determinar los niveles de lípidos moleculares y metabolitos polares en células mononucleares de sangre periférica humana (PBMC) aisladas de muestras prospectivas recolectadas en el estudio de Predicción y Prevención de Diabetes Tipo 1 (DIPP) [6].
- Desarrollo de modelos de discriminación y búsqueda de paneles de marcadores con capacidad para la clasificación de cerdos por su régimen de alimentación. Se utilizaron 80 muestras de tejido adiposo subcutáneo de cerdos ibéricos sometidos a cuatro regímenes de alimentación. Los datos se obtuvieron combinando el método clásico para la determinación de FAs basado en GC–FID y un método para la determinación de las abundancias isotópicas de carbono por espectrometría de masas con relación isotópica (IRMS) [7].

3. Conclusiones

Las conclusiones más destacadas de la Tesis Doctoral de acuerdo con los objetivos inicialmente propuestos son los siguientes:

1. Beneficios de utilizar la versatilidad del analizador de triple cuadrupolo para mejorar la identificación/cuantificación de ciertas familias de metabolitos.

    (i) Se ha desarrollado un método automatizado basado en SPE en línea acoplada a LC–MS/MS para la determinación de FAHFAs en suero. El método permitió identificar y cuantificar en términos relativos 11 FAHFAs en suero mediante el desarrollo de una estrategia de análisis confirmatorio. Los FAHFAs más concentrados en suero fueron PAHSA y PAHOA. Se han

detectado diferencias significativas en los niveles de PAHPA y POHPO en función del estado glucémico, y de POHPA en función del BMI [1].

(ii) Se ha optimizado una estrategia de análisis para la identificación y cuantificación masiva de lípidos polares por LC–MS/MS con un analizador QqQ. La combinación de dos métodos MRM permitió monitorizar 398 lípidos polares en 64 minutos (32 minutos por cada método) después de la aplicación de un protocolo de preparación de muestra simple. La estrategia propuesta puede usarse en el análisis lipidómico de muestras biológicas para obtener información cualitativa y semicuantitativa de familias de lípidos polares.

2. Mejora del proceso analítico (muestreo, preparación de muestra y análisis de datos) a través del desarrollo metodológico en análisis metabolómico no dirigido.

Muestreo

(iii) Es esencial tener en cuenta las alteraciones que se producen en los protocolos experimentales de análisis metabolómico cuando se toman muestras de sangre. Se ha comparado suero y plasma recogidos en tubos con gel polimérico con suero y plasma obtenidos en tubos convencionales utilizando un enfoque no orientado mediante GC–TOF/MS. Se detectaron cambios significativos atribuibles al gel polimérico en suero, mientras que no se observaron diferencias en el plasma, que proporcionó un perfil de metabolitos similar al del plasma recogido en tubos convencionales. Adicionalmente, se han evaluado las diferencias de metabolitos entre suero y plasma recolectado en tubos convencionales del mismo grupo de individuos. Estas diferencias afectaron a rutas importantes como el ciclo del ácido cítrico, el metabolismo de los aminoácidos, de la fructosa y la manosa y de los glicerolípidos, así como las interconversiones pentosa/glucuronato [2].

Preparación de la muestra

(iv) Se ha estudiado la influencia de la preparación de la muestra en el análisis lipidómico de lípidos polares en tejido adiposo. Según los resultados obtenidos, se recomienda la LLE seguida de una etapa SPE para mejorar la detección de glicerofosfolípidos. Respecto al extractante, el MTBE favoreció la detección de lípidos menos abundantes como ceramidas y ácidos grasos insaturados y, por lo tanto, se recomienda para el análisis no dirigido de lípidos polares [3].

(v) Se ha evaluado la influencia de la preparación de la muestra en la capacidad de detección de metabolitos en el análisis de muestras de heces de cerdo mediante LC–MS/MS y GC–MS. Se han identificado tentativamente un total de 303 compuestos mediante combinación de todos los extractantes y plataformas analíticas estudiadas. Según los resultados obtenidos, se recomienda la utilización de MeOH/agua como extractante para el análisis GC–MS. Sin embargo, para LC–MS/MS el análisis combinado de extractos obtenidos con MeOH o MeOH/agua y acetato de etilo puede dar lugar a un

aumento significativo de los compuestos identificados. Respecto a la etapa de derivatización por sililación, la implementación de una metoximación previa hizo posible la identificación de tres α-cetoácidos que no se detectan con las otras estrategias. Con respecto a la complementariedad de las dos plataformas analíticas, es obvio que ambas deben combinarse para obtener una visión integral del metaboloma de las heces de cerdo [4].

Análisis de datos

(vi) Se ha desarrollado una herramienta para la corrección de la variabilidad experimental asociada a la respuesta cuantitativa instrumental y su implementación en flujos de trabajo de metabolómica basados en la detección por MS. El paquete propuesto se basa en funciones que pueden usarse para corregir la variabilidad en los datos obtenidos en estudios de metabolómica con un gran conjunto de muestras. La estrategia incluida en este paquete implica que cada metabolito se corrige de acuerdo con la función que mejor se adapte a su tendencia de variabilidad. Por lo tanto, la corrección se aplica independientemente a cada metabolito [5].

3. Desarrollo de estrategias para la búsqueda de biomarcadores potenciales en aplicaciones clínicas y agroalimentarias.

Aplicaciones clínicas

(vii) Se han estudiado las alteraciones metabólicas producidas en el postprandio del OFTT, mediante la combinación de LC–QTOF MS/MS y GC–QTOF/MS. Las alteraciones metabólicas más importantes afectaron a los procesos inflamatorios y oxidativos, la síntesis de ácidos biliares primarios y secundarios, síntesis de carnitinas y a la producción de cortisol. Este estudio reveló que el OFTT se puede utilizar para interpretar las desviaciones asociadas a enfermedades metabólicas, aumentando la utilidad del OFTT.

(viii) Se ha evaluado la capacidad predictiva de los cambios metabólicos que ocurren en el postprandio del OFTT para predecir el desarrollo de T2DM. Teniendo en cuenta la complejidad de la patogénesis de T2DM, se configuraron dos paneles multimetabolitos para identificar futuros pacientes con T2DM. La combinación de los dos paneles condujo a un modelo con una sensibilidad del 86,6% y una especificidad del 71,6%. El HR obtenido para ambos paneles, 5.4 (3.0-9.6) y 6.5 (3.7-11.4), reveló su poder predictivo y el hecho de que reflejan alteraciones metabólicas asociadas a la oxidación, la secreción de insulina y la actividad mitocondrial y de los peroxisomas.

(ix) Se han estudiado los patrones metabólicos que se producen en las BMPC de niños que han desarrollado autoinmunidad en las células β pancreáticas o T1DM. El análisis de rutas metabólicas sugirió que el metabolismo de alanina, aspartato, glutamato, glicerofosfolípidos y esfingolípidos estaba sobreexpresado en PT1D. Se han desarrollado modelos metabólicos a escala del genoma de las CMSP en la progresión de la T1DM utilizando los datos

transcriptómicos disponibles y los resultados del análisis metabolómico. Los modelos desarrollados mostraron alteraciones en la ruta metabólica de las ceramidas, asociadas específicamente con la progresión de la T1DM [6].

Aplicaciones agroalimentarias

(x) La combinación estadística de los resultados obtenidos por diferentes métodos analíticos (concretamente la determinación de ácidos grasos por GC–FID y de δ13C por análisis IRMS) podría ser la clave para la correcta discriminación de los regímenes alimenticios en el sector del cerdo ibérico. Teniendo en cuenta la demanda de métodos para discriminar los regímenes de alimentación, este enfoque podría implementarse en laboratorios de rutina, ya que el análisis de ácidos grasos se ha realizado tradicionalmente con este fin mediante GC–FID, mientras que el análisis IRMS se usa con frecuencia en laboratorios alimentarios de referencia, por lo que los resultados de ambos conlleva una mayor seguridad en el diagnóstico [7].

4. Bibliografía

1. M.A. López-Bascón, M. Calderón-Santiago, F. Priego-Capote. Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry. Anal. Chim. Acta 943 (2016) 82–88.

2. M.A. López-Bascón, F. Priego-Capote, A. Peralbo-Molina, M. Calderón-Santiago, M.D. Luque de Castro. Influence of the collection tube on metabolomic changes in serum and plasma. Talanta 150 (2016) 681–689.

3. M.A. López-Bascón, M. Calderón-Santiago, J. Sánchez-Ceinos, A. Fernández-Vega, R. Guzmán-Ruiz, J. López-Miranda, M.M. Malagon, F. Priego-Capote. Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue. Talanta 177 (2018) 86–93.

4 M.A. López-Bascón, M. Calderón-Santiago, H. Argüello, J.J Garrido, F. Priego-Capote. Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage. Talanta 199 (2019) 303–309.

5. M. Calderón-Santiago, M.A. López-Bascón, A. Peralbo-Molina, F. Priego-Capote. MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls. Talanta 174 (2017) 29–37.

6. P. Sen, A.M. Dickens, M.A. López-Bascón, T. Lindeman, E. Kemppainen, S. Lamichhane, T. Rönkkö, J. Ilonen, J. Toppari, R. Veijola, H. Hyöty, T. Hyötyläinen, M. Knip, M. Orešič.

Metabolic Alterations in Human Peripheral Blood Mononuclear Cells Associate with Progression to Islet Autoimmunity and Type 1 Diabetes. BioRxiv 658500; doi: https://doi.org/10.1101/658500.

7. M.A. López-Bascón, F. Priego-Capote, M. Calderón-Santiago, V. Sánchez de Medina, J.M. Moreno-Rojas, J.M. García-Casco, M.D. Luque de Castro. Determination of fatty acids and stable carbon isotopic ratio in subcutaneous fat to identify the feeding regime of Iberian pigs. J. Agric. Food Chem. 63 (2015) 692–699.

INDEX

OBJECTIVES

OBJETIVOS

The ***basic objective*** of the research in this Thesis Book was to develop new analytical strategies based on the use of low- and high-resolution mass spectrometry to improve the detection and identification coverage in metabolomic analysis. These new strategies were applied throughout the main steps of the analytical process—sampling, sample preparation, determination and data analysis—and allowed improving basic analytical features such as sensitivity, selectivity and precision of metabolomic analysis methods (targeted and untargeted) and their detection capacity. The achievement of this basic objective leaded to metabolomic analysis methods capable of providing a higher level of information, which is a key milestone for the resolution of biological problems.

This objective was divided into three ***general objectives*** according to the different topics in this research:

- To take benefit from the versatility of the triple quadrupole mass spectrometer to improve the identification/quantification of certain families of metabolites.
- To develop approaches to improve the detection and identification of metabolites by chromatographic techniques coupled to mass spectrometry in high resolution mode.
- To create strategies for searching potential biomarkers in clinical and agro-food studies.

From each general objective, we defined several ***concrete objectives***:

(i) To develop an automated qualitative and quantitative method based on on-line coupling of solid phase extraction (SPE) and liquid chromatography–tandem mass spectrometry (LC–MS/MS) to

maximize sensitivity for determination of fatty acid esters of hydroxy fatty acids (FAHFAs) in serum (Chapter I). The method was further applied to a cohort of individuals to evaluate the influence of glycaemia on FAHFA levels.

(ii) To propose a qualitative/quantitative strategy for determination of polar lipids in human plasma by LC–MS/MS. Two MS/MS acquisition methods were combined to identify and confirm the presence of polar lipids in plasma (Chapter II). Thus, the process was carried out in two steps: (i) identification of lipids through the characteristic fragmentation pattern for each family; and b) confirmation of detected lipids by monitoring product ions corresponding to the fatty acids (FAs) conforming them or other characteristic product ions.

(iii) To study the differences at metabolite level between serum and plasma obtained with conventional tubes (heparin tube for plasma) and polymeric gel tubes by application of an untargeted approach based on gas chromatography coupled to time-of-flight mass spectrometry (GC–TOF/MS) (Chapter III). A cohort of volunteers was selected for blood sampling using four different tubes (plasma, plasma-gel, serum and serum-gel).

(iv) To evaluate the influence of sample preparation on the determination of polar lipids in visceral adipose tissue (Chapter IV). Two different extractants were tested to compare their efficiency for the extraction of polar lipids, but also their inefficiency for extraction of acylglycerides (the main interference in the detection of polar lipids). Additionally, the implementation of an SPE step with a selective sorbent for retention of glycerophospholipids was assessed to check its influence on the subsequent detection of this family of lipids.

(v) To maximize the identification coverage of metabolites found in pig fecal samples through the study of sample preparation (Chapter V). For this purpose, two analytical platforms such as LC–QTOF MS/MS

and GC–TOF/MS were combined to evaluate their additivity in terms of identification. Concerning sample preparation, six solvents with different polarity were tested to evaluate the extraction performance and, in case of GC–MS, two derivatization protocols were compared.

(vi) To develop a new statistical package, called MetaboQC, to study and filtrate instrumental variabililty in data sets generated by MS analysis of sequences processed for several days (Chapter VI). This new tool uses quality controls (QCs) to individually correct any tendency on quantitative signals of metabolites that can be associated to instrumental variability.

(vii) To study, by untargeted metabolomics analysis, the postprandial response to the oral fat tolerance test (OFTT) on plasma metabolomic profile (Chapter VII). Collected plasma samples were analyzed by LC–QTOF MS/MS and GC–TOF/MS. This test can open possibilities for the application of OFTT to the diagnostic of a wide range of pathologies.

(viii) To evaluate the predictive capacity of type 2 diabetes mellitus (T2DM) occurrence by examining the postprandial response (after OFTT) (Chapter VIII). With this aim, plasma samples were collected from 215 patients (CORDIOPREV project) at baseline and four hours after the OFTT. 107 individuals developed diabetes after five years. Collected plasma samples were analyzed by LC–QTOF MS/MS and GC–TOF/MS.

(ix) To elucidate the early events preceding the onset of islet autoimmunity and overt type 1 diabetes mellitus (T1DM). Metabolomics was used to determine levels of molecular lipids and polar metabolites in human peripheral blood mononuclear cells (PBMCs) isolated from prospective samples collected in the Type 1 Diabetes Prediction and Prevention (DIPP) study (Chapter IX).

(x) To develop discrimination models and search for panels of markers with capability to classify slaughtered pigs by their feeding regime (Chapter X). 80 samples of subcutaneous adipose tissue from Iberian pigs subjected to four different feedings were used. Data were obtained from the classical method for the determination of FAs based on GC–FID and from a method for determination of carbon isotopic abundances by isotope ratio mass spectrometry (IRMS).

The formation of the future PhD, which is the ***final objective*** of a Doctoral Thesis, has also included the master on "Chemistry" (Analytical Chemistry Specialization), in which the PhD student developed the mandatory courses. Also, the necessary steps to fulfill the requirements to achieve the International Doctorate mention were developed. In parallel to the above-mentioned tasks and to the research in the main part of this Book, a wider formation of the PhD student has been sought by development of other activities summarized below as annexes:

- Annex I: Collaborations with other members of the group and with other group, which has provided 2 published articles in high-impact international journals.
- Annex II: Book chapter on a subject non-related to the Thesis: Soxhlet extraction.
- Annex III: Oral and poster communications in national or international meetings.
- Annex IV: Co-direction of two Final Degree Projects (TFGs) of Degree in Chemistry students (University of Córdoba, Spain).
- Annex V: Co-direction of two Final Master Projects (TFMs) of Master in Chemistry students (University of Córdoba, Spain).

- Annex VI: Simultaneous research in education, which has provided 1 published article and participation in a teaching innovation project.

El ***objetivo básico*** de la investigación recogida en esta Memoria de Tesis fue desarrollar nuevas estrategias analíticas basadas en el uso de espectrometría de masas de baja y alta resolución para mejorar la detección y la cobertura de identificación en metabolómica. Estas nuevas estrategias se han apliado a lo largo de las principales etapas del proceso analítico (muestreo, preparación de la muestra, detección y análisis de datos) y permitieron mejorar características básicas como la sensibilidad, la selectividad y la precisión de los métodos de análisis metabolómico (dirigido y no dirigido) y su capacidad de detección. El logro de este objetivo básico condujo a métodos de análisis metabolómico capaces de proporcionar un mayor nivel de información, que es un hito clave para la resolución de problemas biológicos.

Este objetivo se dividió en tres ***objetivos generales*** de acuerdo con los diferentes temas de esta investigación:

- Aprovechar la versatilidad del analizador de triple cuadrupolo (QqQ) para mejorar la identificación/cuantificación de ciertas familias de metabolitos.
- Desarrollar nuevas herramientas para mejorar la detección e identificación de metabolitos mediante técnicas cromatográficas acopladas a espectrometría de masas de alta resolución.
- Crear estrategias para buscar biomarcadores potenciales en estudios clínicos y agroalimentarios.

Cada objetivo general ha dado lugar a varios ***objetivos concretos***:

(i) Desarrollar un método cualitativo y cuantitativo automatizado basado en el acoplamiento en línea de extracción en fase sólida (SPE) y cromatografía líquida con detección por espectrometría de masas en tándem (LC–MS/MS) para maximizar la sensibilidad en la

determinación de ésteres de ácidos grasos y ácidos grasos hidroxilados (FAHFAs) en suero (Capítulo I). El método se aplicó a una cohorte de individuos para evaluar la influencia de la glicemia en los niveles de FAHFAs.

(ii) Proponer una estrategia cualitativa/cuantitativa para la determinación de lípidos polares en plasma humano por LC–MS/MS. Se combinaron dos métodos de adquisición MS/MS para identificar y confirmar la presencia de lípidos polares en plasma (Capítulo II). Esta propuesta se llevó a cabo en dos pasos: a) identificación de lípidos a través del patrón de fragmentación característico para cada familia; y b) confirmación de los lípidos detectados mediante la monitorización de iones producto correspondientes a los ácidos grasos (FAs) que los conforman u otros iones característicos.

(iii) Estudiar las diferencias a nivel de metabolitos entre suero y plasma obtenidos con tubos convencionales (tubo de heparina para plasma) y tubos de gel polimérico mediante la aplicación de un enfoque no dirigido basado en cromatografía de gases con detección por espectrometría de masas con analizador de tiempo de vuelo (GC–TOF/MS) (Capítulo III). Se seleccionó una cohorte de voluntarios para el muestreo de sangre utilizando cuatro tipos de tubos (plasma, plasma-gel, suero y suero-gel).

(iv) Evaluar la influencia de la preparación de muestra en la determinación de lípidos polares en tejido adiposo visceral (Capítulo IV). Se probaron dos disolventes para comparar su eficiencia en la extracción de lípidos polares, pero también su ineficiencia para la extracción de acilglicéridos (las principales interferencias en la detección de lípidos polares). Además, se evaluó la implementación de una etapa SPE con un sorbente selectivo para la retención de glicerofosfolípidos con el fin de verificar su influencia en la detección posterior de esta familia de lípidos.

(v) Maximizar la cobertura de metabolitos identificados encontrados en muestras fecales de cerdo a través del estudio de preparación de muestra (Capítulo V). Con este propósito, se combinaron dos plataformas analíticas como LC–QTOF MS/MS y GC–TOF/MS para evaluar su complementariedad en términos de identificación. Respecto a la preparación de la muestra, se probaron seis disolventes con diferente polaridad para evaluar el rendimiento de la extracción y, en el caso de GC–MS, se compararon dos protocolos de derivatización.

(vi) Desarrollar un nuevo paquete estadístico, llamado MetaboQC, para estudiar y filtrar la variabilidad instrumental en conjuntos de datos generados mediante análisis por espectrometría de masas en secuencias desarrolladas durante varios días (Capítulo VI). Esta nueva herramienta utiliza controles de calidad (QCs) para corregir individualmente cualquier tendencia en las señales cuantitativas de metabolitos que puedan estar asociadas a la variabilidad instrumental.

(vii) Estudiar, mediante análisis metabolómico no dirigido, la respuesta posprandial a la prueba oral de tolerancia a la grasa (OFTT) en el perfil metabólico plasmático (Capítulo VII). Las muestras de plasma recolectadas se analizaron por LC–QTOF MS/MS y GC–TOF/MS. Este estudio permite abrir la posibilidad de usar dicho test para el diagnóstico de un amplio abanico de patologías.

(viii) Evaluar la capacidad predictiva de la aparición de diabetes mellitus tipo 2 (T2DM) mediante el examen de la respuesta posprandial (después de la OFTT) (Capítulo VIII). Con este objetivo se recogieron muestras de plasma de 215 pacientes (proyecto CORDIOPREV) justo antes y cuatro horas después de la prueba OFTT al inicio del estudio. 107 personas desarrollaron diabetes después de cinco años. Las muestras de plasma recolectadas se analizaron por LC–QTOF MS/MS y GC–TOF/MS.

(ix) Dilucidar los eventos que preceden al inicio de la autoinmunidad de los islotes y la diabetes mellitus tipo 1 (T1DM) (Capítulo IX). Se utilizó la metabolómica para determinar los niveles de lípidos moleculares y metabolitos polares en células mononucleares de sangre periférica humana (PBMC) aisladas de muestras prospectivas recolectadas en el estudio de Predicción y Prevención de Diabetes Tipo 1 (DIPP).

(x) Desarrollar modelos de discriminación y buscar paneles de marcadores con capacidad para clasificar a los cerdos por su régimen de alimentación (Capítulo X). Se utilizaron 80 muestras de tejido adiposo subcutáneo de cerdos ibéricos sometidos a cuatro regímenes de alimentación diferentes. Los datos se obtuvieron del método clásico para la determinación de FAs basado en GC–FID y de un método para la determinación de las abundancias isotópicas de carbono por espectrometría de masas con relación isotópica (IRMS).

La formación de la futura doctora, que es el ***objetivo final*** de una Tesis Doctoral, también ha incluido el máster en "Química" (especialidad de Química Analítica), en el que la estudiante de doctorado desarrolló los cursos obligatorios. Además, se completaron los pasos necesarios para cumplir con los requisitos para lograr la mención del Doctorado Internacional. Paralelamente a las tareas indicadas anteriormente y a la investigación en la parte principal de la Memoria, se ha buscado una formación más amplia de la estudiante de doctorado mediante el desarrollo de otras actividades resumidas a continuación como anexos:

- Anexo I: Colaboración con otros miembros del grupo y con otro grupo de investigación, que ha proporcionado 2 artículos publicados en revistas internacionales de alto impacto.

- Anexo II: Capítulo de libro sobre un tema no relacionado con la Tesis: extracción Soxhlet.

- Anexo III: Comunicaciones orales y pósters en 20 congresos nacionales o internacionales.

- Anexo IV: Codirección de dos Trabajos Fin de Grado (TFG) de estudiantes de Grado de Química (Universidad de Córdoba, España).

- Anexo V: Codirección de dos Trabajos Fin de Máster (TFM) de estudiantes de Máster en Química (Universidad de Córdoba, España).

- Anexo VI: Investigación simultánea en el área de educación, que ha proporcionado 1 artículo publicado y la participación en un proyecto de innovación docente.

INTRODUCTION

This introduction section is intending to offer an overview of the metabolomics analysis and analytical steps considered in the research that constitutes this PhD Book.

1. Metabolome and metabolomics: basic principles and concepts

From the Latin suffix "ome", the term "omics" means mass or many and is used to differentiate studies that involve a large number of measured parameters, typically genes (genomics), RNA (transcriptomics), proteins (proteomics), or metabolites (metabolomics).

The term *metabolome* first appeared in September 1998, when Oliver *et al*. [1,2] measured the change in the relative concentrations of metabolites as the result of deletion or overexpression of a gene [1]. *Metabolome* was therefore used to address the entire set of metabolites an organism expresses [2]. In 2001, *metabolomics* was defined by Fiehn [3,4] as the comprehensive and quantitative analysis of all metabolites of the biological system under study. In addition, he pointed out that "metabolomic approaches must aim at avoiding exclusion of any metabolite by using well-conceived sample preparation procedures and analytical techniques" [4]. Currently, the most common definition of metabolomics is the systematic identification and quantification of the small molecule (low-molecular-weight, <1 kDa) metabolic products (the metabolome) of a biological system (cell, tissue, organ, biological fluid, or organism) at a specific point in time.

Endogenous metabolites can be classified into primary and secondary. The first ones are metabolism products generated during the growth phase of an organism to perform the physiological functions and support the overall development of the cell. On the other hand, secondary metabolites are the end

products of primary metabolism, synthesized after the growth phase has been completed and are important in ecological and other cell activities.

Currently, 92.000 endogenous metabolites have been identified in human samples according to the Human Metabolome Database (HMDB) [5,6]. These metabolites are a consequence of all processes involved from gene expression to phenotype manifestation, as shows Figure 1.A, which schematizes the traditional central dogma of system biology. As can be seen, the information flow is considered unidirectional and goes from genes to their transcription at mRNA and their translation to proteins, macromolecules in charge of catalyzing metabolic reactions that affect the metabolome.

Alterations in endogenous metabolite levels, which may result from disease processes, drug toxicity, or gene function, could be evaluated in cells, tissues, or biological fluids by metabolomics. Latent biochemical information obtained from metabolomics may be used for diagnostic or prognostic purposes. Such information reflects actual biological events rather than the potential for disease development, data provided by gene expression [7,8].

However, to understand biology at the system level, it is necessary to examine all cell and organism functions, rather than isolated parts of them. The integration of data from the primary omics (*viz.*, genomics, transcriptomics, proteomics and metabolomics) constitutes what was known as systems biology, term coined by Nicholson and Wilson [9,10] to better understand the functioning of a given biological system. At present, systems biology is more frequently known as integrative omics [8,10]. Figure 1.B includes the scheme of the systems biology approach. As can be seen, the biological information flow from the central dogma of molecular biology is complemented with interaction among the different levels, breaking the traditional unidirectional flow of information. Thus, the flow of information takes place not only downstream but also upstream and explains the phenotype of an individual as the direct reflection of its metabolism. Furthermore, lower levels can provide information of upper levels that justify the situation of the former.

The metabolome is the level of this functional cascade that best reflects the physiological state of an organism, being not only the most sensitive to any change, either internal or external, since the metabolites are the real active regulating agents of homeostasis [11,12]. For example, alterations in a single enzyme can lead to a cascade of metabolic perturbations that are functionally related to the given phenotype [12]. Hence, metabolomics is one of the most powerful bioanalytical strategies as it allows obtaining a picture of the metabolites of an organism in the course of a biological process, being considered as a phenotyping tool, thus justifying its usefulness [12].

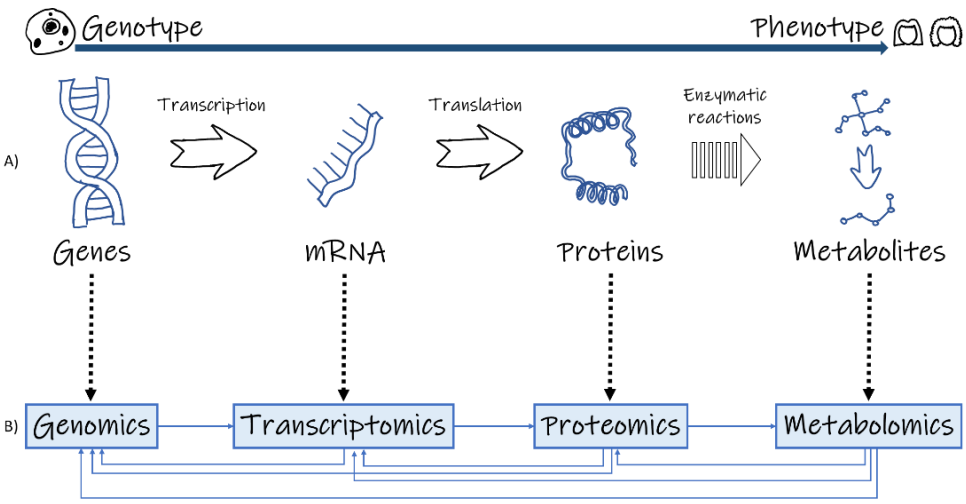


Figure 1. A) Traditional central dogma of molecular biology. B) General scheme of systems biology.

Metabolomics includes different subdisciplines targeted at the study of specific groups of metabolites. Among them, lipidomics is the one with the most salient contributions. Lipidome first appeared in the literature in 2001 [13] and it has greatly advanced in recent years, largely due to the development of mass spectrometry (MS) [14–16]. The lipidome is composed by all biomolecules defined

as lipids, which encompass compounds of very different structural diversity and complexity [17]. Thus, lipidomics can be defined as the study of the structure and function of the complete set of lipids (lipidome) in a given cell or organism, as well as their interactions with other cellular components [16,18].

2. Metabolomic strategies

Metabolomics can be used for two major purposes: (i) to detect differences between global metabolic fingerprints of groups of individuals, or (ii) to understand given metabolic pathways, families or partitioning of metabolic products between cellular compartments and excretion [19]. Metabolomic analysis encompasses different strategies, which depend on the information required from the system under study [3]:

a. Targeted analysis, which aims at qualitative and quantitative study of one or, more frequently, a small group of chemically similar metabolites.

b. Untargeted analysis (also known as global metabolomics profiling), which allows detection of a broad range of metabolites by using a single analytical platform or a combination of complementary analytical platforms to obtain a comprehensive profile of the metabolome [20,21].

c. Metabolomics fingerprinting, a high throughput, fast methodology for analysis of biological samples that provides metabolic snapshots for sample classification and screening [22].

Obviously, the complexity of sample preparation for each strategy is different increasing from (c) to (a). The most recent trend leads to simplification by establishing distinction only between targeted and untargeted analysis [3,12].

Each of these strategies has its own inherent advantages and disadvantages, but they can be highly complementary when used in combination. The detection techniques used also depend on the selected strategy [8]. For example, to obtain a

metabolomic fingerprint the detection technique should allow direct and rapid analysis of the sample. Nuclear magnetic resonance (NMR) spectroscopy, MS (depending on the complexity of the sample) and, to a lesser extent, infrared and Raman spectroscopies, are main tools used in this context.

High-resolution separation approaches (usually GC or LC) are generally involved in both targeted and untargeted analysis prior to individual detection of the metabolites for their identification or quantification. Then, MS is the commonest detection technique in metabolomics, both for targeted and untargeted analysis, as it provides a high spectral resolution —and therefore a great accuracy in the measurement of the *m/z* ratio—, or an excellent sensitivity, depending on the given MS approach [8].

It is worth mentioning that both targeted and untargeted analysis have been applied for development of the research that constitutes the present Thesis Book.

3. General metabolomics workflow

In overall terms, the analytical process can be segmented into five steps: sample collection (sampling), sample preparation, detection, data processing, and statistical analysis. Figure 2 illustrates a general workflow of experiments in metabolomics. This workflow has been restricted to the most common strategies employed for targeted and untargeted analysis.

The workflow starts with sampling, which is usually the limiting step in metabolomics and, ideally, should seek to be non-invasive and ensure representativeness [23]. The next step is the isolation of metabolites, that depends on the selected metabolomic strategy but also on the type of sample. In general, solid–liquid (SLE) and liquid–liquid extraction (LLE) are commonly used for solid and liquid samples, respectively. In some cases, it is necessary to carry out an SPE to concentrate metabolites or remove interferents. For the considered metabolomic approaches, the derivatization step is necessary when the

metabolites are going to be separated by GC and they are not volatile enough; but it can be used also to increase metabolites stability, selectivity or sensitivity.

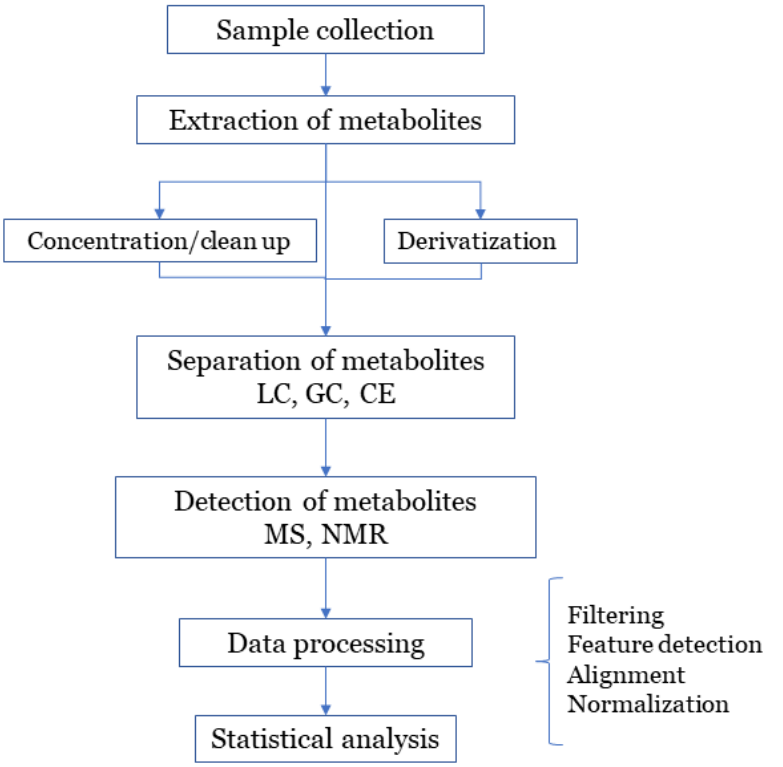


Figure 2. Basic scheme of the workflow in metabolomics analysis.

As can be seen in Figure 2, the next steps are separation and detection of metabolites. As mentioned above, only untargeted and targeted metabolomic analysis have been considered. In these cases, separation of the metabolites (using typically GC [24] or LC [25]) is usually required when working with mass spectrometry as detection technique. So, the most frequently used analytical platforms are NMR, GC–MS and LC–MS.

After detection, the next step is data processing followed by data analysis, two steps that are also dependent on the metabolomic strategy [20,26]. In the following sections of this introduction all the steps of this workflow are discussed.

3.1. Sampling in metabolomics

In metabolomics, an experiment starts with selection of the biological material and sampling, which are frequently limiting steps that affect the data quality and, therefore, the accuracy of the biological interpretation [27]. The main challenges in sampling dealing with metabolomics analysis are:

a. Selection of the biological material –usually biofluids (*e.g.*, blood or urine) or tissues. Ideally, sampling of biofluids should seek to be non-invasive and reproducible [23].

b. Selection of the number of samples. A power analysis should be performed to ensure that a sufficient number of samples are included in the study to reduce the influence of biological variability and obtain statistically validated data [21]. However, in clinical studies to get a large cohort with a specific pathology and similar characteristics is difficult. The influence of diet, gender, age, and genetic factors have to be considered [21].

c. Selection of the sampling strategy. Biofluids are collected by using different approaches. For example, plasma can be collected in tubes containing different commercial anticoagulant agents. Also, sampling conditions (including processing prior to sample preparation) can significantly affect metabolites stability and should be attentively evaluated.

The most common biofluid used in clinical and nutrition studies is blood (serum/plasma) [28,29]. The main reasons justifying the clinical applicability of blood are its minimally invasive sampling, its homogeneity as compared to saliva or urine, which are strongly influenced by the collection volume, and its direct relationship with systemic changes in the metabolome [28]. Plasma is obtained

from whole blood by addition of an anticoagulant followed by centrifugation to separate the aqueous plasma from the blood cells [29]. On the other hand, serum is achieved by collection into tubes without addition of anticoagulants, followed by clotting and centrifugation to eliminate not only blood cells, but also fibrinogen, then isolating the supernatant [29]. Collection of plasma is more reproducible as the extra step related to blood clotting is omitted [29].

Thus, the sampling strategy is specially of interest for blood, since there are several types of commercial tubes (with different tube wall, stopper, stopper lubricant, separator gel, clot activator, etc. [27]) for plasma or serum collection, which are widely used in metabolomics analysis. The selection of the commercial tubes is a key factor to be considered as it can introduce analytical bias and influence the quality of the observed metabolic profiles. This influence has been studied in the research of this Thesis (Chapter III). A similar situation was found with PBMCs used in Chapter IX. PBMCs are blood cells with a single round nucleus. These cells include T cells (~70%), B cells (~15%), monocytes (~5 %), dendritic cells (~1%) and natural killer (NK) cells (~10%). There are commercial kits with different anticoagulants for sampling and isolation of PBMCs, Figure 3 shows an example of PBMCs isolation from blood. Thus, in this case it is also important to select the most adequate commercial kit for each particular study.

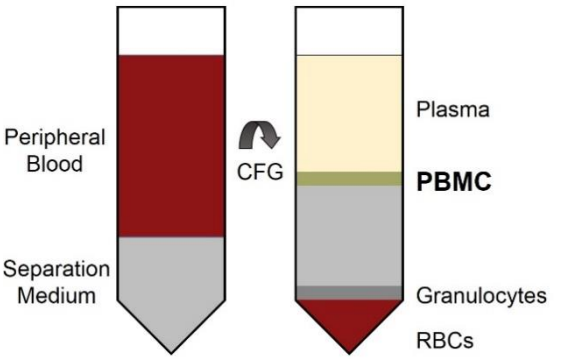


Figure 3. Example of PBMCs isolation from blood [30].

Concerning tissues as metabolomic samples, the sampling technique is conditioned by the location of the target tissue [31]. For example, adipose tissue is located in several anatomical locations. The abdominal adipose tissue (AT) is divided into subcutaneous adipose tissue (SCAT) and visceral adipose tissue (VAT) [32], being SCAT and VAT the most studied samples in metabolomic studies related to metabolic diseases [33–36]. In general, adipose tissue samples are collected by surgical procedure (incision), which allows collection of both VAT and SCAT, or by needle biopsy, the preferred option for SCAT [32,37]. Both VAT and SCAT are used in Chapters IV and X, respectively.

Apart from biofluids and tissues, there are other less common biological samples like feces, a particular biological material, easily accessible, that provides a non-invasive [27] window to study the outcome of the diet–gut microbiota–host interaction through the analysis of remaining unabsorbed metabolites [38]. In spite of the relevance of the information that could be obtained from feces, the analysis of fecal samples for metabolic profiling has received attention only in the last years. In general, the protocol for sampling human feces is the self-collection [39–41]. In some cases, feces are collected prior to surgery or endoscopic examination [41]. On the other hand, feces from animals are collected directly from their intestine after sacrifice, or as pellets from their cages [41]. In this Thesis, pig fecal samples were used in Chapter V.

3.2. Sample preparation in metabolomics

The selection of the sample preparation approach is crucial to the success of metabolomic experiments [23]. Sample preparation aims to enrich metabolites of interest and remove interfering substances. The ideal sample preparation method should be simple, robust and with capability to preserve the integrity of the metabolites [42]. The choice of a sample preparation method affects both the observed metabolite profile and data quality.

Different approaches for sample preparation can be used in metabolomics; these approaches include extraction of metabolites into a solvent, preconcentration, clean-up of interferents, and derivatization of metabolites if required. Sample preparation depends on the selected strategy and the platform used [26]. Thus, targeted metabolomic analysis requires a highly selective extraction protocol that provides clean and concentrated extracts preferably containing only the compounds of interest. The extraction protocol is mainly conditioned by the target biological sample: for solid samples, SLE is the preferred technique, while, for liquid samples, metabolites could be extracted by LLE, SPE or solid-phase microextraction (SPME) [26]. SPE and SPME play a dual role: preconcentration of metabolites and removal of interferents. On the other hand, in untargeted analysis an unselective sample preparation approach is selected by the need to analyze a range of metabolites as wide as possible [43]. This is generally accomplished by using the simplest sample preparation procedures; for example, dilution or solvent precipitation [43]. Metabolite losses are further avoided by minimizing the overall number of steps in the procedure, and short sample-preparation times facilitate high sample throughput [43].

Apart from metabolites extraction, additional sample preparation steps are needed in some cases. For example, the low volatility of many metabolites makes necessary derivatization (which is usually performed by silylation [26,44]) prior to GC–MS analysis.

The sample preparation approach has been of paramount importance in the development of the research planned in the present PhD, concretely for the optimization of methods for analysis of feces and adipose tissue.

3.3. Analytical platforms in metabolomics

Currently, a range of analytical platforms are used for metabolomic analysis, including NMR, direct infusion in MS (DI–MS), GC–MS, LC–MS and capillary electrophoresis coupled to MS (CE–MS) [45]. For the purpose of the Thesis, LC–

MS and GC–MS were preferentially used. In the following sections the separation and detection techniques used in the development of the Thesis are detailed.

a. Separation techniques used in metabolomics: LC and GC

The high number of metabolites in any vegetal or animal sample makes mandatory in targeted and untargeted methods an appropriate separation of the sample components before being subjected to detection. The main benefits ascribed to the implementation of separation techniques are the improvement of the sensitivity and the resolution power of the analysis. In addition, chromatographic separation provides extra information (retention time, RT) that facilitates the identification of metabolites, especially in untargeted metabolomic strategies.

The most used separation techniques in metabolomics are GC and LC, selected depending on the nature of the target metabolites. GC allows the separation of volatile compounds, such as aromatic compounds, but also compounds that can become volatile after derivatization. The type of derivatization depends on the chemical structure of the target metabolites. The majority of metabolites to be separated by GC requires a laborious sample preparation step (hydrolysis, derivatization); thus, its use for large-scale studies is more limited. GC is essential for determination of certain families of metabolites such as amino acids, fatty acids, carboxylic acids or carbohydrates. On the other hand, in LC the most determining component is the analytical column, whose characteristics define the chromatographic mode. Reverse-phase columns, mainly those packed with silica (C18 or C8), are characterized by strong interaction with low-polarity compounds, whereas normal phase columns, such as those used in hydrophilic interaction liquid chromatography (HILIC), are more effective for separation of polar compounds, thus covering a wide range of metabolites with very different features. However, most metabolomics applications use reverse-phase columns [45,46].

The combination of methods based on GC and LC provides a higher level of analytical information due to their complementarity. As examples, there are

metabolites that are not volatile and do not yield volatile products, do not ionize adequately at atmospheric pressure or are thermolabile. Therefore, both GC and LC should be combined to attain the best detection coverage in a biological sample.

b. Detection equipment used in metabolomics: mass spectrometry

Metabolomics analysis demands detectors with high sensitivity and wide dynamic range to detect metabolites present at low concentrations and with variable concentrations in complex matrices, and high-resolution power for identifying them [47]. Attending to these requirements, MS is the most competitive detection technique [21,48].

The main performance characteristics of a mass analyzer are [8,21,49,50]:

(a) Mass accuracy of the measured *m/z* value provided by the mass analyzer, directly related to the mass resolving power and stability of the instrument.

(b) Mass resolving power or ability of an MS to provide a specific value of mass resolution (*i.e.*, the instrument generates distinct signals for two ions with a small *m/z* difference).

(c) Mass range or limits of *m/z* within which an MS can detect ions or record a mass spectrum.

(d) Transmission efficiency or ratio of the number of ions reaching the detector and the number of ions leaving the mass analyzer, related to the sensitivity of the mass spectrometer (*i.e.*, the minimal concentration of a compound leading to a peak intensity greater than a specified signal-to-noise ratio).

(e) Scan speed or rate at which the analyzer measures over a certain mass range.

(f) Scan cycle time or the time required to obtain a mass spectrum, also

known as duty cycle.

Simplistically, a mass spectrometer consists of a sample inlet, an ion source, a mass analyzer, a detector and a data management system. The sample inlet has as function to introduce the sample molecules into the ion source, where they are ionized. The ion source generates gas-phase ions via an ionization technique, the mass analyzer separates the ions according to their mass-to-charge ratio (*m/z*), and the detector generates an electric current from the incident ions proportional to their abundances [21].

The required ionization of the analytes prior to MS detection can be produced by different devices. The most common ionization sources in metabolomics are electron impact ionization (EI) and chemical ionization (CI), commonly used in GC–MS, and electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI), frequently employed in LC–MS [21]. Ionization is crucial for detection of metabolites in MS, being directly related to sensitivity. Figure 4 shows the applicability of different types of ionization according to the polarity and molecular weight of the metabolites to be detected [51–53]. For the purpose of this Thesis, ESI and EI were used for LC–MS and GC–MS analysis, respectively. Thus, these two ionization sources are explained below.

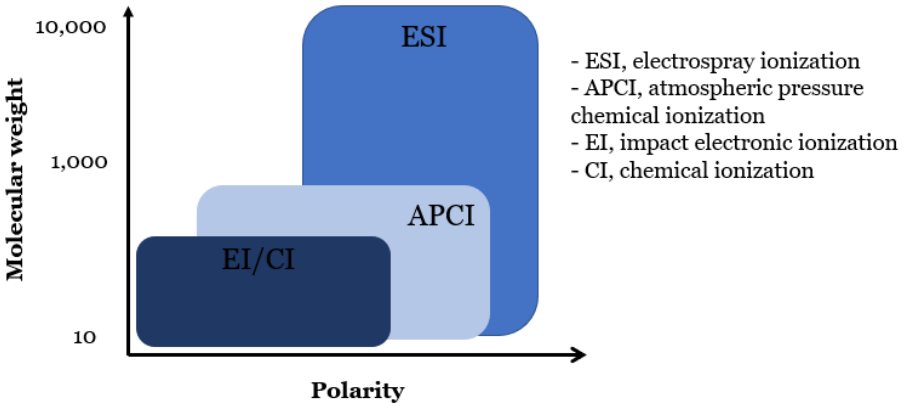


Figure 4. Applicability of ionization sources based on metabolite molecular weight and polarity.

(i) Electron impact ionization (EI)

EI is well established and is considered the most common ionization technique in GC–MS [53]. The sample in gas phase enters the ion source from the GC–MS interface. Electrons emitted by a filament enter the ionization chamber, guided by a magnetic field. The high energy electrons interact with the sample molecules, ionizing and fragmenting them. For this reason, EI is considered a strong ionization technique. The positive voltage on the repeller pushes the positive ions into the lens stack, where they pass through several electrostatic lenses. These lenses focus the ions on a tight beam, which is directed into the mass filter. EI typically produces single-charged molecular ions and fragment ions (smaller parts of the original molecules), which are used for structural elucidation [54].

(ii) Electrospray ionization (ESI)

ESI is a soft ionization technique in which the LC eluent is sprayed (nebulized) into a spray chamber at atmospheric pressure in the presence of a strong electrostatic field and heated drying gas. The electrostatic field occurs between the nebulizer, which is at ground in this design, and the capillary, which is at high voltage (Figure 5).

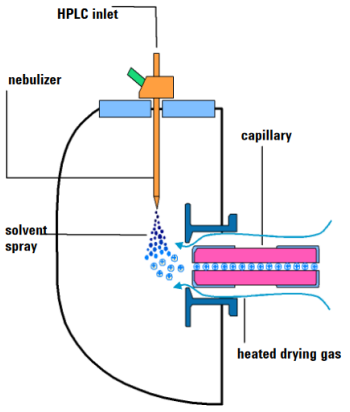


Figure 5. Electrospray ion source [53].

Once the sample has been ionized, it is transported to the mass spectrometer. There are several types of mass analyzers, but we will focus on those used in the research of this Thesis:

(i) Triple quadrupole mass spectrometer (QqQ)

This MS detector is commonly used for quantitative analysis in targeted metabolomics [2]. A QqQ mass detector consists of the parts shown in Figure 6: an ion source followed by a set of lenses for ion transfer to the first quadrupole (Q1), consisting of four parallel bars to which specific direct current values and radio frequency voltages are applied to filter ions with one or more *m/z* values. The applied voltage is variable, so that sequentially some ions can be filtered and only they pass reaching the collision cell in which they are fragmented. This cell, which is generally called the second quadrupole (q2), is actually a hexapole filled with an inert gas, nitrogen or argon, in which the ion is fragmented by application of a determined voltage, and the fragments are sent to the third quadrupole (Q3) in which a second filtering stage allows isolating fixed product ions. Since the fragments are parts of the precursor molecule, they represent portions of its overall structure. This mode is called Multiple Reaction Monitoring (MRM) and is the preferred mode for quantitative and confirmatory analysis of a set of metabolites. The MRM mode delivers excellent selectivity (minimizes interferences) and great reduction in background chemical noise (yields higher signal-to-noise).

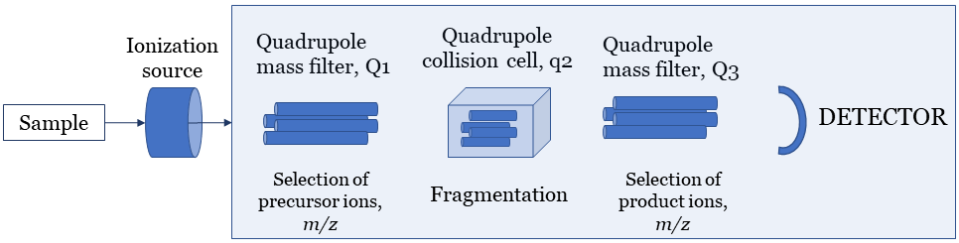


Figure 6. General scheme of a triple quadrupole mass spectrometer.

Nevertheless, the QqQ can be operated in other modes, some of them also used for confirmation in targeted analysis through the development of data–dependent methods (DDM). The different operating modes for QqQ are described below [53,55,56]:

- Full Scan mode. This mode provides a full spectrum with universal detection of known or unknown compounds. Only Q3 is working, and this mass analyzer scans a concrete range of masses without interruption in a given segment of time. Full scan experiments are used to determine the *m/z* of the precursor ion of a compound or a mixture of compounds.

- Selected Ion Monitoring mode (SIM). This mode is used for target detection of known compounds. Similarly to full scan mode, only Q3 is working, but in this case isolates one or a few selected *m/z* values. SIM generally provides higher sensitivity than the full scan mode as data acquisition is focused on concrete *m/z* values.

- Precursor Ion Scan mode (PrIS). This acquisition mode is used to monitor the presence of compounds with a common fragmentation pattern. For this purpose, the Q3 is tuned at one or several specific *m/z* values, which are produced by fragmentation of these compounds in q2. Complementarily, Q1 operates in scan mode. Thus, the chromatograms obtained in PrIS mode are representative for all the precursor ions that produce a selected product ion by activation.

- Product Ion Scan mode (PIS). Contrarily to the PrIS mode, ions formed in the ion source enter Q1 that filters precursor ions with specific *m/z* ratios. The selected precursor ions are then fragmented in the collision cell (q2), to yield product ions by meta-stable ion decomposition or by collision-induced dissociation. These ions enter Q3 operating in scan mode. Therefore, PIS chromatograms are useful to optimize MRM methods and to obtain MS/MS information for a list of precursor ions.

- Neutral Loss Scan mode (NLS). In this case, both the Q1 and Q3 are tuned by filtering over a fixed mass difference between precursor and

product ions. This mass difference corresponds to neutral losses produced in the fragmentation of the precursor ions in q2. Thus, NL chromatograms will contain peaks corresponding to precursor ions that loss a particular neutral mass by fragmentation in q2 and, thus, generate specific product ions that are filtered in Q3.

(ii) Quadrupole-time of flight (QTOF) mass spectrometer

Figure 7 shows the diagram of a QTOF mass analyzer. It is based on the same configuration as the QqQ, but replacing the last quadrupole by an acceleration tube as mass analyzer (usually in an orthogonal configuration) to filter out ions according to the kinetic energy equations.

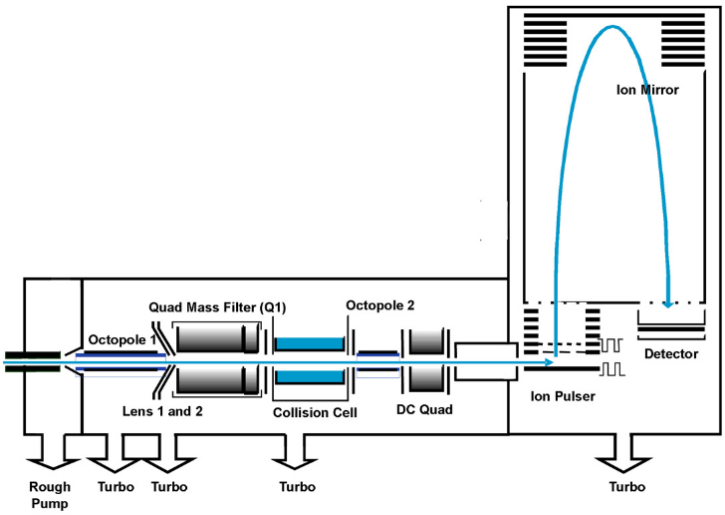


Figure 7. General scheme of a quadrupole time-of-flight mass spectrometer [57].

The QTOF can operate in MS mode with the TOF as scanning tool or in MS/MS mode for structural elucidation, in both cases by taking benefit from the high mass accuracy. This hybrid mass analyzer offers better selectivity than triple quadrupoles, meanwhile sensitivity is considerably lower. On the other hand, thanks to the good mass accuracy (below 2 ppm) highly reliable identification can be achieved, thus allowing its use for global metabolomic profiling [12].

(iii) Isotope-ratio mass spectrometer (IRMS)

IRMS is a specialized mass spectrometry mode because it measures the relative abundances of stable isotopes. In general, there are five main sections in an IRMS instrument: a sample introduction system, an electron ionization source, a magnetic sector analyzer, a Faraday-collector detector array, and a computer-controlled data acquisition system [58]. Several different interfaces are used to introduce samples into the IRMS, the most common being elemental analyzers (EA–IRMS) [58].

Figure 8 shows the diagram of EA–IRMS used in this Thesis. EA-IRMS is a bulk measurement technique that provides representative data for the average isotopic signal of the entire sample. Sample preparation in this technique is very simple as the bulk sample is weighed and placed into a tin or silver capsule, which is lowered into a combustion furnace through an autosampler carousel where the sample is combusted at elevated temperatures under a flow of oxygen into NO_x, CO_2, SO_2, or H_2O [58].

Depending on the isotopes of interest, the combustion products may need to be specifically treated to reduce interferents. In the case of the research in this Thesis, carbon was the atom of interest to obtain a carbon isotope ratio analysis. The combusted sample is transported by a helium gas stream into a reduction chamber where nitrous oxides are converted into N_2 and O_2 in excess is removed. The chemical species are carried through a chemical trap to remove water that was produced from combustion, and then into the gas chromatograph where CO_2 and N_2 are separated. Effluent from the elemental analyzer is then sent to the IRMS [58]. The detector is a Faraday cup, especially useful to measure highly precise ratios of specific ion species as in IRMS [50].

The ratios of these isotopes are always measured relative to an isotopic standard to eliminate any bias or systematic error in the measurements [50]. These standards are internationally recognized standards, such as the Vienna Pee Dee Belemnite (VPDB) for carbon analysis.

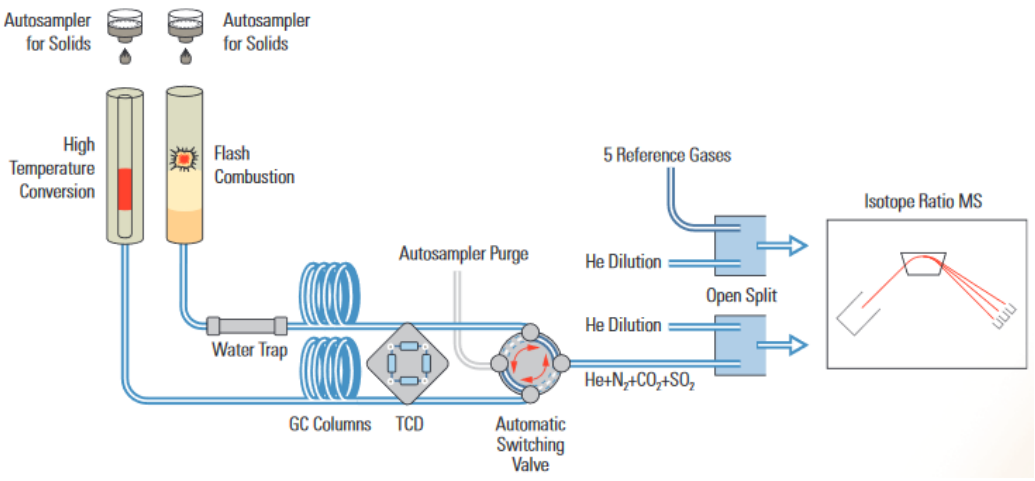


Figure 8. General scheme of an isotope ratio mass spectrometer [59].

3.4. Statistical analysis in metabolomics

Chemometrics is the discipline concerned with the application of statistical and mathematical methods to chemistry [60]. This discipline has been fundamental for the development of metabolomics, while growing with it [61,62]. Data treatment approaches employed in metabolomics are designed according to the selected strategy, being slightly different between targeted and untargeted analysis.

3.4.1. Data pretreatment

Data pretreatment is the first step of the workflow that attempts to prepare raw data for statistical analysis. Metabolomics data pretreatment encompasses several preprocessing steps crucial for data quality and interpretation of the results. The purpose of data pretreatment is the transformation of the raw data (those provided by the analytical instrument that are exclusive of its trademark) into clean data (universal format) for processing by suited softwares. The workflow

depends on the metabolomic strategy, with critical differences between targeted and untargeted analysis [20]:

a. Data pretreatment in targeted metabolomic analysis

In this case, the process is simple, since it mainly consists of obtaining the peak area corresponding to the target analytes. Later, these areas could be transformed into concentrations through calibration models and, then, different normalization strategies can be applied prior to statistical analysis. Thus, the data set contains the concentration of all metabolites in all samples from the cohort under study.

b. Data pretreatment in untargeted metabolomic analysis.

This scenario is more complex. The most critical step is the extraction from the raw data files of signals corresponding to all potential metabolites (typically called molecular features, MFs). The whole process could be divided into different steps [63]:

- Filtering: The amount of data adquired in untargeted analysis makes mandatory a filtering step to remove undesired signals. One of the most common operations is to remove all signals below a specific peak height (number of counts). Other frequent filters are targeted at specific RT or *m/z* ranges.
- Peak detection: the next step is to find potential metabolites present in the chromatogram. This process could be done by detection and extraction of potential MFs from LC–MS and GC–MS data. A critical task is the deconvolution, which is applied in both cases. However, for GC–MS data all ions with similar chromatographic profile are associated to a unique molecular feature, while for LC–MS data ions need to be also related through a main base peak. Therefore, all ions associated to an MF represent different adducts or neutral losses of the base peak.
- Peak alignment. After extraction of all MFs in the set of samples,

it is neccesary to carry out an alignment of signals among samples; thus providing a data set containing the peak area values of all MFs in all samples.

- Recursive extraction. The matrix obtained after alignment of chromatographic peaks could be improved through the implementation of an additional step for extraction of potential metabolites in samples in which they were not initially detected, probably because of the filtering step. This treatment allows reducing missing values, which is a crucial task prior to statistical analysis.
- Identification of metabolites. There are different databases that could be employed for tentative identification of metabolites. In the case of GC–MS data, the most important database used is the Mass Spectral Library from U.S. National Institute of Standards and Technology (NIST). However, this database does not contain high resolution information, but identification can also be supported on the retention index (RI), which is calculated easily by analysis of a mixture of linear hydrocarbons [44]. On the other hand, there are several databases for LC–MS but none of them is complete enough for an integral identification. Furthermore, the use of MS/MS information to complete the assignation of metabolites in biological samples in LC–MS is mandatory. The most common databases for LC–MS analysis of biological samples are the HMDB [5,6], "Metabolites and Tandem MS Database (METLIN) [64] and LipidMaps [65].

3.4.2. Data normalization

The main objective of the normalization process is to remove unwanted variations among samples allowing quantitative comparison of them [63,66]. The selection of a proper normalization method depends on the type of biological sample to be analyzed and the analytical platform [66]. Furthermore, there are multiple normalization processes that could be applied simultaneously. In the case

of MS data the application of a logarithm transformation is frequently used to remove data heteroscedasticity [21,67]. Additionally, the instrumental variability in batches can be minimized by the normalization to the MS "total useful signal" (MSTUS), which uses the total intensity of peaks that are present in all samples under study as correction factor [66]. Both MSTUS and logarithmic transformation are the most common normalization strategies used in the research that constitutes this PhD Book. On the other hand, a common practice for culture cells is to normalize the data using the total protein content in cells [68], because it is commonly assumed that the metabolites concentration increases with the cells number. This strategy was used in Chapter IX by application of the Bradford method [69].

The evaluation of normality influences the selection of statistical tests, which can be parametric or non-parametric [70]. The easiest way to check normality assumption is by using graphical methods, but it should be supported by numerical methods [71]. In the development of the present Thesis, both histograms representations and normality tests were employed.

There are three normality tests commonly used from small to medium sized set of samples: Shapiro-Wilk (SW), Kolmogorov-Smirnov (KS) and Lilliefors (LF) tests [72]. These tests can only be applied under a certain condition, depending on the number of samples (n). If $n<50$ the SW test is the preferred option, while if $n>50$ both KS or LF tests can be applied. LF test is a modification of the KS test since the latter is only appropriate in a situation where the parameters of the hypothesized distribution are completely known. If these parameters are not known, the LF test should be preferably selected.

3.4.3. Statistical analysis

Different strategies can be used to assess the influence of a known factor (disease, diet, time, gender, etc.) in a batch of samples and identify the most altered metabolites. In this step many similarities and differences between targeted and untargeted metabolomic analysis are found [20,26].

a. Statistical analysis for targeted metabolomics

Statistical analysis in targeted metabolomics is typically initiated with the preparation of regression models to obtain the calibration curve for each metabolite. Once the target metabolites have been quantified, basic statistics can be carried out to look for significant differences between their levels found in classes. Univariate statistical tests such as the Student's *t*-test or Wilcoxon test are used to identify the relevant variables after exploratory analysis.

b. Statistical analysis for untargeted metabolomics

The most used statistical analysis for untargeted metabolomics are:

(i) Exploratory analysis by unsupervised analysis

Unsupervised learning refers to those methods for analysis of data without measured/defined outcome (response) or when the outcome measure is not of primary concern [73]. Unsupervised learning uses procedures that attempt to find the natural patterns to facilitate the understanding of the relationship between the samples and to highlight the variables that are responsible for these relationships. By providing means for visualization, unsupervised learning aids in the discovery of unknown but meaningful categories of samples or variables that naturally fall together [26]. The main unsupervised analysis are clustering analysis (CA), principal component analysis (PCA), multidimensional scaling (MDS) and self-organizing maps (SOM) [74]; being PCA the most used in metabolomics studies [26].

PCA is an orthogonal transformation of multivariate data, first formulated by Pearson (1901), and mostly used for exploratory analysis by extracting and displaying systematic variations [73]. PCA attempts to describe the maximal variance of the data, being a very useful tool for displaying purposes as it provides a low-dimension projection of the data (*i.e.,* a window into the original K-dimensional space) by transformation into a new coordinate system [26].

(ii) Classification by supervised multivariate analysis

Supervised learning uses labeled data to classify samples or objects according to an observed response that generally includes several variables or metabolites [26]. This classification aims at producing general hypotheses based on a training set of samples identified by known labels corresponding to the existing classes.

The list of supervised multivariate analysis included the *k*-nearest neighbor (KNN), supporting vector machine (SVM), probabilistic neural networks (PNN), linear discriminant analysis (LDA), random forests and partial least squares discriminant analysis (PLS-DA). Random forest and PLS-DA are the most used supervised multivariate analysis.

Random forests were introduced by Breiman in 2001, and it is basically defined as a combination of tree predictors [75]. Each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [75,76]. The random forest classifier uses bagging or bootstrap aggregating to form an ensemble of classification and regression tree (CART)-like classifiers [75]. Each node is split using the best among a subset of predictors randomly chosen at that node. This strategy performance could be better than many other classifiers, including LDA and SVM, and is robust against overfitting [75,77]. However, it does not provide a visual result of samples classification since it is mainly used to identify preset number of compounds with the highest frequency of occurrence in the tree branches generated by the application of this algorithm. This is the reason why this analysis is mainly used to rank the top potential biomarkers in metabolomics studies [78–83].

PLS-DA was introduced by Barker and Rayens in 2003 [84]. Currently, it is included into most of the packages used by chemometricians, especially in metabolomics [85]. PLS-DA applies partial least squares (PLS) regression using as the response variable a dummy index that associates the same integer number to samples belonging to the same class [84]. This regression is particularly suitable for situations where the number of observations or samples (N) is lower than that

of measured variables (*e.g.,* detected MFs, K). Its popularity has increased thanks to its ability to deal with noisy and intercorrelated variables. PLS also builds a low dimensional sub-space as the PCA, but the sub-space is based on combinations of the measured variables. PLS adjusts the model to capture the variation due to the classification of the samples, which can be expressed by a variation in *X*. Thus, the PLS-based classification generates data structures with intrinsic prediction power by maximizing the covariance between the measured variables and the classification assigned to the samples. The decomposition relies on latent variables that are computed sequentially to provide a good correlation with the remaining unexplained fraction of *Y*. In the context of classification, PLS discriminant analysis (PLS-DA) is performed to sharpen the partition between groups of observations to obtain the maximum separation between them. The model can then be analyzed to understand which variables carry the class-separating information. PLS-DA has demonstrated to be a potent tool for classification of metabolomics data.

(iii) Univariate analysis: parametric and non-parametric tests

There are multitude of test for performing univariate analysis. Nevertheless, the assumption or not of a normal distribution in data makes mandatory to distinguish between parametric and non-parametric tests. The former tests make certain assumptions about a data set; namely, the data are drawn from a population with a specific (normal) distribution. For this reason, these tests are only adequate to normal-distributed data sets, while non-parametric tests, which make fewer assumptions about the data set, are used in case of non-normal distribution of data [86,87]. In both cases there are parallel tests that can be used depending on the purpose of the study. Table 1 lists parametric and non-parametric tests –they can be used to identify statistical differences between samples from different groups or time points [86]. The predictive power of each variable is assessed by finding statistically significant differences between the mean intensity values of a given signal, of which the calculated *p*-value is a straightforward indicator. Such procedures are easily understandable, but their

use is rather limited in dealing with thousands of highly correlated variables [26]. False positives (type I error) are likely to occur when performing multiple comparisons. To reduce them, different strategies for *post hoc* analysis have been proposed [86]. Adjustments such as the Bonferroni or Benjamini correction have been introduced and are widely used to reduce false positives in statistical analysis [88].

Table 1. Summary of analog parametric and non-parametric tests.

	Parametric test	**Non-parametric test**
One sample	One sample *t*-test	Wilcoxon test
Two samples	Paired *t*-test	Wilcoxon test
	Unpaired *t*-test	Mann-Whitney test
Many samples	Analysis of variance (ANOVA)	Kruskal-Wallis test
	2-way ANOVA	Friedman test

c. Statistical analysis to discover potential biomarkers

There are some specific statistical strategies for discovery of potential biomarkers. In general, the biomarker selection is performed before building a definitive predictive model. However, whereas long lists of metabolites or large multivariate models are quite useful for understanding pathways and biological processes, they are not useful for developing cost-effective biomarker panels. Rather, a short list of 1–10 biomarkers is mathematically much more robust and far more practical for clinical testing purposes. For this reason, iterative methodologies for biomarker panel development, like the employed in the PanelomiX bioinformatics tool [89] or in the ROCCET on-line tool [90], are recommended.

Apart from the tools for biomarker discovery, there are other statistical tools that can be used to determine the predictive power of a single or a group of metabolites; some of them are described below.

(i) Receiver operating characteristic (ROC) curve analysis

The ROC curve analysis is the standard method to describe and assess the performance of diagnostic tests with binary classification [91,92]. ROC curves consider the frequency for production of true positives, true negatives, false positives and false negatives. These values are summarized into the proportion of actual positive cases that are correctly classified as positive (sensitivity), and the proportion of actual negative cases that are correctly classified as negative (specificity) [26]. An example of ROC curve is shown in Figure 9.

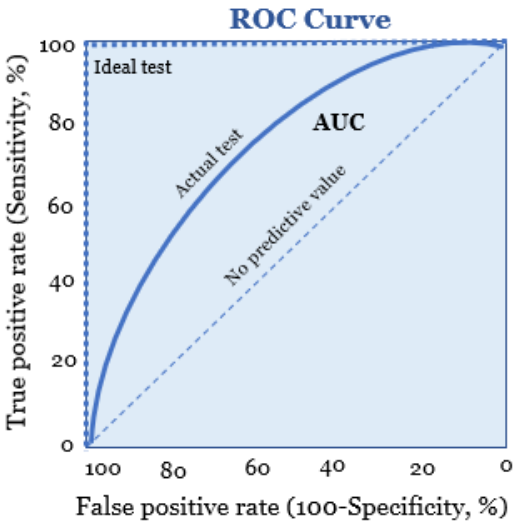


Figure 9. ROC representation including the curve described for an ideal test, a test without predictive value and an actual test.

Sensitivity can be considered as the probability of true positives (the positive result of the test from a subject that has an actual positive outcome), and specificity can be considered as the probability of true negatives (the negative result of the test from a subject that has an actual negative outcome) [26,91]. As the sensitivity and specificity of a test can vary depending on the biomarker decision boundary,

the best way to observe how this decision threshold affects sensitivity and specificity is through a ROC curve.

The area under the curve (AUC) is the main parameter considered from a ROC curve analysis (Figure 9). This parameter can be interpreted as the probability of the diagnostic test or classifier for randomly ranking positive or negative values [91,92]. If all positive samples are ranked before negative samples (example of a perfect classifier), the AUC is 1.0. An AUC of 0.5 is equivalent to randomly classify subjects as either positive or negative (non-useful classifier). However, using the AUC may not be always appropriate and the partial AUC would be most useful when only certain regions of the ROC space (high sensitivity or high specificity) are of particular interest [54].

(ii) Cox regression analysis

In long-term studies dealing with the diagnostic of a disease, it is important to evaluate the association between the concentration of a potential biomarker and the incidence of the pathology.

The Cox (Cox's proportional hazards) model [93] is the most commonly used multivariate approach for analyzing survival or incidence time data in clinical research (intervention and control groups). It is a regression model that describes the relationship between the event incidence, as expressed by the hazard function and a set of covariates [94]. Thus, this model is analogous to a multiple regression but the dependent variable is the hazard function that depends on all explanatory variables under consideration [95]. This hazard function described how hazard ratio (HR) varies in relation to changes in the explanatory variables and it is assumed that the HR does not depend on time [94,96]. The null value for the HR is 1, which indicates equality in the probability of occurrence of the event in both groups in the monitored time interval. An HR > 1 indicates more risk of occurrence of the event in the intervention group, and an HR < 1 a lower risk in the intervention group [95].

4. Current limitations and challenges of metabolomic analysis

One of the main limitations of metabolomics is the scarce number of completed databases as compared to those existing in other "omics" disciplines [97]. Probably, this limitation is not going to be solved in a short time period, since the number of metabolites is high enough to make this a difficult task owing to a long list of unknown metabolites in most biological samples.

There are other more affordable limitations in metabolomics. For example, there are still no universal protocols for metabolomic analysis of specific types of samples, mainly owing to the lack of standardization. On the other hand, the identification of metabolites is influenced by other difficulties apart from the absence of updated databases, since there are not protocols that guarantee a good identification level without an exhaustive training of the personnel involved in this task. One other pending aspect is the lack of validation protocols that ensure the analytical quality of the results.

All metabolomic limitations are a consequence of the great complexity of biological samples, which is basically explained by the huge variety of metabolites to be detected, the wide chemical diversity and ranges of concentrations covering several orders of magnitude. Thus, it is technically unfeasible to develop an analytical method that in a single analysis provides the detection of glucose, a polar metabolite present at the millimolar level, and calcitriol, a non-polar metabolite present at the picomolar level. In this sense, the presence of major compounds complicates the detection of metabolites present at low concentration. It is evident that there is a demand for more sensitive, more selective and more precise methodologies that provide greater detection capacity (also known as metabolomic coverage).

Sample preparation has not received enough attention in metabolomics compared to detection, particularly by NMR and/or MS. However, metabolomics analysis cannot be efficiently completed without a well-planned sample-

preparation protocol [23], and the quality of the results is largely related to the efficiency of sample preparation. Regarding the detection stage, the combination of analytical techniques ensures an increase in the detection and identification capacity in targeted analysis. However, a less implemented strategy is the development of data acquisition modes that allow improving the detection capacity.

Finally, data analysis is other of the blocks to improve the capacity of detection and identification, also promoting adequate strategies for data processing. Data analysis in "omics" studies represents the new bottleneck in the workflow owing to the large amount of generated information. In this sense, the combination of bioinformatic tools with detection approaches supposes another strategy that can improve the detection capacity of metabolomics analysis methods. The combination of this type of innovations in the different stages of the analytical process applied in metabolomics can lead to a remarkable improvement of the metabolomic coverage. The highest present challenge in metabolomics is to maximize the detection capability of the analytical methods to achieve unequivocal identification of thousands of metabolites, thus making possible to generate representative results of complex metabolomes.

The existing ways to increase the detection capability in metabolomics analysis make mandatory integration of the results obtained using complementary analytical techniques such as NMR and MS, either with direct infusion or coupled to a separation technique such as GC–MS or LC–MS platforms, and in a lesser extension, to CE–MS platforms. However, the integrated mode is only accessible to high budget laboratories because of the high acquisition and maintenance costs of these platforms.

With these premises, the major current challenges in metabolomics analysis are:

a. To improve the sample preparation protocols.
b. To maximize the detection capacity of analytical methods to achieve the unambiguous identification of thousands of metabolites.

c. To develop strategies for identification of unknown metabolites.

d. To improve the identification capability by creation of new databases with NMR or MS/MS information.

The research developed in this PhD contributes to the development of strategies, tools and methods to address these challenges in metabolomics analysis, which will provide a greater capacity to obtain information from different samples and with higher analytical quality. Figure 10 shows the general workflow of the analytical process in metabolomics and indicates the Chapters that contain analytical contributions to address these challenges.

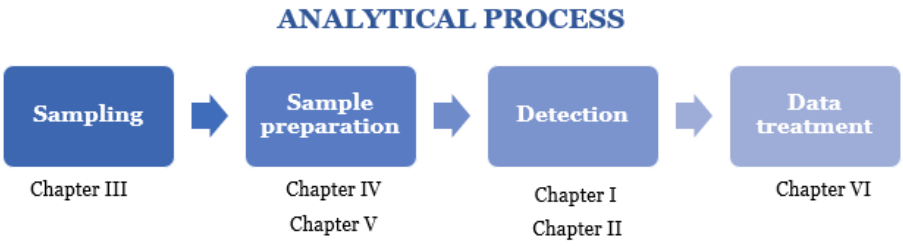


Figure 10. General workflow of the analytical process and the chapters in which improvements in given steps have been done.

References

[1] S.G. Oliver, M.K. Winson, D.B. Kell, F. Baganz, Systematic functional analysis of the yeast genome, Focus 16 (1998) 373–378.

[2] D.B. Kell, S.G. Oliver, S.G. Oliver, The metabolome 18 years on: a concept comes of age, Metabolomics 12 (2016) 1–8.

[3] O. Fiehn, Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks, Comp. Funct. Genomics 2 (2001) 155–168.

[4] O. Fiehn, Metabolomics – the link between genotypes and phenotypes, Plant Mol. Biol. 48 (2002) 155–171.

[5] D.S. Wishart, C. Knox, A.C. Guo, R. Eisner, N. Young, B. Gautam, D.D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J.A. Cruz, E. Lim, C.A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H.J. Vogel, I. Forsythe, HMDB: a knowledgebase for the human metabolome, Nucleic Acids Res. 37 (2009) 603–610.

[6] Human Metabolome Database, www.hmdb.ca (accessed on September 2019)

[7] A. Sussulini, Metabolomics: from Fundamentals to Clinical Applications, Elsevier, 2017.

[8] A. Mena-Bravo, Expansion, improvement and application of metabolomic platforms for the determination of vitamin D and its metabolites, Doctoral Thesis, University of Córdoba, Spain, 2019.

[9] J.K. Nicholson, I.D. Wilson, Understanding 'global' systems biology: metabonomics and the continuum of metabolism, Nat. Rev. Drug Discov. 2 (2003) 411–420.

[10] J.C. Lindon, E. Holmes, J.K. Nicholson, *Global Systems Biology through Integration of "Omics" Results*, in Handb. Metabonomics Metabolomics, Elsevier, 2007.

[11] A. Zhang, H. Sun, H. Xu, S. Qiu, X. Wang, Cell metabolomics, Omi. A J. Integr. Biol. 17 (2013) 495–501.

[12] G.J. Patti, O. Yanes, G. Suizdak, Metabolomics: the apogee of the omics trilogy, Nat. Rev. Mol. Cell Biol. 13 (2012) 263–269.

[13] K. Kishimoto, R. Urade, T. Ogawa, T. Moriyama, Nondestructive quantification of neutral lipids by thin-layer chromatography and laser-fluorescent scanning: suitable methods for "Lipidome" analysis, Biochem. Biophys. Res. Commun. 281 (2001) 657–662.

[14] K. Yang, X. Han, Lipidomics: techniques, applications, and outcomes related to biomedical sciences, Trends Biochem. Sci. 41 (2016) 954–969.

[15] N. Zehethofer, D.M. Pinto, Recent developments in tandem mass spectrometry for lipidomic analysis, Anal. Chim. Acta 627 (2008) 62–70.

[16] M.R. Wenk, The emerging field of lipidomics, Nat. Rev. Drug. Discov. 4 (2005) 594–610.

[17] O. Quehenberger, E.A. Dennis, The human plasma lipidome, N. Engl. J. Med. 365 (2011) 1812–1823.

[18] K. Sandra, P. Sandra, Lipidomics from an analytical perspective, Curr. Opin. Chem. Biol. 17 (2013) 847–853.

[19] O. Fiehn, Cellular Metabolomics: the Quest for Pathway Structure, in Handbook Metabonomics Metabolomics, Elsevier, 2007: pp. 35–54.

[20] E. Gorrochategui, J. Jaumot, S. Lacorte, R. Tauler, Data analysis strategies for targeted and untargeted LC-MS metabolomic studies: Overview and workflow, Trends Anal. Chem. 82 (2016) 425–442.

[21] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass spectrometry-based metabolomics, Mass Spectrom. Rev. 26 (2007) 51–78.

[22] A.K. Kosmides, K. Kamisoglu, S.E. Calvano, S.A. Corbett, I.P. Androulakis, Metabolomic fingerprinting: challenges and opportunities., Crit. Rev. Biomed. Eng. 41 (2013) 205–21.

[23] B. Alvarez-Sanchez, F. Priego-Capote, M.D. Luque de Castro, Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation, Trends Anal. Chem. 29 (2010) 111–119.

[24] H. Kanani, P.K. Chrysanthopoulos, M.I. Klapa, Standardizing GC–MS metabolomics, J. Chromatogr. B 871 (2008) 191–201.

[25] M. Biosyst, B. Zhou, LC-MS-based metabolomics, Mol. Biosyst. 8 (2012) 470–481.

[26] M. Calderón-Santiago, Mass spectrometry for the identification and quantitation of metabolomic biomarkers in clinical analysis, Doctoral Thesis, University of Córdoba, Spain, 2014.

[27] M.J.N. de Paiva, H.C. Menezesa, Z. de L. Cardeala, Sampling and analysis of metabolomes in biological fluids epidemiology is important for identifying associations between, Analyst 139 (2014) 3683–3694.

[28] M.A. López-Bascón, F. Priego-Capote, A. Peralbo-Molina, M. Calderón-Santiago, M.D. Luque de Castro, Influence of the collection tube on metabolomic changes in serum and plasma, Talanta 150 (2015) 681–689.

[29] V.V. Hernandes, C. Barbas, D. Dudzik, A review of blood sample handling and pre-processing for metabolomics studies, Electrophoresis 38 (2017) 2232–2241.

[30] K. Ota, S. Tanaka, Animal cell culture medium and culture container, Patent 20170022477 A1, 2017.

[31] K. Yang, Adipose Tissue Protocols, Humana Press, 2008.

[32] A.E. Pricope-Veselin, V. Mocanu, D. Timofte, Open surgical and needle biopsy to study abdominal subcutaneous adipose tissue in obesity, J. Surg. 14 (2018) 101–103.

[33] I. Schlecht, W. Gronwald, G. Behrens, S.E. Baumeister, J. Hertel, J. Hochrein, H.U. Zacharias, B. Fischer, P.J. Oefner, M.F. Leitzmann, Visceral adipose tissue but not subcutaneous adipose tissue is associated with urine and serum metabolites, PLoS One 12 (2017) e0175133.

[34] M. Orešič, Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction, Nutr. Metab. Cardiovasc. Dis. 19 (2009) 816–824.

[35] L.D. Roberts, J.A. West, A. Vidal-Puig, J.L. Griffin, Methods for Performing Lipidomics in White Adipose Tissue, in: Methods Enzymology, Elsevier, 2014: pp. 211–231.

[36] R.C. Baker, Y. Nikitina, A.R. Subauste, Analysis of Adipose Tissue Lipid using Mass Spectrometry, in: Methods Enzymology, Elsevier, 2014: pp. 89–105.

[37] K.L. Campbell, K.W. Makar, M. Kratz, K.E. Foster-Schubert, A. McTiernan, C.M. Ulrich, A pilot study of sampling subcutaneous adipose tissue to examine biomarkers of cancer risk, Cancer Prev Res. 23 (2008) 1–7.

[38] S. Matysik, C.I. Le Roy, G. Liebisch, S.P. Claus, Metabolomics of fecal samples: a practical consideration, Trends Food Sci. Technol. 57 (2016) 244–255.

[39] C. Haro, M. Montes-Borrego, O.A. Rangel-Zúñiga, J.F. Alcalã-Diaz, F. Gamez-Delgado, P. Pérez-Martinez, J. Delgado-Lista, G.M. Quintana-Navarro, F.J. Tinahones, B.B. Landa, J. Lapez-Miranda, A. Camargo, F. Pérez-Jiménez, Two healthy diets modulate gut microbial community improving insulin sensitivity in a human obese population, J. Clin. Endocrinol. Metab. 101 (2016) 233–242.

[40] C. Haro, S. Garcia-Carpintero, J.F. Alcala-Diaz, F. Gomez-Delgado, J. Delgado-Lista, P. Perez-Martinez, O.A. Rangel Zuñiga, G.M. Quintana-Navarro, B.B. Landa, J.C. Clemente, J. Lopez-Miranda, A. Camargo, F. Perez-Jimenez, The gut microbial community in metabolic syndrome patients is modified by diet, J. Nutr. Biochem. 27 (2016) 27–31.

[41] O. Deda, H.G. Gika, I.D. Wilson, G.A. Theodoridis, An overview of fecal sample preparation for global metabolic profiling, J. Pharm. Biomed. Anal. 113 (2015) 137–150.

[42] Z.G. Gong, J. Hu, X. Wu, Y.J. Xu, The recent developments in sample preparation for mass spectrometry-based metabolomics, Crit. Rev. Anal. Chem. 47 (2017) 325–331.

[43] D. Vuckovic, Current trends and challenges in sample preparation for global metabolomics using liquid chromatography – mass spectrometry, Anal. Bioanal. Chem. 403 (2012) 1523–1548.

[44] O. Fiehn, Metabolomics by gas chromatography-mass spectrometry: the combination of targeted and untargeted profiling, Curr. Protoc. Mol. Biol. 114 (2017) 1–43.

[45] G. Theodoridis, H.G. Gika, I.D. Wilson, LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics, Trends Anal. Chem. 27 (2008) 251–260.

[46] G. Theodoridis, H.G. Gika, I.D. Wilson, A. Park, Mass spectrometry-based holistic analytical approaches for metabolite profiling in systems biology studies, Mass Spectrom. Rev. 30 (2011) 884–906.

[47] F.A. Mellon, Mass spectrometry: Principles and Instrumentation Applications, in: Encyclopedia of Food Sciences and Nutrition, 1995: pp. 3739–3749.

[48] Z. Lei, D. V Huhman, L.W. Sumner, Mass spectrometry strategies in Metabolomics, J. Biol. Chem. 286 (2011) 25435–25442.

[49] K.K. Murray, R.K. Boyd, M.N. Eberlin, G.J. Langley, L. Li, Y. Naito, Definitions of terms relating to mass spectrometry (IUPAC Recommendations 2013), Pure Appl. Chem. 85 (2013) 1515–1609.

[50] E. De Hoffmann, V. Stroobant, Mass Spectrometry Principles and Applications, John Wiley & Sons, 2007.

[51] C. Morris, C.M. O'Grada, M.F. Ryan, M.J. Gibney, H.M. Roche, E.R. Gibney, L. Brennan, Modulation of the lipidomic profile due to a lipid challenge and fitness level: A postprandial study, Lipids Health Dis. 14 (2015) 1–11.

[52] Y. Wang, S. Liu, Y. Hu, P. Li, J.B. Wan, Current state-of-the-art of mass spectrometry-based metabolomics studies - A review focusing on wide coverage, high throughput and easy identification, R. Soc. Chem. Adv. 95 (2015) 78728–78737.

[53] A. Technology, Mass spectrometry fundamentals–theory, 2016 (accessed on 12th May 2019).

[54] F.G. Kitson, B.S. Larsen, C.N. McEwen, *What is GC/MS? A Practical Guide*, in: Gas Chromatography and Mass Spectrometry, Academic Press, 1996: pp. 3–23.

[55] Agilent Technologies, Triple quadrupole manual (Concepts guide) (accessed on 17th May 2019).

[56] X. Han, K. Yang, R.W. Gross, Multi-dimensional mass spectrometry-based shotgun lipidomics and novel strategies for lipidomic analyses, Mass Spectrom. Rev. 31 (2012) 134–178.

[57] Agilent Technologies, QTOF manual (Components) (accessed on 22th May 2019).

[58] Z. Muccio, G.P. Jackson, Isotope ratio mass spectrometry, Analyst 134 (2009) 213–222.

[59] Thermo Scientific, Thermo Scientific FLASH HT Plus Fully Automated Multi-

Element Isotope Analysis for C, N, S, O and H (accessed on 15th May 2019).

[60] B. Lavine, J. Workman, Chemometrics, Anal. Chem. 82 (2010) 4699–4711.

[61] J. Trygg, T. Lundstedt, Chemometrics Techniques for Metabonomics, Elsevier, 2007.

[62] R. Madsen, T. Lundstedt, J. Trygg, Chemometrics in metabolomics — a review in human disease diagnosis, Anal. Chim. Acta 659 (2010) 23–33.

[63] M. Katajamaa, M. Oresic, Data processing for mass spectrometry-based metabolomics, J. Chromatogr. B 1158 (2007) 318–328.

[64] X. Domingo-Almenara, J.R. Montenegro-Burke, C. Guijas, E.L.-W. Majumder, H.P. Benton, G. Siuzdak, Autonomous METLIN-guided in-source fragment detection increases annotation confidence in untargeted metabolomics, Anal. Chem. (2019) 1–9.

[65] E. Fahy, M. Sud, D. Cotter, S. Subramaniam, LIPID MAPS online tools for lipid research, Nucleic Acids Res. 35 (2007) 606–612.

[66] Y. Wu, L. Li, Sample normalization methods in quantitative metabolomics, J. Chromatogr. A 1430 (2015) 80–95.

[67] R.A. Van Den Berg, H.C.J. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. Van Der Werf, Centering, scaling, and transformations: improving the biological information content of metabolomics data, BMC Genomics 7 (2006) 1–15.

[68] Y. Kam, G.W. Rogers, N. Jastromb, B.P. Dranka, Methods and strategies for normalizing XF metabolic data to cellular parameters, Agil. Technol. (2018). www.agilent.com/cs/library/applications/5991-7908EN.pdf (accessed on 15th February 2019).

[69] H. van den Berg, F.A. Hommes, A rapid and sensitive method for the determination of short chain fatty acids in serum, Clin. Chim. Acta 51 (1974) 225–232.

[70] D.J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, CRC Press, 2003.

[71] N.M. Razali, Y.B. Wah, M. Sciences, Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests, J. Stat. Model. Anal. 2 (2011) 21–33.

[72] H. Kim, Statistical notes for clinical researchers: assessing normal distribution using skewness and kurtosis, Open Lect. Stat. 7658 (2013) 52–54.

[73] J. Bible, S. Datta, S. Datta, Cluster Analysis: Finding Groups in Data, in: Informatics for Materials Science and Engineering, Elsevier, 2013: pp. 53–70.

[74] T.M.D. Ebbels, Non-linear Methods for the Analysis of Metabolic Profiles, Elsevier, 2007.

[75] L.E.O. Breiman, Random Forests, Mach. Learn. 45 (2001) 5–32.

[76] X. Chen, H. Ishwaran, Genomics random forests for genomic data analysis, Genomics 99 (2012) 323–329.

[77] A. Liaw, M. Wiener, Classification and regression by random forest, R News 2 (2002) 18–22.

[78] M. Calderón-Santiago, M.A. López-Bascón, Á. Peralbo-Molina, F. Priego-Capote, MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls, Talanta 174 (2017) 29–37.

[79] K.M. Sas, A. Karnovsky, G. Michailidis, S. Pennathur, Metabolomics and diabetes: analytical and computational approaches, Diabetes 64 (2015) 718–732.

[80] R. Wang-Sattler, Z. Yu, C. Herder, A.C. Messias, A. Floegel, Y. He, K. Heim, C. Meisinger, C. Gieger, C. Prehn, W. Roemisch-margl, Novel biomarkers for pre-diabetes identified by metabolomics, Mol. Syst. Biol. (2012) 1–11.

[81] A. Peralbo-Molina, M. Calderon-Santiago, F. Priego-Capote, B. Jurado-Gamez, M.D. Luque de Castro, Identification of metabolomics panels for potential lung cancer screening by analysis of exhaled breath condensate, J. Breath Res. 10 (2016) 1–12.

[82] P. Moreno, C. Jimenez-Jimenez, M. Garrido-Rodriguez, M. Calderon-Santiago, S. Molina, M. Lara-Chica, F. Priego-Capote, E. Muñoz, M.A. Calzado, Metabolomic profiling of human lung tumor tissues – nucleotide metabolism as a candidate for therapeutic interventions and biomarkers, Mol. Oncol. 12 (2018) 1778–1796.

[83] M. del M. Delgado-Povedano, M. Calderon-Santiago, B. Jurado-Gamez, F. Priego-capote, M.D. Luque De Castro, Recent advances in human sweat metabolomics for lung cancer screening, Metabolomics 12 (2016) 1–12.

[84] P. Oliveri, C. Malegori, M. Casale, Chemometrics and Statistics. Multivariate Classification Techniques, in: Encyclopedia of Analytical Science, Elsevier, 2019: pp. 481–486.

[85] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away, J. Chemom. 28 (2014) 213–225.

[86] K. Kapur, Principles of Biostatistics, in: Clinical and Translational Science, Elsevier, 2016: pp. 243–260.

[87] F.S. Nahm, Nonparametric statistical tests for the continuous data: the basic concept and the practical use, J. Anesthesiol. 69 (2016) 8–14.

[88] J. Popper Shaffer, Multiple hypothesis testing, Annu. Rev. Psychol. 46 (1995) 561–584.

[89] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J. Sanchez, M. Müller, PanelomiX : A threshold-based algorithm to create panels of biomarkers, Biochem. Pharmacol. 1 (2013) 57–64.

[90] J. Xia, D.I. Broadhurst, M. Wilson, D.S. Wishart, Translational biomarker discovery

in clinical metabolomics: an introductory tutorial, Metabolomics 9 (2013) 280–299.

[91] K. Søreide, Receiver-operating characteristic curve analysis in diagnostic, prognostic and predictive biomarker research, J. Clin. Pathol. 62 (2009) 1–5.

[92] S.D. Walter, The partial area under the summary ROC curve, Stat. Med. 24 (2005) 2025–2040.

[93] D.R. Cox, Regression models and life-tables, J. R. Stat. Soc. 34 (1972) 187–220.

[94] M.J. Bradburn, T.G. Clark, S.B. Love, D.G. Altman, Tutorial paper survival analysis part II: Multivariate data analysis – an introduction to concepts and methods, Br. J. Cancer 89 (2003) 431–436.

[95] M.Á. Martínez-gonzález, Á. Alonso, J. López, ¿Qué es una hazard ratio? Nociones de análisis de supervivencia, Med. Clin. 34227 (2008) 1–9.

[96] V. Bewick, L. Cheek, J. Ball, Statistics review 12: Survival analysis, Crit. Care 8 (2004) 389–394.

[97] A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B.S. Kristal, B. Van Ommen, V. Elwin, E. Pujos-Guillot, D. Wishart, S. Wopereis, Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research, Metabolomics 5 (2009) 435–458.

ANALYTICAL TOOLS AND EQUIPMENT

This section of the Thesis Book describes the different analytical tools and equipment used in the experimental part of the Thesis, which are described in more detail in the subsequent chapters.

1. Samples

The global objective of the Thesis research was to develop analytical strategies for improving the detection and identification by mass spectrometry in targeted and untargeted (global) metabolomic analysis applied to clinical and agro-food areas. Samples used in these two fields were:

Clinical samples. Human serum and plasma from volunteers were used in Chapters I, II and VI. Serum samples used in Chapter III were obtained from patients recruited at the Cardiology Unit of Miguel Servet Hospital (Zaragoza, Spain). Human plasma from CORDIOPREV study (ClinicalTrials.gov, NTC00924937) recruited at the Reina Sofia University Hospital (Córdoba, Spain) were used in Chapters VII and VIII. PBMCs used in Chapter IX were isolated from prospective samples collected in the Finnish Type 1 Diabetes Prevention and Prediction Study (DIPP). Visceral adipose tissue was used in Chapter IV. These samples were obtained from a pool of individuals with morbid obesity undergoing bariatric surgery and recruited by the Lipids and Atherosclerosis Unit of the Reina Sofia University Hospital (Cordoba, Spain). All clinical experiments were carried out in accordance with the ethical principles of human medical research (World Medical Association, Helsinki Declaration, 2004).

Agro-food samples. A pool of pig feces was prepared by homogenous mixing of samples from five pigs, then used in the research in Chapter V. All procedures of sample extraction were performed in accordance with the European regulations regarding the protection of animals used for experimental and other scientific purposes, under the supervision of the Ethical and Animal Welfare Committee of

the University of León (Spain). Subcutaneous pig fat, used in Chapter X, was collected following the established Spanish official method (B.O.E. 2004). In this case, eighty Iberian pigs were selected from different farms located at North of the Sevilla province (Spain).

2. Systems for sample preparation

The systems used for sample preparation can be classified into continuous and discontinuous systems.

2.1. Continuous systems for sample preparation

A continuous system was used for sample preparation prior to analysis of FAHFAs in serum, as described in Chapter I. It was an automated SPE workstation Prospekt-2 system from Spark Holland (Emmen, The Netherlands), which included an automatic cartridge exchanger (ACE) and a high-pressure syringe dispenser (HPD) for solvent delivery. The automated system was coupled to a Midas autosampler furnished with a 1 mL sample-loop. Full automation of the extraction step is controlled by Sparklink software version 2.10 from Spark Holland. This system allows direct connection with the chromatographic system.

2.2. Discontinuous systems for sample preparation

The majority of the research in the Thesis was based on untargeted analysis (metabolomics profiling). For this reason, a minimally invasive sample preparation was always adopted as strategy. In general terms, sample preparation consisted of extraction, centrifugation and preconcentration, the last step only if required (evaporation and reconstitution).

However, for the extraction and preconcentration of phospholipids conventional SPE cartridges, packaged with zirconia coated silica, were necessary, as detailed in Chapter IV. In this chapter, the influence of the SPE cartridges on sample preparation is discussed.

3. Separation and detection systems

The methods developed in the experimental part of this Doctoral Thesis have been based on a chromatographic separation (using LC or GC) and subsequent detection by mass spectrometry.

3.1. Targeted analysis

An Agilent (Agilent Technologies, Palo Alto, CA, USA) 1200 Series LC system coupled to an Agilent 6460 QqQ furnished with an Agilent JetStream Technology ESI was used in Chapter I for targeted analysis of FAHFAs. Chromatographic separation was performed with a C18 Mediterranea column (10 × 4.6 mm i.d.; packed with 3 μm particle diameter size). Agilent MassHunter Workstation LC–QqQ (version B.3.01) was the software for data acquisition.

A Thermo Scientific UltiMate 3000 series LC system coupled to a Thermo Scientific QqQ TSQ Quantum™ Access MAX detector (Waltham, MA, USA) was used in Chapter II for polar lipid analysis. The QqQ detector was equipped with a heated electrospray ionization (HESI) probe for spraying of chromatographic eluate. Chromeleon™ software (version 6.80) was used for controlling the LC system, TSQ Tune software (version 1.2.1) was used to control the detector parameters and, finally, Thermo Xcalibur™ software (version 3.0.63) was used for methods and worklists creation.

The levels of polar metabolites —the target of Chapter IX— were measured by a GC–Q/MS. For this purpose, an Agilent 7890A Series GC system coupled to an Agilent 5977B single quad mass spectrometer equipped with an EI source was used. The gas separation was carried out by a fused silica HP-5 column (30 m × 0.25 mm i.d., 0.25 μm film thickness capillary column) from Agilent Technologies. Agilent MassHunter Workstation GC–Q (version B.07.02) was the software for data acquisition.

Two different platforms were used in Chapter X: GC–FID and IRMS. For separation and detection of FAMEs an Agilent 7820A GC System equipped with an

FID was selected. The GC was furnished with an autosampler and a split/splitless injector. An SP™-2380 fused silica capillary column (60 m × 0.25 mm i.d., 0.2 μm film thickness capillary column) provided by Supelco (Bellefonte, PA, USA) was used as analytical column. EZ Chrom Elite Compact software (version 3.3.2 from Agilent Technologies) was used for acquisition and processing of data in the GC–FID system. A Delta V Advantage Isotope Ratio Mass Spectrometer from Thermo Fisher Scientific (Bremen, Germany) was employed for determination of $^{13}C/^{12}C$ (carbon isotope ratio, $\delta^{13}C$). This system was equipped with a ConFlo IV Universal Interface for continuous flow analysis and a Flash 2000 HT elemental analyzer. A molecular sieve packed column (5 Å, 1 m × 1/8" × 2 mm) from Thermo Scientific (Bremen, Germany) was used for sample clean-up. Isodat Gas Isotope Ratio MS Software (version 3.0 from Thermo Scientific, Bremen, Germany) was used to acquire and process the signal obtained by IRMS analysis.

3.2. Untargeted analysis

An Agilent 1200 Series LC system coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer was used in Chapters V, VI, VII and VIII. The QTOF detector was equipped with a dual electrospray ionization source for simultaneous spraying of chromatographic eluate and a reference solution to calibrate continuously the detected *m/z* ratios. The liquid chromatographic separation was carried out with different columns in each Chapter, depending on the objective. Agilent MassHunter Workstation LC–QTOF acquisition software (version B.06) was used to control the instrument and acquire the data.

An Agilent 7890A Series GC system coupled to an Agilent 7200 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with an EI source was used in Chapters III, V, VII and VIII. In all untargeted metabolomic studies, the gas chromatographic separation was carried out with a fused silica DB-5MS-UI (30 m × 0.25 mm i.d., 0.25 μm film thickness capillary column) from Agilent Technologies. The analytical sample was thus monitored in high resolution mode. Agilent MassHunter Workstation GC–QTOF Acquisition software (version B.06)

was used to control data acquisition and set the parameters for optimum operation.

The levels of molecular lipids, in Chapter IX, were measured by an LC–QTOF MS/MS. For this purpose, an Agilent 1290 Infinity LC system coupled to an Agilent 6545 QTOF detector furnished with a dual JetStream Technology (dual ESI) was used. Chromatographic separation was performed using an Acquity UPLC BEH C18 column (100 mm × 2.1 mm i.d., 1.7 μm particle size) from Waters Corporation (Wexford, Ireland). Agilent MassHunter Workstation LC–QqQ Acquisition software (version B.08.00) was the software for data acquisition and set the parameters for optimum operation.

4. Data processing

4.1. Data processing for targeted analysis

In Chapter I a quantitative method for analysis of FAHFAs was developed. MassHunter Workstation software was used to process all data obtained by LC–QqQ in MRM mode. This software allowed generating a data set containing the peak area and RT of each metabolite in each sample. Calibration models were built using the peak areas obtained by analysis of serum spiked with multistandard solution at different concentration levels. The limit of detection (LOD) for each analyte was calculated as $3.3 \times \sigma/S$, where σ the standard deviation of the response of a blank analysis and S is the slope of the calibration curve for each analyte. The limit of quantitation (LOQ) was calculated as $10 \times \sigma/S$.

In Chapter II a quantitative method for analysis of 398 polar lipids was developed. TraceFinder™ software (version 3.2.512.0, ThermoFisher Scientific) was used to process the data obtained by LC–QqQ in different work modes. A data set containing the peak area and RT of each lipid in each sample was obtained.

In Chapter IX, the GC–Q/MS data were processed in MassHunter Quant (version 8, from Agilent Technologies). This software allowed generating a data set

containing the peak area and RT of each metabolite in all samples. The peaks were manually checked and corrected if needed for correct integration. The calibration models were built using the peak areas obtained by analysis of multistandard solution at different concentration levels.

In Chapter X, EZ Chrom Elite Compact software (version 3.3.2 from Agilent Technologies) was used for acquisition and processing of data in the GC–FID system. This software allowed generating a data matrix containing the peak area and RT of each FAME. The concentrations of target FAs were expressed as a percentage in relative terms. Isodat Gas Isotope Ratio MS Software (version 3.0 from Thermo Scientific, Bremen, Germany) was used to acquire and process the signals obtained by IRMS analysis. The results of $\delta^{13}C$ analysis were reported as the per mil (‰) enrichment relative to the international standard, V-PDB for carbon isotope ratio, according to the equation: $X(‰) = [(R_{sample}/R_{reference}) - 1] \times 1000$, where X is the ratio of the heavy to the light stable isotope (e.g. $^{13}C/^{12}C$) in the sample (R_{Sample}) and in the standard ($R_{Reference}$). Finally, a matrix with data from both techniques was built.

4.2. Data processing for untargeted analysis

Untargeted analysis was performed in Chapters II, IV, V, VII and VIII. The strategy carried out was different depending on whether the data comes from LC–QTOF MS/MS or GC–TOF/MS.

4.2.1. Data from LC–QTOF MS/MS

MassHunter Workstation software (version B7.00, Qualitative Analysis, Agilent Technologies) was used to process all data obtained by LC–QTOF in data-dependent acquisition MS/MS mode. Treatment of raw data files started by extraction of potential MFs with the suited algorithm included in the software. For this purpose, the extraction algorithm considered all ions exceeding 500 counts for both polarities with a single charge state for the obtained chromatograms. These cut-off values were established taking into account the chromatographic background noise. Additionally, the isotopic distribution to consider MFs as valid

should be defined by two or more ions (with a peak spacing tolerance of *m/z* 0.0025, plus 7.0 ppm in mass accuracy). Ions and adducts formation in the positive (+H, +Na, +K, +NH_4) and negative ionization (−H, +HCOO, +Cl) modes, as well as neutral loss by dehydration were included to identify features corresponding to the same potential metabolite.

A library (.cdb) with all identified metabolites (see section 5) was used to perform a targeted compound extraction analysis using a tolerance window of 0.8 min and 5 ppm mass accuracy. This step was performed with Profinder Analysis (version B8.00, Agilent Technologies) in all chapters, except Chapter IX. A table with the peak area of all identified compounds in the different samples was obtained as a result.

In the case of Chapter IX, a database (.csv file) was created with all identified lipids to perform a targeted compound extraction. For this purpose, open source software MZmine 2.33 was used using a tolerance window of 0.1 min and 5 ppm mass accuracy.

4.2.2. Data from GC–TOF/MS

Unknown Analysis software (version 7.0, Agilent Technologies) was used to unzip all data files obtained by GC–TOF/MS in full scan mode. Then, MassHunter software was used to process GC–TOF/MS data files. Treatment of raw data files started by deconvolution of chromatograms to obtain a list of MFs considered as potential compounds defined by the *m/z* value of one representative ion for each chromatographic peak and its RT. For this purpose, the deconvolution algorithm was applied to each sample by considering all ions exceeding 1500 counts for the absolute height parameter, the accuracy error at 50 ppm and the window size factor at 150 units. The list of MFs obtained for each analysis was exported as data files in compound exchange format (.cef files).

5. Identification of metabolites detected by untargeted analysis

5.1. Identification of metabolites detected by LC–QTOF MS/MS

Tentative identification of metabolites was supported on MS and MS/MS information that was searched in the METLIN MS and MS/MS databases (http://metlin.scripps.edu), the HMDB (version 3.6) and the LIPID MAPS website ((http://www.lipidmaps.org), using in all cases the MFs obtained in data processing step. Once the tentative identification was finished, PCDL Manager software (version B0.07.00, Agilent Technologies) was used to create a library with the MS/MS spectra obtained for each metabolite.

5.2. Identification of metabolites detected by GC–TOF/MS

Tentative identification of metabolites was performed by searching each mass spectrum in the NIST database (version 11) using the RI value. The identification was firstly carried out by searching MS spectra on the NIST database. Only those identifications with a match factor and a reverse match factor higher than 700 were considered as valid. The RI values included in the NIST database were also taken into account to support identifications. An RI calibration model was built by plotting the RT obtained by analysis of the alkane standard mixture (C10 to C40 with an even number of carbons) with the chromatographic method used in this research and the RI values provided for each alkane by the NIST database. Then, the RI value was experimentally estimated for each identified compound by using the RT and the calibration equation. The requirement to accept NIST database identifications was that the difference between the experimental RI and the theoretical value provided by the NIST for each target compound should be below 100 units. The NIST database does not contain high resolution MS information as provided by the TOF detector. For this reason, a third step was included to validate identification of each compound by using high resolution MS. Thus, the molecular formula for the $[M]^+$ ion and the most intense fragments for each MF should fit the NIST 11 identification by setting a cut-off value in mass

accuracy of 5 ppm.

A library was created with the MS spectra obtained for each metabolite with the Library Editor (version B0.07.00, Agilent Technologies) software.

6. Data pretreatment and statistical analysis

Datasets including the peak area for all metabolites or potential metabolites detected in all samples were then processed for statistical analysis. According to the objective of the study different statistical tools were used. The main software and tools used in the development of the Thesis are described below:

a. Normalization of the data was carried out with MassProfiler Professional software package (MPP, version 2.2, Agilent Technologies) and Metaboanalyst (version 4.0, URL: https://www.metaboanalyst.ca/). According to the objective of the study, the main standardizations were based on logarithmic transformation, Z-transform, autoscaling, and MSTUS.
b. Statgraphics Centurion XVI (version 16.1.18), R program (interface R-studio, version 1.0.136) and PASW Statistics (version 18) allow performing different parametric and non-parametric statistical analysis, and the graphic representation of the data.
c. R program (interface R-studio, version 1.0.136) was used to apply and develop MetaboQC package by combination of different algorithms.
d. MPP software and Metaboanalyst allow unsupervised analysis by PCA and supervised analysis by PLS-DA and random forests.
e. Evaluation of the prediction capacity of metabolites and creation of prediction models was carried out using Random Forests Analysis and the ROC curves provided by the PanelomiX software (URL: https://www.panelomix.net/), ROCCET (URL: https://www.metaboanalyst.ca/rocanalysis) and Metaboanalyst

(URL: https://www.metaboanalyst.ca/), which allows the creation of marker panels and the evaluation of their predictive capacity.

f. The evaluation of the occurrence incidence of a disease or condition was evaluated using PASW Statistics (version 18) by Cox regression analysis.

g. Mapping Metabolites on Kyoto Encyclopedia of Genes and Genomes (KEGG) (URL: http://www.genome.jp/kegg/), Metaboanalyst software and Ingenuity Pathway Analysis (IPA®) were used to know the metabolic pathways in which the significant metabolites were involved in order to find connections among them.

EXPERIMENTAL PART

SECTION I

Methodological development in targeted metabolomics analysis: versatility of the triple quadrupole mass spectrometer

Section I of this PhD Book is devoted to methodological development in targeted metabolomic analysis with the aim to confirm the presence of metabolites in biological samples as a preliminary step to their quantitative analysis. Low resolution mass spectrometry, with a QqQ detector, is an ideal tool to undertake quantitative and confirmatory analysis in metabolomics and, therefore, to work in targeted analysis. Also, the QqQ offers additional features that make it a suitable detector to play an important role in qualitative analysis. For this purpose, the development of DDM can be used to increase the identification power in certain families of compounds. The QqQ operational modes, once the fragmentation mechanisms of a particular family of compounds are known, can be used to confirm their presence by monitoring in the first or third quadrupole representative precursor or product ions, respectively. It is possible to do several scans to identify all the compounds belonging to a family present in a biological sample and, once confirmed their presence, proceed to its quantification by means of the MRM mode. With these premises, Chapter 1 and 2 were developed.

Chapter 1 is devoted to confirm and quantify the presence of FAHFAs in human serum. The low nanomolar concentrations at which FAHFAs are present in serum make necessary the development of a method with enough sensitivity to quantify these compounds. For this purpose, an automated method based on on-line SPE coupled to LC–MS/MS was developed. In Chapter 2 a qualitative/ quantitative strategy for determination of polar lipids was developed for application to biological samples. LC–MS/MS allowed detecting polar lipids in the presence of other concentrated fractions in plasma such as acylglycerides.

CHAPTER I

Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry

Analytica Chimica Acta

943 (2016) 82–88

Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry

María Asunción López-Bascón[a,b,c], Mónica Calderón-Santiago[a,b,c], Feliciano Priego-Capote[a,b,c,*]

[a]*Department of Analytical Chemistry, University of Córdoba, Annex C-3, Campus of Rabanales, Córdoba, Spain*

[b]*University of Córdoba Agroalimentary Excellence Campus, ceiA3, Córdoba, Spain*

[c]*Maimónides Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba Córdoba, Spain*

Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry

María Asunción López-Bascón, Mónica Calderón-Santiago, Feliciano Priego-Capote

ABSTRACT

A novel class of endogenous mammalian lipids endowed with antidiabetic and anti-inflammatory properties has been recently discovered. They are fatty acid esters of hydroxy fatty acids (FAHFAs) formed by condensation between a hydroxy fatty acid and a fatty acid. FAHFAs are present in human serum and tissues at low nanomolar concentrations. Therefore, high sensitivity and selectivity profiling analysis of these compounds in clinical samples is demanded. An automated qualitative and quantitative method based on on-line coupling between solid phase extraction and liquid chromatography–tandem mass spectrometry has been developed for determination of FAHFAs in serum with the required sensitivity and selectivity. Matrix effects were evaluated by preparation of calibration models in serum and methanol. Recovery factors ranged between 73.8 and 100% in serum. The within-day variability ranged from 7.1 to 13.8%, and the between-days variability varied from 9.3 to 21.6%, which are quite acceptable values taking into account the low concentration levels at which the target analytes are found. The method has been applied to a cohort of human serum samples to estimate the concentrations profiles as a function of the glycaemic state and obesity. Statistical analysis revealed three FAHFAs with levels significantly different depending on the glycaemic state or the body mass index. This automated method could be implemented in high-throughput analysis with minimum user assistance.

Keywords: Fatty acid esters of hydroxy fatty acids, solid phase extraction, serum, liquid chromatography–mass spectrometry, targeted analysis

1. Introduction

In the last decades the prevalence of metabolic syndrome has increased in developed countries, mainly because of the rise of obesity [1]. Metabolic syndrome is a collection of risk factors associated to a higher risk for cardiac diseases, atherosclerosis, fatty liver disease and diabetes [1–3]. These diseases share risk factors, being the most important obesity and insulin resistance [1]. Nowadays, the most common disease in patients affected by metabolic syndrome is diabetes mellitus. In fact, type-2 diabetes mellitus (T2DM) is increasing worldwide at an epidemic rate, which is expected to reach 592 million inflicted individuals by 2035 as compared to 382 million reached in 2013 [1,4]. Obesity is one of the major risk factors for T2DM, since around 85% of subjects with T2DM are over-weighted or obese [5]. Thus, the body mass index (BMI) is not only associated with chronic low-grade inflammation and increased oxidative stress, but also with insulin resistance and metabolic dysregulation [2,5].

Recently, a novel class of endogenous mammalian lipids endowed with antidiabetic and anti-inflammatory properties has been found [6–11]. Yore *et al.* in 2014 referred to this class of natural-occurring lipids as fatty acid-hydroxy fatty acids, abbreviated as FAHFAs [6,7]. The authors identified 16 FAHFAs by different combinations of the main long-chain fatty acids (palmitate, oleate, stearate or palmitoleate) conjugated to a hydroxylated version of one of the same set of fatty acids [7]. In that research, the FAHFAs content of blood, serum and adipose tissue taken from diabetic mice were compared to that from normal mice. Yore *et al.* reported that 6 of the 16 FAHFA species were upregulated by GLUT4 overexpression, the unique insulin-sensitive glucose transporter protein [7]. The most dramatically upregulated FAHFA in the overexpressing GLUT4 was palmitic acid hydroxystearic acid (PAHSA) [7]. In humans, the authors observed a strong association between PAHSA and insulin sensitivity, being PAHSA levels significantly lower in insulin-resistant individuals as compared to insulin-sensitive cases [6,7]. Additionally, the characterization of the most common FAHFAs found in normal tissues and serum revealed that their levels were similar to other signalling lipids such as prostacyclins, prostaglandins, steroids and endocannabinoids [6]. Research on FAHFAs has revealed multiple effects that improve

glucose-insulin homeostasis, which suggests that restoring PAHSA levels in insulin-resistant individuals could have beneficial metabolic effects. In fact, Yore *et al.* showed that administration of these fatty acids to mice improved glucose uptake from blood, enhanced insulin secretion and relieved obesity-associated inflammation, suggesting that these naturally occurring fats could be used for diabetes therapy [6].

Despite the potential of these novel lipids for diabetes treatment [6,9,10], further studies are needed both to establish their normal physiological levels in humans and to study their evolution after a specific treatment. However, the low nanomolar concentrations at which FAHFAs are present in serum make necessary the development of a method with enough sensitivity to determine FAHFAs, even in deficiency state. One option is to preconcentrate FAHFAs by implementation of a solid-phase extraction (SPE) step, which has been used for quantitation of other compounds found at concentrations similar to FAHFAs such as prostaglandins [10], prostanoids [11,12], steroids [13,14], eicosanoids [15,16] or endocannabinoids [17,18].

In this research an automated qualitative and quantitative method based on on-line coupling of SPE and liquid chromatography–tandem mass spectrometry (LC–MS/MS) to maximize sensitivity has been developed for determination of FAHFAs. The method has been further applied to a cohort of individuals to evaluate the influence of the glycaemic state on FAHFA levels.

2. Materials and methods

2.1. Reagents

Palmitic acid-12-hydroxy-stearic acid (12-PAHSA), palmitoleic acid-12-hydroxy-stearic acid (12-POHSA), stearic acid-12-hydroxy-stearic acid (12-SAHSA) and oleic acid-12-hydroxy-stearic acid (12-OAHSA), all them in solutions, were purchased from Cayman Chemicals (Ann Arbor, MI, USA). Concretely, standard solutions contained 1 mg of each FAHFA in 200 μL of methyl acetate (5000 mg L^{-1}). Working solutions of each standard were prepared diluting the commercial standards at 500 μg L^{-1} in chromatographic grade methyl acetate from Scharlab (Barcelona, Spain).

Chromatographic grade methanol and ammonium hydroxide were purchased from Scharlab (Barcelona, Spain), while mass spectrometry grade ammonium acetate was purchased from Fluka (Spain). Deionized water from a Millipore Milli-Q water purification system was used for preparation of all aqueous solutions.

2.2. Instruments and apparatus

A microcentrifuge Sorvall Legend Micro 21R from Thermo Scientific (Waltham, MA, US) was used to separate the phases after protein precipitation. On-line sample preparation was carried out with an automated SPE workstation Prospekt-2 system from Spark Holland (Emmen, The Netherlands), which included an automatic cartridge exchanger (ACE) and a high-pressure syringe dispenser (HPD) for solvent delivery. The automated system was coupled to a Midas autosampler furnished with a 1 mL sample-loop. Peek tube (0.25 mm i.d.) from VICI (Houston, TX, USA) was used for all connections between the different valves. Full automation of the extraction step was controlled by Sparklink software version 2.10 from Spark Holland.

The SPE workstation system was on-line connected to an Agilent (Palo Alto, CA, USA) 1200 Series LC system for chromatographic separation furnished with a C18 Mediterranea column (10 × 4.6 mm i.d.; packed with 3 μm particle diameter size) from Teknokroma (Barcelona, Spain). The chromatograph was coupled to an Agilent (Palo Alto, CA, USA) 6460 triple quadrupole detector (QqQ) furnished with an Agilent JetStream Technology electrospray ion source (ESI). Nitrogen as source gas was provided by a high purity generator from CLAN Tecnológica (Sevilla, Spain), while ultrapure nitrogen (99.999%) used as collision gas was from Carburos Metálicos (Sevilla, Spain). Agilent MassHunter Workstation was the software for data acquisition, qualitative and quantitative analysis. Both the SPE and LC systems were configured for complete automation of analysis sequences.

2.3. Patient recruitment and selection

Serum samples were obtained from patients recruited at the Cardiology Unit of Miguel Servet Hospital (Zaragoza, Spain). The cohort was formed by a total of 84 individuals, who were diagnosed by atherosclerosis after evaluation through cardiac

catheterization (angiographic stenosis revealed a reduction of the arterial lumen ≥70%). The patients were classified according to their glucose level in blood and BMI. The distribution of individuals according to the created groups is shown in Table 1. The glucose level in blood was obtained by quantitative analysis using the autoanalyzers AU2700 and AU5400 from Beckman Coulter (Brea, CA, USA). All steps from blood extraction to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki of 2004 [19]. Individuals selected for this study were previously informed to obtain consent.

Table 1. Classification of patients according to body mass index (BMI) and glycaemic state.

Body mass index (kg m^{-2})	**Glycaemic state**[a] (mg dL^{-1})	**Individuals**
Normal weight 18.5<BMI<24.99	Normoglycaemic	12
	Prediabetic	7
Overweighed 25 <BMI<30	Normoglycaemic	12
	Prediabetic	10
	Diabetic	7
Obese BMI >30	Normoglycaemic	12
	Prediabetic	10
	Diabetic	7

[a]Levels of glucose in blood: normoglycaemic: between 60 and 100 mg dL^{-1}; prediabetic: between 100 and 126 mg dL^{-1}; and diabetic: >126 mg dL^{-1}.

2.4. Sample pretreatment

Venous blood was collected in evacuated sterile serum tubes without additives (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA). The blood samples were incubated for 30 min at room temperature to allow coagulation and centrifuged at 2000 × *g* for 15 min at 4 ºC to isolate serum fraction (processing within 2 h after collection). Samples were placed in plastic ware tubes and stored at −80 ºC until analysis.

Serum samples (500 µL) immersed in an ice bath were treated for deproteination by adding 1 mL methanol. The mixture was shaken for 1 min and the precipitate removed after centrifugation for 5 min at 4 ºC and 14500 × *g*. The upper phase was collected in a vial and stored at −80 ºC until analysis.

2.5. SPE–LC–MS/MS analysis

Deproteinized serum samples were subjected to automated SPE coupled to LC–MS/MS for FAHFAs analysis. Hysphere C8 cartridges (7 μm, 10 × 2.0 mm, Spark Holland) were used for the SPE step. Pumping of all solutions involved in the different SPE steps was carried out by the HPD unit. The analysis starts by solvation of the sorbent phase with 1 mL methanol and subsequent equilibration with 2 mL 20:80 (v/v) methanol–water before sample loading. After filling the 1-mL loop of the injection valve with the deproteinized sample, this is loaded into the C8 cartridge with 2 mL 20:80 (v/v) methanol–water as carrier solution, which leads the injected volume to the SPE sorbent. Under these conditions, the FAHFAs are retained in the cartridge, which is rinsed with 2 mL methanol to remove potential interferents. Then, the mobile phase pumped by the chromatograph passes through the cartridge for 7 min for elution of FAHFAs using the programmed chromatographic gradient. Then, the cartridge is washed with 2 mL of organic mobile phase.

The chromatographic conditions are based on the research proposed by Yore *et al.* [6], although this method was partially modified to reduce the analysis time. The chromatographic mobile phases are 5 mM ammonium acetate and 0.01% ammonium hydroxide in 98:2 methanol–water as organic mobile phase and 5 mM ammonium acetate and 0.01% ammonium hydroxide in 98:2 water–methanol as aqueous phase. The gradient used for LC separation starts with 97% organic mobile phase (maintained for 4 min) to 100% organic mobile phase in 12 min (also maintained for 4 min) at a flow rate of 0.8 mL min^{-1}. The total analysis time is 20 min, 5 additional min being required to re-establish the initial conditions and equilibrate the column. The compartment for the analytical column is maintained at 25 ºC.

The chromatographic eluate is directly introduced in the triple quadrupole detector QqQ for MS/MS detection in negative mode. The ESI parameters were set as follows: gas temperature, 350 ºC; drying gas, nitrogen 10 L min^{-1}; nebulizer pressure, 34 psi; sheath gas temperature, 350 ºC; sheath gas flow, 12 L min^{-1}; capillary voltage, 4000 V; and nozzle voltage, 300 V.

2.6. Quantitative analysis

Calibration models were built using the peak areas obtained by analysis of multistandard solutions of commercial 12-FAHSAs at different concentrations. Two calibration models were prepared for each 12-FAHSA by spiking methanol and a serum pool prepared from the set of samples used in this study with the target analytes. The calibration range was built with twelve concentrations ranging from 0.01 to 25 mg L^{-1}, which were analyzed in triplicate. Calibration curves obtained with commercial standards were used for quantitative analysis of the corresponding FAHSA. On the other hand, non hydroxy-stearic acid FAHFAs were relatively quantified by using the calibration model of the corresponding hydroxy-stearic acid FAHFAs. For example, the calibration model of PAHSA was used for quantitation of palmitic acid-hydroxy-oleic acid (PAHOA).

The limit of detection (LOD) for each analyte was calculated as $3.3 \times \sigma/S$, where σ is the standard deviation of the response of a blank analysis and S is the slope of the calibration curve for each analyte. The limit of quantitation (LOQ) was calculated as $10 \times \sigma/S$.

3. Results and discussion

3.1. Optimization of the LC–MS/MS method

A standard solution containing the four commercial 12-FAHSAs (12-PAHSA, 12-POHSA, 12-OAHSA and 12-SAHSA) was used to optimize the LC–MS/MS method. The multistandard solution was injected by direct infusion for selection of the precursor ion from all the commercial FAHSAs. All precursor ions, which are listed in Table 2 together with the optimum voltage set in Q1 for their efficient isolation, were identified as $[M-H]^-$ ions.

Fragmentation of the different FAHSAs was studied using collision energies from 0 to 45 eV. Table 2 shows the product ions selected for each commercial FAHSA and their optimum collision energy values.

Table 2. Main parameters of the LC–MS/MS MRM method for FAHSAs determination.

Analyte	Retention time (min)	Precursor ion (*m/z*)	Q1 voltage (V)[a]	Collision energy (eV)	Quantitative transition (*m/z*)[b]	Qualifier ions (*m/z*)
PAHSA	6.4	537	160	20	537→255	299, 281
POHSA	5.5	535	160	20	535→ 253	299, 281
SAHSA	8	565	200	20	565→283	299, 281
OAHSA	6.35	563	180	20	563→281	299, 281

[a]Voltage for filtration of precursor ions; [b]Transition selected for quantitation: Precursor ion (*m/z*) → Product ion (*m/z*).

Figure 1 shows the fragmentation scheme for PAHSA as a model compound to explain the product ions selected for each FAHFA. As can be seen, three main product ions were formed according to this fragmentation scheme. The most intense product ion corresponded to the fatty acid —in this example to palmitic acid carboxylate ion at *m/z* 255. The hydroxy fatty acid moiety fit the second product ion, which was formed by cleavage of the ester bound. This ion was obtained at *m/z* 299 for PAHSA. A third representative fragment was obtained by dehydration of the last ion—in this case at *m/z* 281. According to the sensitivity of the different transitions from precursor to product ions, those leading to the fatty acid carboxylate fragments were selected for quantitative purposes, while the other two transitions were selected with qualitative interest in the multiple reaction monitoring (MRM) method. Additionally, FAHFAs were confirmed using the Neutral Loss scan mode (NL), considering two neutral losses for each compound. These were formed by generation of the two ions selected as qualifiers. The information of the NL selected for each compound is reported as Supplementary Table 1. As can be seen, FAHFAs with the same main fatty acid in their structure were characterized by the same neutral losses.

Once the MRM method was configured, different chromatographic gradients were tested to attain an appropriate resolution. The best separation was obtained with methanol–water (98:2) as organic mobile phase, and water–methanol (98:2) as

aqueous mobile phase using 5 mM ammonium acetate and 0.1% (v/v) ammonium hydroxide as ionization agent in both phases.

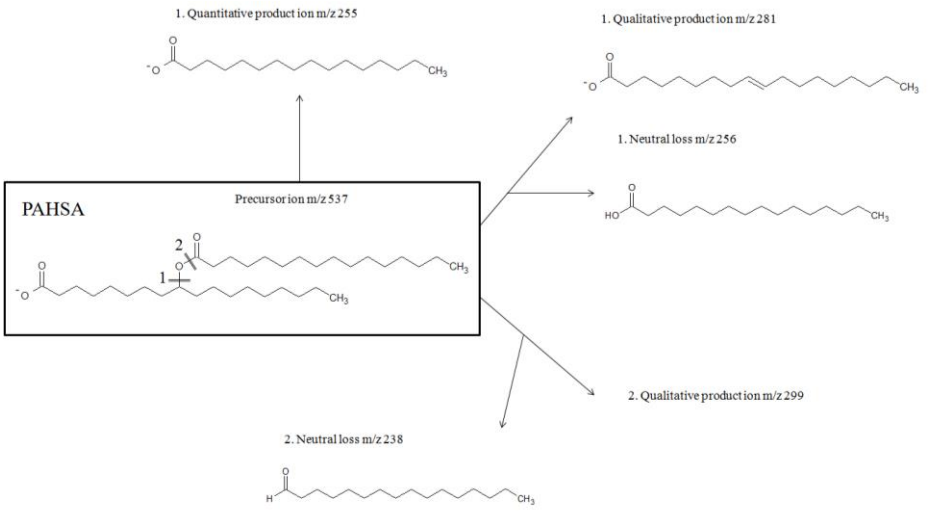


Figure 1. PAHSA precursor ion, product ions and neutral losses in negative ionization mode.

3.2. Development of the SPE protocol

Low concentrations of FAHFAs, below 20 nmol L^{-1} and 150 pmol g^{-1}, have been reported in serum and adipose tissue, respectively [8]. For this reason, an on-line SPE–LC–MS/MS configuration was adopted to endow the method with enough sensitivity to analyze FAHFAs.

An optimization study of the SPE process was designed by using a serum pool spiked with the four FAHSAs standards at 100 mg L^{-1} using the peak area as quantitative response. The optimization of the SPE protocol was divided into three main blocks: tests with sorbent materials, evaluation of the variables involved in each step of the SPE protocol and characterization of analytical features.

The four commercial sorbents tested, packed with the same technology from Spark Holland (Emmen, The Netherlands) and based on non-polar interactions, were C8 (end-capped silica- based octyl phase, particle size 10 μm), C18 (end-capped silica-based octadecyl phase, particle size 7 μm), C18HD (end- capped silica based phase with a high density of octadecyl chains, particle size 7 μm) and Resin SH (strong-hydrophobic modified polystyrene-divinyl benzene phase, particle size 20–50 μm).

Table 3. Variables studied, range tested for each variable and selected value after optimization of the SPE protocol for preconcentration and clean-up of FAHFAs in deproteinized human serum prior to on-line elution to LC–MS/MS.

Step	Variable	Range tested	Selected value
SPE sorbent		C8EC, C18EC, C18HD and Resin SH	C8EC
Solvation	Composition (%)		Methanol 100%
	Volume (mL)		1
	Flow rate (mL min^{-1})		5
Equilibration	Composition (%)	Methanol 0–30%	Methanol 20%
	Volume (mL)	1–2.5	2
	Flow rate (mL min^{-1})		5
Sample application	Composition (%)	Methanol 0–30%	Methanol 20%
	Volume (mL)	1–2.5	2
	Flow rate (mL min^{-1})		2
Wash cartridge	Composition (%)	Acetonitrile and methanol 50–100%	Methanol 100%
	Volume (mL)	0.5–3	2
	Flow rate (mL min^{-1})		1
Elution	Elution time (min)	5–15	7
Wash cartridge	Composition (%)	Mobile phase and methanol	Mobile phase
	Volume (mL)		2
	Flow rate (mL min^{-1})		1

A generic reversed-phase protocol was applied to evaluate the retention/elution capability of these sorbents. It consisted of sorbent conditioning with 2 mL of methanol,

equilibration with 2 mL of water, sample loading with 1 mL of methanol and elution with the chromatographic gradient. The sorbent with the highest retention/elution capability for FAHSAs was C8, the less polar of the studied sorbents (Supplementary Figure 1). According to the result, this sorbent was selected for development of the SPE protocol.

The variables involved in the SPE protocol were then optimized for C8 cartridge. The composition, volume and flow rate of solvents tested, as well as the optimum values obtained for each step are shown in Table 3. The most critical steps were sample loading, cartridge washing and elution of the target analytes. As can be seen, deproteinized samples were loaded in 20:80 (v/v) methanol–water so non-polar interactions of the FAHFAs and the sorbent are enhanced. A washing step with 2 mL methanol allowed removal of interferents such as glycerophospholipids that could promote ionization suppression effects. Losses of the target compounds by partial elution were not detected. Finally, the elution step was carried out according to the programmed chromatographic gradient thanks to the moderate alkaline pH of the mobile phase (pH 9) that is above the pKa of long-chain fatty acids. Therefore, the pH change favoured the elution of the FAHFAs as carboxylate forms. The elution time was also varied from 5 to 15 min and the optimum was 7 min.

3.3. Validation of the SPE method

The SPE method was characterized by preparation of calibration curves for the commercially available FAHSAs spiked in serum and methanol. In serum, the concentration of each FAHSA detected in non-spiked aliquots was taken into account to build the calibration models. The correlation coefficients, all above 0.99, and calibration ranges are shown in Table 4. The comparison of the slopes of the two calibration models revealed the presence of matrix effects. As can be seen, the sensitivity for compounds with long retention times was more affected by matrix effects. The sensitivity of the method was evaluated by estimation of the LOD and LOQ as described above. The absolute concentrations estimated as LOD and LOQ, expressed as pg mL^{-1}, are also listed in Table 4.

Table 4. Calibration models prepared by SPE–LC–MS/MS analysis of methanol and human serum both spiked with variable concentrations of the target analytes.

Methanol				
Analyte	PAHSA	POHSA	SAHSA	OAHSA
Slope	13432	34102	9327.3	41186
Intercept	536.64	6603.8	311.4	6406.2
Coefficient of regression (R^2)	0.9986	0.9988	0.992	0.9988
Limit of detection (LOD)[a]	0.065	0.026	0.059	0.041
Limit of quantification (LOQ)[a]	0.196	0.0796	0.180	0.0126
Serum pool				
Analyte	PAHSA	POHSA	SAHSA	OAHSA
Slope	8329	24929	2707	25023
Intercept	23245	58815	7203.5	62848
Coefficient of regression (R^2)	0.9942	0.9962	0.9954	0.9957
Limit of detection (LOD)[a]	0.065	0.003	0.098	0.004
Limit of quantification (LOQ)[a]	0.198	0.01	0.296	0.126

[a]LOD and LOQ expressed as ng mL^{-1}

The recovery factor for the FAHSAs was assessed by analysis of non-spiked and spiked serum aliquots at three concentrations (1, 5, and 10 μg L^{-1}). The recovery factor for the four compounds was studied with a two-cartridge configuration [20], in which two C8 cartridges were located in-series. Thus, after sample injection, the FAHSAs are mainly retained in the first cartridge, while the non-retained FAHSAs enter into the second cartridge to be trapped. The eluted fraction from both cartridges is sequentially injected into the QqQ analyzer to estimate the concentration retained in each cartridge. The recovery factor was calculated as amount retained in cartridge 1/(amount retained in cartridge 1 + amount retained in cartridge 2). Supplementary Table 2 lists the recovery factor (%) estimated for the four FAHSAs. These recoveries ranged between 73.8 and 100% depending on the spiked concentration. The general tendency was a lower recovery factor for higher spiked concentrations. In addition, metabolites with longer retention times provided lower recoveries. The same strategy was adopted to evaluate the extraction efficiency by analysis of a serum pool. The extraction efficiency

was estimated for all FAHFAs detected in the serum pool. SAHSA was not detected in serum samples, and the extraction efficiency for the rest varied from 77.8 to 100%, which is quite acceptable taking into account their concentrations.

The precision of the method was assessed by estimation of the within-day and between-days variability, both expressed as relative standard deviation (RSD). For this purpose, a single experiment with a serum pool and duplicate analysis per day [21] was planned for one week. The within-day variability ranged from 7.1 to 13.8%, and the between-days variability varied from 9.3 to 21.6%.

3.5. Application of the SPE–LC–MS/MS method to human serum samples

3.5.1. Sample analysis

As previously emphasized, Yore *et al.* evidenced that FAHFAs levels can be related to diabetes and obesity. For this reason, the association among FAHFAs with glycaemic state and BMI has been here studied by application of the developed method to the selected human cohort. First of all, the SPE–LC–MS/MS method was applied to identify FAHFAs present in the samples. The analysis was carried out by the MRM method including all potential transitions for all FAHFA combinations considering palmitic acid, palmitoleic acid, oleic acid and stearic acid. The application of the method to the studied cohort allowed detecting 11 FAHFAs (Supplementary Table 2) in at least 75% of the samples; 4 of them were confirmed by comparison with commercial standards. Supplementary Figure 2 shows MRM chromatograms for the six most concentrated FAHFAs detected in a serum sample. All samples were analyzed in triplicate and the mean value was employed for absolute or relative quantitation as described.

Figure 2 shows the profile of relative concentrations calculated as the percentage in the total amount of detected FAHFAs, for each group of individuals considering glycaemic state and BMI. As can be seen, similar levels were found for all detected FAHFAs in the three groups obtained considering obesity or glycaemic state. The most concentrated FAHFAs in serum were PAHSA and PAHOA, with concentrations around 35 and 25%, respectively; followed by stearic acid-hydroxy-oleic acid (SAHOA), oleic

acid-hydroxy-oleic acid (OAHOA) and oleic acid-hydroxy-stearic acid (OAHSA), with relative concentrations from 5 to 15%.

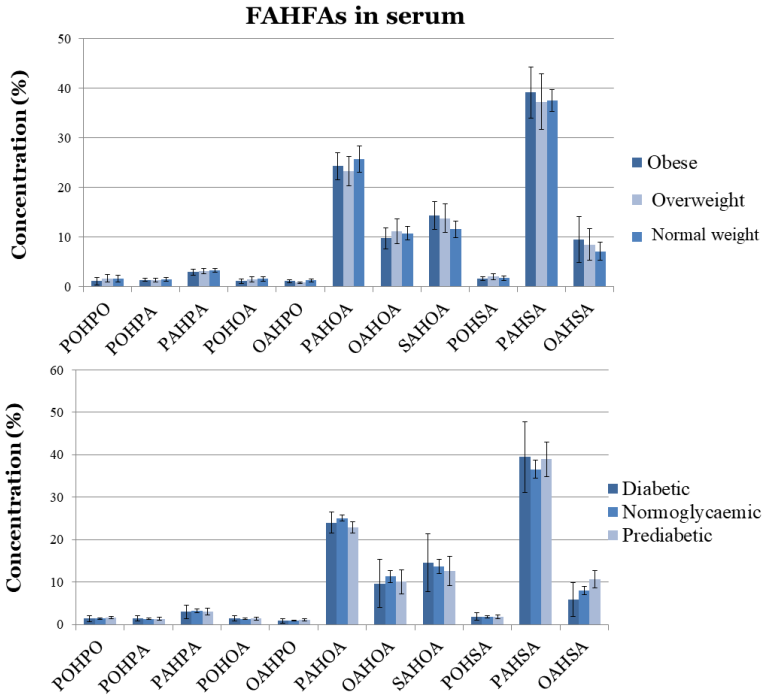


Figure 2. Relative concentration profile expressed in percentage of FAHFAs in individuals classified according to glycaemic state and BMI.

3.5.2. Statistical analysis

Spearman Rank correlation analysis, summarized in Table 5, revealed four high correlations (p-value < 0.0001 and R > 0.6), all of them positive, between the levels of the most concentrated FAHFAs. One of the correlations, corresponding to PAHOA–OAHOA (R = 0.64), was between compounds with the common hydroxy fatty acid moiety; while three correlations, PAHPA–PAHSA (R = 0.72), OAHOA–OAHSA (R = 0.70), and PAHSA–PAHOA (R = 0.66) were between compounds sharing the same fatty acid moiety.

Table 5. Spearman Rank correlation results for the pairs of serum FAHFAs showing high correlation.

Analyte		PAHSA
PAHPA	R^2	0.7221
	p-value	<0.001
		OAHOA
PAHOA	R^2	0.6454
	p-value	<0.001
		OAHSA
OAHOA	R^2	0.6986
	p-value	<0.001
		PAHOA
PAHSA	R^2	0.6598
	p-value	<0.001

Statistical analysis by the Kruskal-Wallis test using the normalized quantitative response (peak areas normalized by logarithmic transformation) revealed that the level of three FAHFAs was significantly different among individuals with different glycaemic state or BMI. Considering glycaemic state, two compounds reported significant differences in their concentrations: palmitic acid-hydroxy-palmitic acid (PAHPA), which resulted significant between diabetic and prediabetic individuals (*p*-value 0.0149); while palmitoleic acid-hydroxy-palmitoleic acid (POHPO) was significant for the comparison between prediabetic and normoglycaemic individuals (*p*-value 0.0251).

Figure 3 illustrates the variability found in the concentration of both FAHFAs as a function of the glycaemic state. As can be seen, the POHPO levels were higher in prediabetics as compared to normoglycaemic individuals, while PAHPA was found in diabetics at concentration significantly higher than in prediabetics.

On the other hand, BMI only contributed to explain the variability in the concentration of palmitoleic acid-hydroxy-palmitic acid (POHPA) when normal and overweighed individuals were compared (*p*-value 0.0300). As Figure 3 shows, the level of POHPA was found lower in overweight individuals as compared to normal weight individuals.

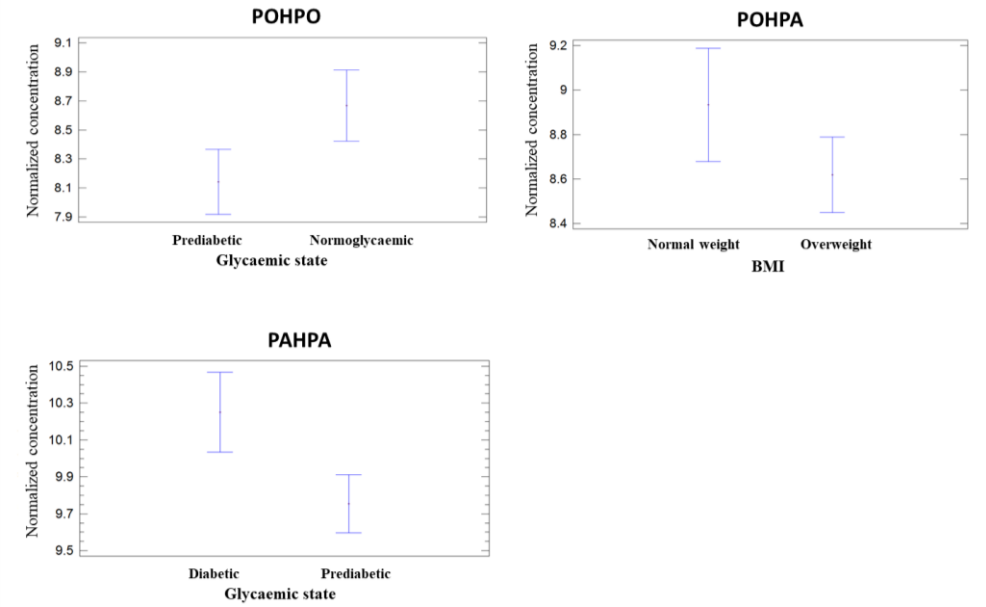


Figure 3. Means plots with 95% confidence level of significant FAHFAs for glycaemic state or BMI.

4. Conclusions

FAHFAs are endogenous mammalian lipids that had been associated to insulin sensitivity in humans [6]. These compounds are present at low nanomolar concentration in serum, thus hindering their detection. In this research, an automated method based on SPE on-line coupled to LC–MS/MS has been developed for determination of FAHFAs in serum. The developed research has provided the following conclusions: (i) the method shows excellent analytical features; (ii) eleven FAHFAs have been detected in serum by application of a confirmatory test; (iii) PAHSA and PAHOA were the most concentrated FAHFAs in serum; (iv) strong correlation has been detected between some FAHFAs; (v) PAHPA and POHPO reported significant differences among different glycaemic states, while only POHPA was found significantly different considering BMI.

Akcnowledgments

This research has been financed jointly by the Spanish Ministerio de Economía y Competitividad, ISCIII-Subdirección General de Evaluación, Fondo Europeo de Desarrollo Regional (FEDER) and Junta de Andalucía through Projects CTQ2015-68813-R, PIE14/00005/1 (FLEXI-MET) and FQM-1602. The authors are thanked to J.G. Galache (Miguel Servet Hospital, Zaragoza) for providing the serum samples used in this research, which were collected for development of the CTQ2009-07430 project. M.A. Lopez-Bascon also thanks the MICINN for an FPU scholarship (FPU15/02359).

References

[1] L. Guariguata, D.R. Whiting, I. Hambleton, J. Beagley, U. Linnenkamp, J.E.Shaw, Global estimates of diabetes prevalence for 2013 and projections for 2035. Diabetes Res. Clin. Pract. 103 (2013) 137–149.

[2] H. Cao, K. Gerhold, J.R. Mayers, M.M. Wiest, S.M. Watkins, G.S. Hotamisligil, Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism, Cell 134 (2008) 933–944.

[3] J.P. Després, I. Lemieux, Abdominal obesity and metabolomic syndrome, Nature 444 (2006) 881–887.

[4] J. Grünberg, WISP2 – A novel adipokine related to obesity and insulin resistance, Thesis University of Gothenburg, Doctoral Thesis, 2015.

[5] C. AustinPickens, L.M. Sordillo, S.S. Comstock, W.S. Harris, K. Hortos, B. Kovan, J.I. Fenton, Plasma phospholipids, non-esterified plasma polyunsaturated fatty acids and oxylipids are associated with BMI, Prostaglandins Leukot. Essent. Fat. Acids 95 (2015) 31–40.

[6] M.M. Yore, I. Syed, P.M. Moraes-Vieira, T. Zhang, M.A. Herman, E.A. Homan, R.T. Patel, J. Lee, S. Chen, O.D. Peroni, A.S. Dhaneshwar, A. Hammarstedt, U. Smith, T.E. McGraw, A. Saghatelian, B.B. Kahn, Discovery of a class of endogenous. mammalian lipids with anti-diabetic and anti-inflammatory effects, Cell 159 (2014) 318–332.

[7] D.M. Mouio, C.B. Newgard, The good in fat, Nature 516 (2014) 49.

[8] J. Ogawa, New lipid science in our inner ecosystem, Eur. J. Lipid Sci. Technol. 117 (2015) 577–578.

[9] C. Greenhill, Lipids: New class of endogenous lipids identified in mammals, Nat. Rev. Endocrinol. 11 (2014) 3–3.

[10] S. Rinne, C.R. Kleiveland, M. Kassem, T. Lea, E. Lundanes, T. Greibrokk, Fast and simple online sample preparation coupled with capillary LC-MS/MS for determination of prostaglandins in cell culture supernatants, J. Sep. Sci. 30 (2007) 1860–1869.

[11] C. Ferreiro-Vera, J.M. Mata-Granados, F. Priego-Capote, M.D. Luque de Castro, Automated method for targeting analysis of prostanoids in human serum by on-line solid-phase extraction and liquid chromatography–mass spectrometry in selected reaction monitoring. J. Chromatogr. A 1218 (2011) 2848–2855.

[12] M.S. Gachet, P. Rhyn, O.G. Bosch, B.B. Quednow, J. Gertsch, A quantitiative, LC-MS/MS method for the measurement of arachidonic acid, prostanoids, endocannabinoids, N-acylethanolamines and steroids in human plasma, J. Chromatogr. B 6 (2015) 976–977.

[13] U. Ceglarek, L. Kortz, A. Leichtle, G.M. Fiedler, J. Kratzsch, J. Thiery, Rapid quantification of steroid patterns in human serum by on-line solid phase extraction combined with liquid chromatography–triple quadrupole linear ion trap mass spectrometry, Clin. Chim. Acta 401 (2009) 114–118.

[14] F. Guo, J. Shao, Q. Liu, J. Shi, G. Jiang, Automated and sensitive determination off our anabolic androgenic steroids in urine by online turbulent flow solid-phase extraction coupled with liquid chromatography–tandem mass spectrometry: A novel approach for clinical monitoring and doping control, Talanta 125 (2014) 432–438.

[15] M.A. Fernández Peralbo, F. Priego-Capote, J.G. Galache-Osuna, M.D. Luque de Castro, Targeted analysis of omega-6-derived eicosanoids in human serum by SPE-LC-MS/MS for evaluation of coronary artery disease, Electrophoresis 34 (2013) 2901–2909

[16] C. Ferreiro-Vera, F. Priego-Capote, M.D. Luque de Castro, Integrated identification/ confirmatory and targeted analysis of epoxyeicosatrienosic acids in human serum by LC–TOFMS and automated on-line SPE–LC–QqQ MS/MS, Talanta 106 (2013) 440–447.

[17] E. Jagerdeo, J.E. Schaff, M.A. Montgomery, M.A. LeBeau, A semi-automated solid-phase extraction liquid chromatography/tandem mass spectrometry method for the analysis of tetrahydrocannabinol and metabolites in whole bloody, Rapid Commun. Mass Spectrom. 23 (2009) 2697–2705.

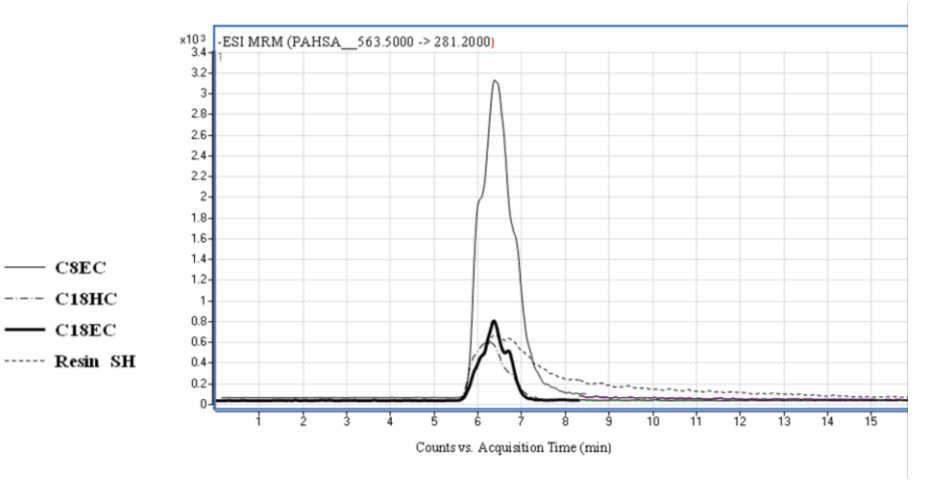
[18] M.G.J. Balvers, K.C.M. Verhoeckx, R.F. Witkamp, Development and validation of a quantitative method for the determination of 12 endocannabinoids and related compounds in human plasma using liquid chromatography–tandem mass spectrometry, J. Chromatogr. B 877 (2009) 1589–1590.

[19] Off. J. World Med. Assoc. 50 (2004) 50–51.

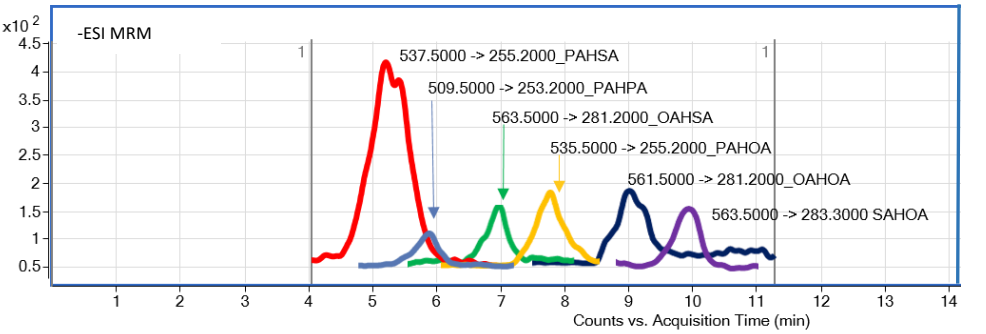
[20] J.A. Bert Ooms, G.J. Mark Van Gils, A.R. Duinkerken, O. Halmingh, Development and validation of protocols for solid-phase extraction coupled to LC and LC– MS, Am. Lab. 32 (2000) 52–57.

[21] D.L. Massart, B.G.M. Vanderginstre, L.M.C. Buydens, S. De Jong, P.J. Lewi, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics. Part A, 1st edition, Elsevier, Amsterdam, 1997.

Supplementary Information



Supplementary Figure 1. MRM chromatograms obtained with the eluate from the SPE cartridges tested by SPE–LC–MS/MS in negative ionization mode for PAHSA.



Supplementary Figure 2. MRM chromatograms of the six most concentrated FAHFAs detected in a serum sample. The name of the FAHFAs as well as the MRM transitions is specified above each peak.

Supplementary Table 1. Ions selected for identification and confirmatory analysis of FAHFAs in serum samples

Analyte	Precursor ion (*m/z*)	Product ions (*m/z*)[a]	Neutral loss (*m/z*)[b]
POHSA	535	253	254
		281, 299	236
OAHSA	563	281	280
		281, 299	264
PAHSA	537	255	256
		299, 281	238
SAHSA	565	283	284
		299, 281	266
POHPO	505	253	254
		251, 269	236
POHOA	533	253	254
		279, 297	236
OAHPO	533	281	280
		251, 269	264
POHPA	507	253	254
		253, 271	236
PAHPA	509	255	256
		253, 271	238
PAHOA	535	255	282
		279, 297	238
OAHOA	561	281	280
		279, 297	264
SAHOA	563	283	284
		279, 297	266
SAHPA	537	253	284
		253, 271	266
PAHOA	535	255	256
		279, 297	238

[a]First column contains product ion for quantification; second column lists product ions for qualitative confirmation; [b]First neutral loss is formed by cleavage of fatty acid; second neutral loss is formed by cleavage of the hydroxy-fatty acid (see Figure 1).

Supplementary Table 2. Recovery factor, expressed as %, estimated for each analyte with two-cartridges configuration for spiked and non-spiked serum samples.

			Extraction efficiency (%)	**Recovery factor (%)[b]**		
Analyte	**Retention time (min)**	**Transition[a]**	**Non-spiked**	**1 ng mL^{-1}**	**5 ng mL^{-1}**	**10 ng mL^{-1}**
POHSA	5.5	535→253	100	73.8	86.5	87.9
OAHSA	6.35	563→281	100	91	93.4	90.5
PAHSA	6.4	537→255	77.8	98.2	90	89.8
SAHSA	8	565→281	-	100	100	90.6
POHPO	3.5	505→253	93.2			
POHOA	4	533→253	78.5			
OAHPO	4.2	533→281	82.8			
POHPA	5.7	507→253	100			
PAHPA	6.7	509→253	93			
PAHOA	7.5	535→255	89.7			
OAHOA	9.3	561→281	100			
SAHOA	9.7	563→283	84.4			

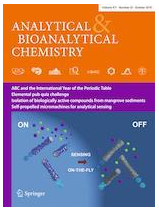
[a]Transition selected in Q1 → Q3. Precursor ion (*m/z*) → Product ion (*m/z*); [b] Calculated as amount retained in cartridge 1/(amount retained in cartridge 1 + amount retained in cartridge 2)*100.

CHAPTER II

Development of a qualitative/quantitative strategy for comprehensive determination of polar lipids by LC–MS/MS in human plasma

Springer

Accepted in
Analytical and Bioanalytical Chemistry



Development of a qualitative/quantitative strategy for comprehensive determination of polar lipids by LC–MS/MS in human plasma

María Asunción López-Bascón[a,b,c], Azahara Díaz-Lozano, Mónica Calderón-Santiago[a,b,c,], Feliciano Priego-Capote[a,b,c*]

[a]*Department of Analytical Chemistry, University of Córdoba, Annex C-3, Campus of Rabanales, Córdoba, Spain*

[b]*University of Córdoba Agroalimentary Excellence Campus, ceiA3, Córdoba, Spain*

[c]*Maimónides Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba Córdoba, Spain*

Development of a qualitative/quantitative strategy for comprehensive determination of polar lipids by LC–MS/MS in human plasma

María Asunción López-Bascón, Azahara Díaz-Lozano, Mónica Calderón-Santiago, Feliciano Priego-Capote

ABSTRACT

Polar lipids, especially glycerophospholipids, constitute the main components of cell membranes and are precursors of signaling molecules in many cellular and physiological processes. For this reason, the development of methods with high capability for detection of polar lipids in biological samples is required. In this research, the objective was to develop a method for comprehensive qualitative/quantitative determination of polar lipids in plasma by combination of acquisition methods with a triple quadrupole mass analyzer. The strategy was optimized in two steps: a) a first step for detection of lipids by monitoring selective fragmentation patterns representative of each lipid family; and b) a second step for confirmation of lipid species by detection and identification of product ions associated with the conjugated fatty acids. The acquisition list was divided in two MRM methods to ensure the detection of all transitions with suitable instrumental sensitivity according to chromatographic retention time and relative abundance in plasma. The combination of the two MRM methods allowed the detection of 398 polar lipids in plasma in 64 min. This strategy has been applied to a cohort of 384 individuals in order to obtain a qualitative and quantitative distribution of polar lipids in human plasma. The most concentrated lipid families in relative terms were lysophospholipids, plasmalogens and phosphatydilcholines, with mean relative concentration of 58.0, 17.1 and 8.3%, respectively. Then, sphingomyelins and phosphatidylethanolamines reported a relative concentration of 2.0%, followed by phosphatidylserines, with 1.1%.

Keywords: Lipidomics; polar lipids; LC–MS/MS; plasma; multiple reaction monitoring

SECTION II

Methodological development in untargeted metabolomics analysis: improvement of the analytical process

Section II of this PhD Book is devoted to methodological development in untargeted metabolomics analysis with the main aim of helping to overcome some of the limitations of this strategy. The influence of aspects related to sampling, sample preparation, detection and data analysis on the identification coverage has been studied in this section.

There is a great diversity of blood collection tubes for sampling serum or plasma, which are widely used in metabolomics analysis. In Chapter III, samples of serum or plasma collected in polymeric gel tubes were compared with those sampled in conventional tubes from a metabolomics perspective using an untargeted GC–TOF/MS approach.

The influence of sample preparation on untargeted lipidomics analysis of polar lipids in adipose tissue by LC–MS/MS was the aim of Chapter IV. Two common extractants used for lipids isolation, methanol:chloroform ($MeOH:CHCl_3$) and methyl *tert*-butyl ether (MTBE), were qualitatively and quantitatively compared for the extraction of the main families of lipids. Also, the implementation of an SPE step for selective isolation of glycerophospholipids prior to LC–MS/MS analysis was assayed to evaluate its influence on lipids detection coverage as compared to direct analysis.

In Chapter V an analytical strategy was planned to maximize the identification coverage of metabolites found in pig fecal samples. For this purpose, two complementary platforms such as LC–QTOF MS/MS and GC–TOF/MS were used. Concerning sample preparation six extractant with different polarity grade were tested to evaluate the extraction performance and, in the particular case of GC–MS, two derivatization protocols were compared. A total number of 303 compounds were tentatively identified by combination of all the extractants and analytical platforms.

The use of quality controls (QC) to remove instrumental variability has been studied in Chapter VI. A statistical package MetaboQC has been developed to study and filtrate instrumental variability in data sets generated by MS analysis of large sequences (programmed for several days). This tool uses QCs to individually correct any tendency on quantitative signals of metabolites that can be associated to instrumental variability.

CHAPTER III

Influence of the collection tube on metabolomic changes in serum and plasma

Talanta

150 (2015) 681–689

Influence of the collection tube on metabolomic changes in serum and plasma

María Asunción López-Bascón[a,b,c], Feliciano Priego-Capote[a,b,c,*], Ángela Peralbo-Molina[a,b,c], Mónica Calderón-Santiago[a,b,c], María Dolores Luque de Castro[a,b,c]

[a]*Department of Analytical Chemistry, University of Córdoba, Annex C-3, Campus of Rabanales, Córdoba, Spain*

[b]*University of Córdoba Agroalimentary Excellence Campus, ceiA3, Córdoba, Spain*

[c]*Maimónides Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba Córdoba, Spain*

Influence of the collection tube on metabolomic changes in serum and plasma

María Asunción López-Bascón, Feliciano Priego-Capote, Ángela Peralbo-Molina, Mónica Calderón-Santiago, María Dolores Luque de Castro

ABSTRACT

Major threats in metabolomics clinical research are biases in sampling and preparation of biological samples. Bias in sample collection is a frequently forgotten aspect responsible for uncontrolled errors in metabolomics analysis. There is a great diversity of blood collection tubes for sampling serum or plasma, which are widely used in metabolomics analysis. Most of the existing studies dealing with the influence of blood collection on metabolomics analysis have been restricted to comparison between plasma and serum. However, polymeric gel tubes, which are frequently proposed to accelerate the separation of serum and plasma, have not been studied. In the present research, samples of serum or plasma collected in polymeric gel tubes were compared with those taken in conventional tubes from a metabolomics perspective using an untargeted GC–TOF/MS approach. The main differences between serum and plasma collected in conventional tubes affected to critical pathways such as the citric acid cycle, metabolism of amino acids, fructose and mannose metabolism and that of glycerolipids, and pentose and glucuronate interconversions. On the other hand, the polymeric gel only promoted differences at the metabolite level in serum since no critical differences were observed between plasma collected with EDTA tubes and polymeric gel tubes. Thus, the main changes were attributable to serum collected in gel and affected to the metabolism of amino acids such as alanine, proline and threonine, the glycerolipids metabolism, and two primary metabolites such as aconitic acid and lactic acid. Therefore, these metabolite changes should be taken into account in planning an experimental protocol for metabolomics analysis.

Keywords: GC–TOF/MS, Metabolomics, plasma, serum, polymeric gel tubes, sampling tubes

1. Introduction

Human blood is the most common biofluid used by clinicians [1]. The main reasons justifying the clinical applicability of blood are its minimally invasive sampling [2], its homogeneity as compared to saliva or urine, which are strongly influenced by the collection volume [3], and its direct relationship with systemic changes in the metabolome [4,5]. Nevertheless, blood presents several limitations ascribed to its complicated and variable composition, as it contains approximately 4000 metabolites covering a wide range of concentrations and chemical diversity [6].

Two types of samples are obtained from blood, plasma and serum, which are used to evaluate various biochemical parameters demanded in the clinical field [4,5,7,8]. The collection of serum or plasma depends on allowing or blocking coagulation, respectively. During coagulation, fibrin clots are formed, then separated by centrifugation from serum together with blood cells and related coagulation factors, while platelets release proteins and metabolites into serum. On the other hand, plasma is obtained by addition of an anticoagulant (EDTA, heparin, citrate, etc.) before removal of blood cells by centrifugation [1]. There are no virtual interferences that can occur in serum owing to postcentrifugal coagulation [9,10]. Nevertheless, the presence of anticoagulants in plasma collection tubes can introduce some interferents, such as enzyme inhibitors, fibrinogen and cations [11]. The coagulation procedure creates some differences in the composition of metabolites between the two biofluids, such as in the levels of inflammatory markers. Thus, serum samples are preferred in assays to determine cardiac troponins, whereas plasma is recommended in glucose tolerance tests [1].

The impact of blood sample collection procedures on variations in the concentration of metabolites present in plasma or serum has been tested in several clinical studies [5,7,12–19]. Thus, Jorgenrud *et al.* (2015) studied the influence of three collection tubes (the conventional tube for serum and those containing citrate or EDTA for plasma) on the stability of some metabolites identified by UHPLC–QTOF MS/MS and GC×GC–TOF/MS [17]. Also, Barri and Dargsted (2013) searched for differences in plasma collected in different tubes (with EDTA, citrate or heparin) and the conventional

tube for serum by using untargeted analysis based on UPLC–QTOF MS/MS [7]. However, this study did not consider the inter-individual variability that could mask the real behavior of certain metabolites. Wedge *et al.* (2011) considered the inter-individual variability, but their study was focused on the differences between plasma and serum samples of patients with small-cell lung cancer by using untargeted analysis based on UPLC–MS and GC–QTOF [14]. Breier *et al.* (2014) reported metabolite measurements generally higher in serum as compared to plasma for saturated acylcarnitines, amino acids, biogenic amines, glycerophospholipids, sphingolipids and hexose. Additionally, the study was targeted at stability conditions finding the majority of metabolites stable for 24 h both on cool packs and at room temperature in non-centrifuged tubes [18].

There are several types of commercial tubes (tube wall, stopper, stopper lubricant, separator gel, clot activator, etc. [16] for collection of plasma or serum whose composition establishes the subsequent applicability of the collected sample. Among them, it is worth mentioning tubes containing a polymeric gel, an inert material that forms a barrier between the target sample (plasma or serum) and the rest of blood (packed cells) by centrifugation. The main benefits of separator gel tubes are an easy use, short processing time through clot activation, higher yield in the isolation of serum or plasma, reduced aerosolization of hazardous substances and a single centrifugation step [17]. On the other hand, some limitations of these tubes have also been described in the case of hydrophobic drugs such as phenytoin, phenobarbital, carbamazepine, quinidine and lidocaine, which could be adsorbed to the separator gel [17,20]. These tubes have been scarcely studied to assess their applicability in metabolomics analysis by comparison with the use of conventional tubes for serum or plasma. Breier *et al.* analyzed serum samples collected into conventional and gel-barrier tubes and found that serum metabolite concentrations were mostly unaffected by tube type, except for methionine sulfoxide that was significantly more concentrated in serum collected with gel-barrier tubes [18].

The objective of the present research was to study the differences at the metabolite level between serum and plasma obtained with conventional tubes (heparin tube for plasma) and polymeric gel tubes by application of an untargeted approach

based on GC–TOF/MS analysis. For this purpose, a cohort of volunteers was selected for blood sampling, then collected using four different tubes (plasma, plasma-gel, serum and serum-gel). Sample preparation was based on protein precipitation with an organic solvent followed by silylation as derivatization procedure. The differences between the two main clinical samples, plasma and serum, and those ascribed to their collection in polymer gel tubes were discussed taking into account the sources of intra-individual and inter-individual variability.

2. Materials and methods

2.1. Reagents

Both MS-grade *n*-hexane from Sigma–Aldrich (St. Louis, MO, USA) and methanol from Scharlab (Barcelona, Spain) were used for sample preparation. Bis-(trimethylsilyl) fluoroacetamide (BSTFA) and trimethylchlorosilane (TMCS) from Sigma–Aldrich were used as silylation agents in the derivatization step. Pyridine from Merck (Darmstadt, Germany) was used as solvent for derivatization. Mass spectrometry grade perfluorotri-*n*-butylamine (PFTBA) from Agilent Technologies (Santa Clara, CA, USA) was used for daily mass calibration. An alkane standard mixture (from C10 to C40) designed for performance test in GC from Sigma–Aldrich was used to establish the retention index (RI) calibration. Triphenyl phospate (TPP, from Sigma–Aldrich) was used as external standard (ES) and standards of D-mannitol, D-fructose, D-sorbitol, myo-inositol, D-glucose and D-mannose were purchased to confirm identification of sugars.

2.2. Apparatus and instruments

A micro-centrifuge Sorvall Legend Micro 21R from Thermo Scientific (Waltham, MA, US) was used to separate the phases after extraction and protein precipitation. A speed-vac Concentrator Plus, from Eppendorf Ibérica (Madrid, Spain), was used to evaporate the methanol phase before derivatization. A block heater from Stuart Equipment (Staffordshire, OKA, USA) was used in the derivatization step.

An Agilent 7890A Series GC system coupled to an Agilent 7200 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with an electron impact (EI) source (Santa Clara, CA, USA) was used for analysis. The analytical sample was thus monitored in high resolution mode. MassHunter GC QTOF Acquisition software (version B.06, Agilent Technologies) was used to control data acquisition and set the parameters for optimum operation.

2.3. Samples

All experiments were carried out in accordance with the ethical principles of human medical research (World Medical Association, Helsinki Declaration, 2004 [21]). The ethical review board of Reina Sofía University Hospital (Córdoba, Spain) approved and supervised the clinical study.

The samples were obtained from thirteen healthy volunteers (3 men and 10 women) at Reina Sofía University Hospital (Córdoba, Spain). The steps from blood extraction to analysis were supervised by specialized personnel pertaining to this hospital. Blood from each volunteer was collected into four different Vacutainer® tubes (Becton Dickinson): plastic serum tubes with spray-coated silica (serum), plastic serum tubes with spray-coated silica and a polymer gel to favor serum separation (serum-gel), spray-coated silica tubes with heparin for plasma (plasma) and heparin tubes with polymer gel to favor plasma separation (plasma-gel).

The collected samples were processed by centrifugation for 15 min at 1000 × *g* for isolation of serum and plasma in conventional tubes, and at 2000 × *g* for gel separation tubes. The serum fraction was processed 1 h after collection to ensure complete coagulation.

2.4. Untargeted analysis

Plasma and serum samples (50 μL) were deproteinized with 150 μL of methanol. The mixture was vortexed for 1 min and subsequently cooled at −20 ºC for 3 min. The resulting precipitate was separated by centrifugation at 14000 × *g* for 15 min at 4 ºC and the methanol–aqueous phase was isolated, was dried by evaporation and the residue reconstituted with 5 μL of pyridine and 1 μL of formic acid. For derivatization,

45 µL of a 98:2 BSTFA–TMCS mixture was added to the reconstituted residue, shaken for 30 s and maintained at 30 °C for 30 min, and a small amount of anhydrous sodium sulfate was added to remove residual water before injection into the chromatograph. 4 µL of each analytical sample was diluted in 192 µL of *n*-hexane and 4 µL of ES (140 µg L^{-1}). All samples were prepared and analyzed in triplicate.

GC–TOF/MS analyses were performed by EI ionization mode at 70 eV. Chromatographic separation was carried out with a fused silica DB-5MS-UI (30 m × 0.25 mm i.d, × 0.25 µm) film thickness capillary column. The GC oven temperature program started at 60 °C (1 min held), followed by a temperature ramp of 10 °C min^{-1} to final 300 °C (2 min held). Post-run time was programmed for 4 min up to 310 °C to assure complete elution of the injected sample. Pulsed splitless injections of 1 µL of sample were carried out at 250 °C and ultrapure grade helium was used as carrier gas at 1.0 mL min^{-1} flow rate. The interface, ion source and quadrupole temperatures were set at 280, 300 and 200 °C, respectively. A solvent delay of 5.5 min was used to prevent damage in the ion source filament. The TOF detector was operated at 5 spectra s^{-1} in the mass range *m/z* 50–550 and the resolution was 8500 (full width half maximum, FWHM) at *m/z* 501.9706. A daily mass calibration was performed with PFTBA. Tentative identification of compounds was performed by searching MS spectra on the National Institute of Standards and Technology (NIST, version 11, 2011) database taking into account the RI values.

2.5. Identification of metabolites

Identification was firstly carried out by searching MS spectra on the NIST11 database. Only those identifications with a match factor and a reverse match factor higher than 700 were considered as valid. The RI values included in the NIST database were also taken into account to support identifications. An RI calibration model was built by plotting the retention times obtained by analysis of the alkane standard mixture (C10 to C40 with an even number of carbons) with the chromatographic method used in this research and the RI values provided for each alkane by the NIST database. Supplementary Figure 1 shows the RI calibration graph and the equation fitting this model. Then, the RI value was experimentally estimated for each identified compound

by using the retention time and the calibration equation. The requirement to accept NIST identifications was that the difference between the experimental RI and the theoretical value provided by the NIST for each target compound should be below 100 units.

The NIST database does not contain high resolution MS information as provided by the TOF detector. For this reason, a third step was included to validate identification of each compound by using high resolution MS. Thus, the molecular formula for the $[M]^+$ ion and the most intense fragments for each molecular feature (MF) should fit the NIST identification by setting a cut-off value in mass accuracy of 10 ppm. Additionally, five derivatized standards (D-mannitol, D-fructose, D-sorbitol, myo-inosiol, D-glucose and D-mannose) were injected for identification of sugars, thus confirming their presence in the samples.

2.6. Data processing and statistical analysis

Unknown Analysis software (version 7.0, Agilent Technologies, Santa Clara, CA, USA) was used to process all data files obtained by GC–TOF/MS in full scan mode. Treatment of raw data files started by deconvolution of chromatograms to obtain a list of MFs, which are considered as potential compounds defined by five *m/z* values of representative ions for each chromatographic peak (one quantifier and four qualifiers) and the retention time. For this purpose, the deconvolution algorithm was applied to each sample by considering all ions exceeding 1500 counts for the absolute height parameter, the accuracy error at 50 ppm and the window of 0.3 min. The list of MFs obtained for each analysis was exported as data files in compound exchange format (*.cef* files). These data files were treated with the Mass Profiler Professional (MPP) software package (version 12.1, Agilent Technologies, Santa Clara, CA, USA) to build the data set including quantitative information (expressed as peak area ratio) of MFs detected in each analysis for further processing.

In the next step, the data were processed by alignment of the potential MFs across samples according to their retention time and *m/z* values using a tolerance window of 0.3 min and an accuracy error of 15 ppm. Finally, the MFs resulting after data pretreatment were exported (*.cef* file) for recursive targeted analysis. For this purpose,

the Quantitative Analysis software (version 7.0, Agilent Technologies, Santa Clara, CA, USA) was used to reintegrate all potential metabolites found in all analyzed samples, using the corresponding characteristic quantifier and qualifiers. The data set was exported and reprocessed with MPP software package (version 2.2, Agilent Technologies, Santa Clara, CA, USA). The data exported were normalized prior to statistical analysis by logarithmic transformation. The MPP software also allowed unsupervised analysis by principal component analysis (PCA) and Mann-Whitney *t*-test corrected by the Benjamini-Hochberg algorithm to evaluate significant differences in concentration of each compound (*p*-value lower than 0.05) in the studied samples. Mapping Metabolites on Kyoto Encyclopedia of Genes and Genomes (KEGG) (URL: http://www.genome.jp/kegg/) was used to know the metabolomic pathways in which the significant metabolites were involved to find connections among them.

3. Results and discussion

3.1. Quality control study

Quality control (QC) studies are crucial in metabolomics experiments dealing with biological samples such as serum or plasma because the variability sources potentially involved. The selection of an internal standard in untargeted analysis is a difficult task as clinical samples typically contain between 100 and 1000 metabolites covering a wide range of concentrations and structures. Even the selection of an internal standard representative for each family of compounds does not ensure to operate under quality conditions. For this reason, it is also frequent the selection of external standards to control external variability sources and accumulate repetitions of each sample to minimize the experimental variability. In this research, TPP was selected as ES due to its exogenous character and its intermediate retention time within the interval of the selected chromatographic method (elution at 21.6 min).

A total number of 65 compounds were identified in the samples analyzed in this study. The compounds were identified at least in all the samples pertaining to one of the four target types (serum, plasma, serum-gel and plasma-gel). These compounds are

well-known metabolites previously detected in human blood such as amino acids, hydroxy acids, carboxyl acids, fatty acids or carbohydrates [1,4,6]. Supplementary Table 1 lists the name of the metabolites, chemical class, code (if exists), retention time and fragments for each entry. As can be seen, the main classes of identified compounds were amino acids (14 metabolites), fatty acids (14 metabolites), carbohydrates (11 metabolites), keto acids (5 metabolites) and carboxylic acids (4 metabolites). It is worth mentioning that several of them, particularly amino acids and derivatives, were exclusively detected in certain types of samples. One amino acid, proline, was only found in serum and serum-gel samples, while valine was detected in serum, serum-gel and plasma samples. No exclusive metabolites were detected in plasma-gel samples.

It is practically impossible to obtain high accuracy and precision for all detected metabolites in untargeted metabolomics analysis [22]. In this context, before comparison of the metabolite profiles detected in the four types of samples, evaluation of the variability of the identified compounds is mandatory. Thus, the relative standard deviation (RSD) was calculated for each compound per individual and per type of sample by considering the thirteen individuals. A great number of metabolites reported an average intra-individual variability, expressed as RSD, above 30%, despite the use of ES. This variability could be assigned to instability according to the protocol used for analysis. Table 1 includes the list of metabolites that gave RSD values higher than 30% in the different samples. The general trend observed for these metabolites was that the relative concentration (expressed as peak area ratio) significantly decreased since the first injection to the third. Thus, metabolites such as urea, valine, proline, succinic acid or arachidonic acid experienced this pattern in one or several samples. This variability source has been pointed out by other authors both in serum and plasma [4,14], and it was attributed to metabolism activity or degradation produced during analysis

The comparison among types of samples revealed that above 20 metabolites suffered this variability in serum (by setting the RSD cut off at 30%); on the other hand, only 8 compounds showed this behavior in plasma-gel samples.

Table 1. Identified metabolites with a relative standard deviation (RSD) higher than 30% in the four types of samples (serum, serum-gel, plasma and plasma-gel).

Compound name[a]	Serum	Serum-gel	Plasma	Plasma-gel
Acetic acid		X	X	X
Alanine	X			
Aminobutyric acid	X			
Aminomalonic acid	X			
Arachidonic acid	X			
Ascorbic acid	X			
Butyric acid		X	X	
Cysteine	X		X	
Erythronic acid	X			
Glycine	X		X	X
Lactic acid	X		X	
Leucine	X	X	X	X
Oxoglutarate	X		X	
Phenylalanine	X	X	X	X
Proline	X	X		
Pyroglutamic acid	X	X	X	X
Serine	X	X	X	X
Succinic acid	X	X	X	
Threonine	X		X	
Urea	X	X	X	X
Uridine	X			X
Valine	X	X	X	X
2,3-Dihydroxybutanoic acid		X		
3-Hydroxybutyric acid			X	

[a]The identified compounds are TMS derivatives.

In general, the samples collected in gel- separator tubes reported a lower number of metabolites affected by this variability source than those collected in conventional tubes, either plasma or serum samples. This result is in agreement with those reported by O'Keane *et al.* (2006) who found an enhanced metabolite stability in samples (serum or plasma) collected in tubes containing a polymeric gel [23]. Six metabolites (leucine, phenylalanine, serine, valine, pyroglutamic acid and urea) provided intra-individual variability higher than 30% in all types of samples. Other three metabolites (lactic acid,

threonine and cysteine) only led to a lack of precision in plasma or serum collected without gel separator. According to these results, the two most important chemical families affected by intra-individual variability were amino acids and carboxylic acids. Supplementary Figure 2 shows the bars diagram of the relative quantitative response of lactic acid, urea and alanine, as provided by the three replicates in the analysis of each of the four samples. As can be seen, the quantitative response for these three metabolites decreased from the first to the third replicate; decrease especially significant for urea, which was not detected in the third replicate of any sample. On the other hand, glycerol was characterized by the absence of variability between replicates. The two families with the lowest intra-individual variability were sugars and long chain fatty acids, except for arachidonic acid in serum. After this study, the quantitative response for the metabolites exceeding the intra-individual variability cut-off was estimated from one or two injections of the target analytical sample.

Concerning sensitivity, the analysis of serum samples provided higher sensitivity than plasma samples, behavior that had been previously reported [1]. Figure 1 shows the base peak chromatograms (BPCs) obtained by analysis of serum, serum-gel, plasma and plasma-gel samples from the same volunteer using the untargeted protocol based on GC–TOF/MS. Some representative metabolites pertaining to different chemical families are numbered to facilitate comparison of the differences in the quantitative response provided by each of them.

3.2. Metabolite differences between serum and plasma

Once the data set was extracted from the raw data files, statistical analysis was applied to evaluate differences in the metabolite profile obtained by untargeted analysis of serum, plasma, serum-gel and plasma-gel. Theoretically, the inclusion of a gel to accelerate the process of serum or plasma separation should not change the metabolite composition of the samples because of the inertness of gel [17].

However, the clinical use of these samples justifies the application of multivariate statistical tools by unsupervised analysis looking for discrimination patterns.

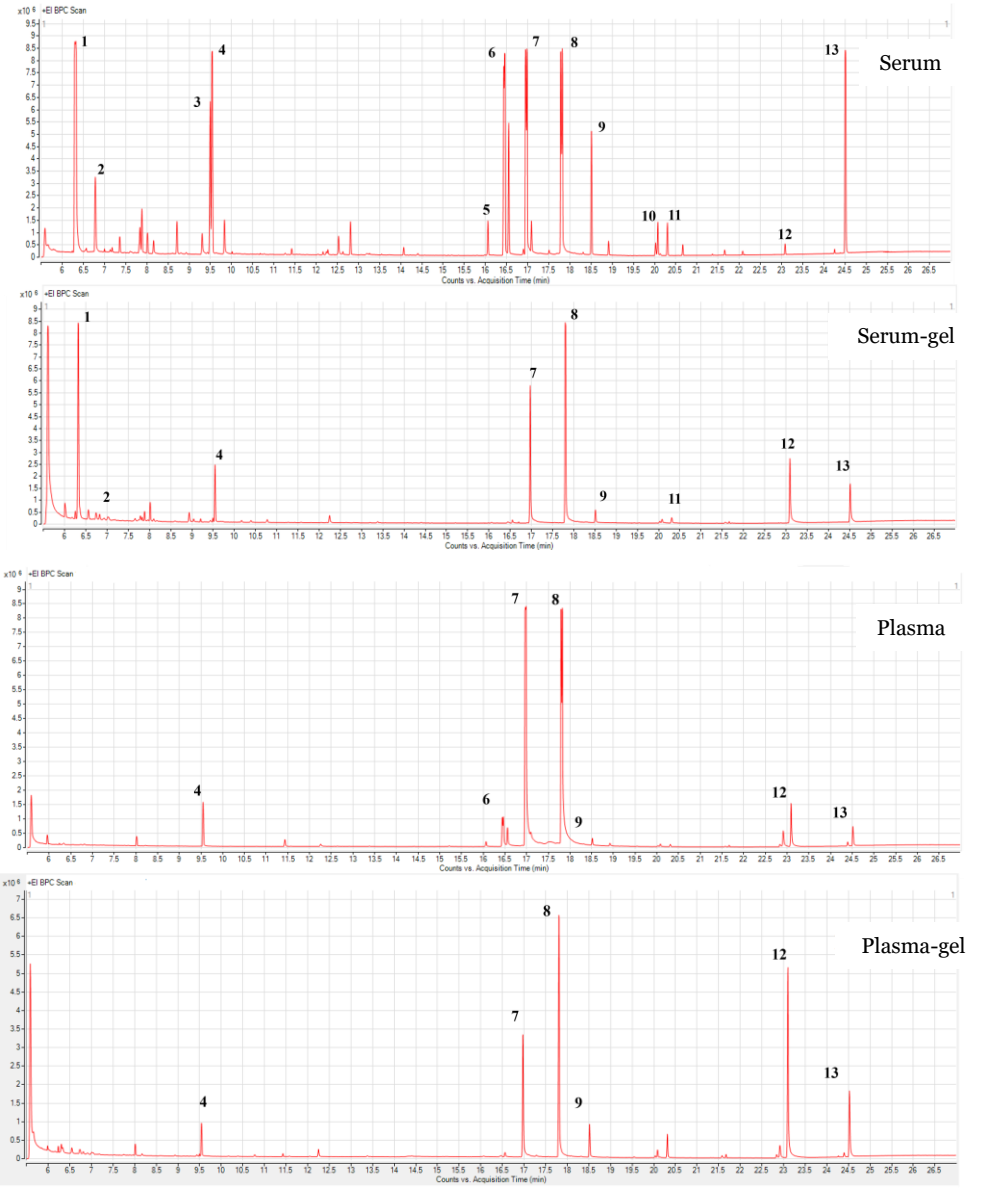


Figure 1. Base peak chromatograms (BPCs) from serum, serum-gel, plasma and plasma-gel samples from the same volunteer obtained by GC–TOF/MS. 1) Lactic acid; 2) valine; 3) phosphoric acid; 4) glycerol; 5) α-D-mannopyranose; 6) sorbitol; 7) β-D-mannopyranose; 8) glucopyranose; 9) palmitic acid; 10) oleic acid; 11) stearic acid; 12) monopalmitin; 13) monostearin.

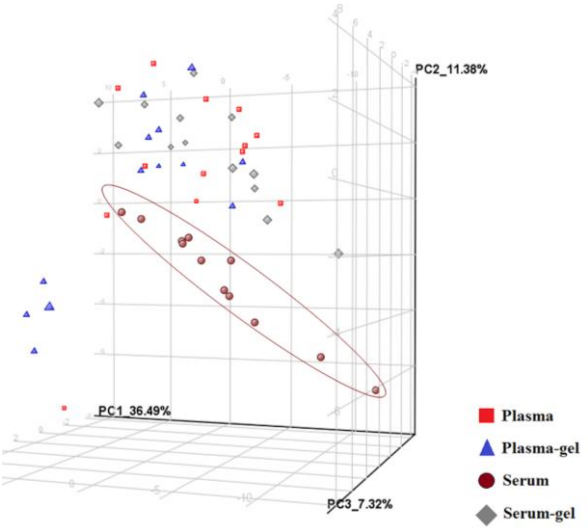


Figure 2. PCA scores plot in a 3D graph obtained from the analysis of the four types of samples: serum, serum-gel, plasma and plasma-gel.

Figure 2 shows the 3D PCA scores plot obtained from the four types of samples that explains a variability of 52.1% and shows a clear separation in three groups that explains the main variability sources. Thus, separation between serum samples and serum-gel samples could be attributed to the effect of the polymeric gel. Apart from this discrimination, plasma and plasma-gel samples overlapped, forming a third group distributed in an area between the two types of serum samples. Therefore, the analysis of plasma or serum provides metabolite differences, but also the use of a polymeric gel to favor serum isolation promotes differences in the profile of the monitored metabolites.

After detecting variability in the metabolite profile of serum and plasma samples, the metabolite differences between serum and plasma collected within the conventional tubes were analyzed. Figure 3 shows the 2D PCA scores plot explaining 46.98% of the total variability.

Discrimination between the metabolite profile of serum and plasma had previously been demonstrated by other authors [1,4,11,14,18]. Among them, our study

can be considered complimentary of that by Breier *et al.* [18], who used LC–MS in contrast to GC–MS used in our case. As discussed in the introduction, the main result described by Breier *et al.* was that metabolite measurements were higher in serum as compared to plasma, with special emphasis on certain families of metabolites such as saturated short- and medium-chain acylcarnitines, amino acids, biogenic amines, glycerophospholipids, sphingolipids and hexose.

Table 2. Identified metabolites with significant differences in their levels in plasma and serum samples. The pathways in which they are involved are also listed.

Compound[a]	***p*-value**[b]	**FC**[c]	**Pathways**
Phenylacetic acid	0.0463	-1.4318	Phenylalanine metabolism
Sorbitol	0.0401	3.6497	Metabolims: fructose/mannose and galactose
Lactic acid	0.0401	1.0148	Glycolysis/gluconeogenesis and metabolism: pyruvate, propanoate and fructose/mannose
Tyrosine	0.0401	-16	Metabolism: tyrosine, nitrogen, phenylalanine, ubiquinone and thiamine
Mannose	0.0346	4.6908	Metabolism: fructose/mannose and galactose
Myo-inositol	0.0334	2.9793	Pentose and glucuronate interconversions
Proline	0.0331	-16	Biosynthesis amino acids and arginine and proline metabolism
Glycerol	0.0234	-2.3692	Pentose and glucuronate interconversions and metabolism: galactose and glycerolipid
Monopalmitin	0.0234	10.808	Glycerolipid metabolism
Dodecanol	0.0234	1.684	-
Derivate of cholesterol	0.018	16	Biosynthesis of steroids by oxidation of colesterol or biosynthesis of primary bile acids
Uridine	0.0078	-2.9412	Pentose and glucuronate interconversions and galactose metabolism
Threonine	6.00E-04	-6.9353	Biosynthesis amino acids and glycine, serine and threonine metabolism
Oxoglutaric acid	0	-5.7524	Biosynthesis of amino acids and metabolism: 2-Oxocarboxylic acid, carbon, glyoxilate/dicarboxylate, Alanine/aspartate/glutamic, ascorbate/aldarate, vitamin B6, lisien, butanoate and histidine

[a]The identified compounds are TMS derivatives; [b]Value corrected by the Benjamini-Hochberg method; [c]FC: Fold change.

Most of these families, except for amino acids and hexose, are preferentially detected by LC–MS and, for this reason, the results obtained in the research proposed here could complement those reported by Breier *et al.* [18]. A Mann-Whitney *t*-test with Benjamini-Hochberg algorithm correction was applied to identify the metabolites with the highest statistical contribution to differentiate serum and plasma samples. Fourteen metabolites, listed in Table 2, presented significant different concentration in serum and plasma with a p-value < 0.05. As can be seen, most of them provided p-values between 0.05 and 0.01, except for three metabolites: threonine (p-value 0.0006), uridine (p-value 0.008) and a cholesterol derivative (p-value 0.0002).

Table 3 lists the main pathways in which the identified metabolites are involved, and Supplementary Figure 3A shows these pathways and the connection among significant metabolites. Among them, it is worth mentioning oxoglutarate, involved in the citric acid cycle, and found at higher concentration in serum than in plasma, as Supplementary Figure 4 shows. The involvement of the metabolite in this cycle allows connecting this critical pathway to others dealing with the metabolism of amino acids, particularly ascorbate/aldarate, alanine/aspartate/glutamate and glutamine/glutamate pathways (Supplementary Figure 3A). One other metabolite reporting significant differences of its levels in plasma and serum was myo-inositol that takes part in the ascorbate/aldarate pathway as a precursor of glucuronate. The ascorbate/aldarate metabolic pathway is strongly linked to the fructose and mannose metabolism. Three metabolites involved in the last pathway (sorbitol, mannose and lactate) were found significantly different by comparing serum and plasma levels. Sorbitol, mannose, myo-inositol and lactate were more concentrated in serum than in plasma, as Supplementary Figure 4 illustrates. Mannose is a key substrate of glycolysis, which is other relevant pathway also including the third metabolite previously mentioned (lactate, a precursor of pyruvate). Lactate is related to one other metabolite of the citric acid cycle: oxalacetate, which is first converted to pyruvate, and then to lactate. Oxalacetate is the precursor of aspartate, involved in the formation of proline and threonine, two amino acids with also different levels in plasma and serum in the selected cohort. In fact, proline was not detected in plasma samples, while threonine provided a small signal in this type of sample. The metabolism of phenylalanine showed also significant

differences depending on the sample; thus, this amino acid, precursor of tyrosine, was found significant and similarly to other amino acids not detected in plasma, and phenylacetate, a metabolite of phenylalanine that can be hydroxylated prior to participate in the metabolism of tyrosine. This derivative was found at higher concentration in plasma than in serum.

Table 3. Identified metabolites with significant differences in their levels in serum and serum-gel samples. The pathways in which they are involved are also listed.

Compound[a]	***p*-value**[b]	**FC**[c]	**Pathways**
Aminomalonic acid	0.0251	-4.9247	-
Glycerol	0.0251	-1.7515	Pentose and glucuronate interconversions and metabolism: galactose and glycerolipid
Monopalmitin	0.0222	9.9298	Glycerolipid metabolism
Monostearin	0.0185	-3.6099	Biosynthesis: fatty acids and unsaturated fatty acids
Proline	0.0166	-16.0000	Biosynthesis amino acids and arginine and proline metabolism
Aconitic acid	0.0166	-38.2050	-
Dodecanol	0.0166	1.9543	-
Lactic acid	0.0143	-1.6320	Glycolysis/gluconeogenesis and metabolism: pyruvate, propanoate and fructose/mannose
Threonine	0.0134	-7.4579	Biosynthesis of amino acids and glycine and serine/threonine metabolism
Alanine	0.0018	-16.0000	Biosynthesis of amino acids and metabolism: Alanine/aspartate/ glutamate, cysteine/methionine, D-alanine, taurine/hypotaurine and selenocompound
Derivate of cholesterol	0.0002	-3.8270	Biosynthesis of steroids by oxidation of colesterol or biosynthesis of primary bile acids

[a]The identified compounds are TMS derivatives; [b]Value corrected by the Benjamini-Hochberg method; [c]FC: Fold change.

Oxoglutarate is involved in the pentose/glucuronate interconversion pathway, in which one important metabolite such as glycerol was also found at lower concentration in plasma than in serum samples. Glycerol, which is essentially produced by glycolysis, is the precursor of glycerophospholipids and acylglycerols, as illustrates Supplementary Figure 3A. Among them, monopalmitin was found at higher concentration in plasma than in serum. On the contrary, glycerol, inositol, sorbitol and mannose, products from galactose by the action of α-galactosidase, were characterized for a higher concentration in serum than in plasma. α-Galactosidase takes part in the galactose metabolism in

which UDP-glucose participates, which is the precursor of uridine, this latter also detected at different levels in both clinical samples.

One other significant compound (*p*-value 0.0002) was a cholesterol derivative not found in serum. According to the databases, two candidates could fit the molecular formula $C_{27}H_{44}O_2$ assigned by MS in high resolution mode. The two candidates are hydroxy derivatives obtained from cholest-4-en-3-one, which participate in two different pathways. Thus, 26-hydroxycholest-4-en-3-one is produced in the biosynthesis of steroids by oxidation of cholesterol, while 7-hidroxycholest-4-en-3-one is formed in the biosynthesis of primary bile acids by a similar mechanism. According to the structure of both compounds, it is impossible to elucidate the exact structure of this compound with the experimental method used in this study. The last compound was dodecanol, the origin of which should be mainly ascribed to exogenous sources through the diet. Anyway, this fatty alcohol was found at higher concentration in plasma than in serum samples (Supplementary Figure 4).

3.3. Influence of the polymeric gel on the metabolomic analysis of plasma and serum

The next step was targeted at detecting metabolite differences in serum and plasma collected in polymeric gel tubes as compared to conventional tubes. Despite the benefits ascribed to these tubes concerning pretreatment time, no analytical studies had been carried out to identify metabolite differences when compared with serum and plasma collected in conventional tubes. Initially, unsupervised analysis by PCA was applied to compare both the metabolite profile in serum *vs*. that found in serum-gel and plasma *vs*. plasma-gel. The variability explained was 67.01%. As Figure 3B shows, a clear discrimination exists between serum and serum-gel samples, which was previously seen in the PCA dealing with the four groups of samples. On the other hand, no discrimination seems to exist between plasma and plasma-gel samples; therefore, no statistical differences in the metabolite composition could be attributed to the use of plasma tubes or those containing the polymeric gel.

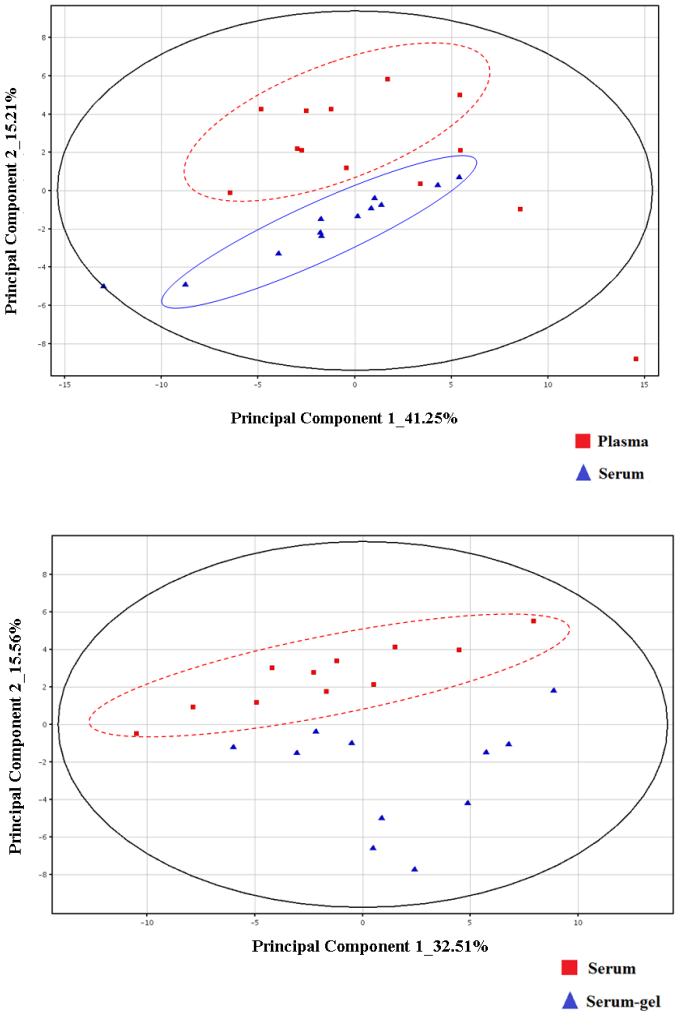


Figure 3. PCA scores plot in a 2D graph obtained from serum and plasma samples (A) and from serum and serum-gel samples (B).

A Mann-Whitney *t*-test corrected by the Benjamini-Hochberg algorithm was performed to identify the metabolites with the highest contribution to differentiate serum from serum-gel samples. The results showed that eleven metabolites were found at significant different concentration in serum samples collected with both types of tubes, with p-value<0.05. Table 3 lists these metabolites together with the main

pathways in which they are involved. These significant metabolites were present at higher relative concentration in serum than in serum-gel samples, except for monopalmitin and dodecanol that reported the opposite situation. The most representative case corresponded to aconitic acid that was hardly detected in serum collected in polymeric gel tubes, thus indicating a high activity of the enzymes acting on this metabolite, a substrate of the citric acid cycle, as shows Supplementary Figure 3B. Other metabolites involved in the citric acid cycle were related to pathways in which other significant metabolites detected in this study play a key role. Thus, acetyl-CoA is related to the fatty acids metabolism in which a critical pathway is the synthesis of glycerolipids. The differences in concentration of glycerol and two monoacylglycerols (monostearin and monopalmitin) resulted significant in the comparative analysis between serum obtained with conventional and polymeric gel tubes. Other two carboxylic acids such as oxalacetic acid and fumaric acid allow connecting the citric acid cycle with two pathways dealing with the metabolism of the amino acids alanine and proline, which were also significantly different in the two types of serum samples. Similarly, pyruvate, which justifies the connection between the citric acid cycle and lactate through the glycolysis, is connected to the threonine metabolism that could explain the different level of this amino acid in serum collected with the two tubes.

One other metabolite that provided significant differences between serum obtained with conventional tubes and polymeric gel tubes was aminomalonic acid, a dicarboxylic acid with one methylene group less than aspartic acid. Some studies have tentatively identified the origin of aminomalonic acid by elimination of sulfur containing cysteine [24]. The conversion of cysteine into aminomalonate could induce conformational changes important for ubiquitinylation and degradation of iron regulatory protein 2 (IRP2) [24]. Finally, dodecanol and the cholesterol derivative, which reported significant differences between serum and plasma samples, also provided levels statistically different in serum and serum-gel samples.

Concerning plasma samples, no overall differences were found between plasma obtained with conventional tubes and polymeric gel tubes, as deduced from the PCA test. Nevertheless, the Mann-Whitney *t*-test allowed identifying five metabolites (four

amino acids: alanine, valine, tyrosine and glutamic acid, and phosphate) with different concentration in the two types of plasma samples.

4. Conclusions

Blood collection is a frequently forgotten aspect that can be responsible for uncontrolled variability sources in metabolomics analysis. Most of the studies dealing with the influence of blood collection tubes on metabolomics profiling have been restricted to conventional tubes for plasma and serum. However, the behavior of polymeric gel tubes had not been comparatively assessed. In this research, serum and plasma collected in polymeric gel tubes were compared with serum and plasma collected in conventional tubes using a GC–TOF/MS untargeted approach. Significant changes attributable to the polymeric gel were only detected in serum, while no differences were observed in plasma, which in overall terms provided a metabolite profile similar to that of plasma collected in conventional tubes. Changes occurring in serum were mainly found in the metabolism of amino acids, particularly, alanine, proline and threonine; in the metabolism of glycerolipids, detected through changes in the levels of glycerol and two important monoglycerides such as monopalmitin and monostearin, and in two metabolites (aconitic acid and lactic acid) involved in primary pathways. Although no overall changes were observed in plasma, significant differences in the concentration of five metabolites, four amino acids and phosphate, were found. It is essential to take into account these alterations when an experimental protocol for metabolomics analysis is planned.

An additional issue was to evaluate the metabolite differences between serum and plasma collected from the same group of individuals in conventional tubes. These differences affected to critical pathways such as the citric acid cycle, the metabolism of amino acids, the fructose and mannose metabolism and that of glycerolipids, and pentose/glucuronate interconversions.

Akcnowledgments

The Spanish Ministerio de Economía y Competitividad (MINECO) and European FEDER program are thanked for financial support through projects CTQ2012-37428 and, particularly, "Development of methods for early cancer detection, December 29, 2011", that also covered A. Peralbo-Molina and M. Calderón-Santiago research contracts. F.P.C. is also grateful to Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal contract (RYC-2009-03921).

References

[1] Z. Yu, G. Kastenmuller, Y. He, P. Belcredi, G. Moller, C. Prehn, J. Mendes, Simone Wahl1, W. Roemisch-Marg, U. Ceglarek, A. Polonikov, N. Dahmen, H. Prokisch, L. Xie, Y. Li, H.E. Wichmann, A. Peters, F. Kronenberg, K. Suhre, J. Adamski, T. Illig, R. Wang-Sattler, Differences between human plasma and serum metabolite profiles, PLoS ONE 6 (2011) e21230.

[2] K.A. Lawton, A. Berger, M. Mitchell, K.E. Milgram, A.M. Evans, L. Guo, R.W. Hanson, S.C. Kalhan, J.A. Ryals, M.V. Milburn. Analysis of the adult human plasma metabolome, Pharmacogenomics 9 (2008) 383–397.

[3] R. Lundblad. Considerations for the use of blood plasma and serum for proteomic analysis, The Internet Journal of Genomics and Proteomics 1 (2003) 1–8.

[4] Linsheng Liu, Jiye A, Guangji Wang, Bei Yan, Ying Zhang, Xinwen Wanga, Chunyan Zhao, Bei Cao, Jian Shi, Mengjie Li, Tian Zheng, Yuanting Zheng, Gang Hao, Fang Zhou, Jianguo Sun, Zimei Wub, Differences in metabolite profile between blood plasma and serum, Anal. Biochem. 406 (2010) 105–112.

[5] Y. Yanga, C. Cruickshanka, M. Armstronga, S. Mahaffeyb, R. Reisdorpha, N. Reisdorpha, New sample preparation approach for mass spectrometry-based profiling of plasma results in improved coverage of metabolome, J. Chromatogr. A 1300 (2013) 217– 226.

[6] N. Psychogios, D.D. Hau, J. Peng, A. Chi Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, The human serum metabolome, PLoS ONE 6 (2011) 2:e16957.

[7] R. Madsena, T. Lundstedtc, J. Trygga, Chemometrics in metabolomics—A review in human disease diagnosis, Anal. Chim. Acta 659 (2010) 23–33.

[8] A. Zhang, H. Sun, X. Wang, Serum metabolomics as a novel diagnostic approach for disease: a systematic review, Anal. Bioanal. Chem. 404 (2012) 1239–1245.

[9] B.T. Doumas, L.L Hause, D.M. Slmuncak, D. Breltsnteld, Differences between values for plasma and serum in tests performed in the Ektachem 700 XR Analyzer, and evaluationof "Plasma Separator Tubes (PST)", Clin. Chem. 35 (1989) 151–153.

[10] K. Dettmer-Martin, F. Almstetter1-Inka J. Appel, N. Nurnberger, G. Schlamberger, W. Gronwald, H.H.D. Meyer, P.J. Oefner, Comparison of serum versus plasma collection in gas chromatography – Mass spectrometry-based metabolomics, Electrophoresis 31 (2010) 2365–2373.

[11] T. Barri, L.O. Dragsted. UPLC-ESI-QTOF/MS and multivariate data analysis for blood plasma and serum metabolomics: Effect of experimental artefacts and anticoagulant, Anal. Chim. Acta 768 (2013) 118–128.

[12] M.J. Evans, J.H. Livesey, M.J. Ellis, T. G. Yandle, Effect of anticoagulants and storage temperatures on stability of plasma and serum hormones, Clin. Biochem. 34 (2001) 107–112.

[13] M.B.F. Kohek, C.R.M. Leme1, I.T. Nakamura, S.A. de Oliveira, V. Lando, B.B. Mendonca, Effects of EDTA and sodium citrate on hormone measurements by fluorometric (FIA) and immunofluorometric (IFMA) methods, BMC Clinical Pathology 2 (2002) 2.

[14] J.R. Denery, A.A.K. Nunes, T.J. Dickerson, Characterization of differences between blood sample matrices in untargeted metabolomics, Anal. Chem. 83 (2011) 1040–1047.

[15] D.C. Wedge, J.W. Allwood, W. Dunn, A.A. Vaughan, K. Simpson, M. Brown, L. Priest, F.H. Blackhall, A.D. Whetton, C. Dive, R. Goodacre, Is serum or plasma more appropriate for intersubject comparisons in metabolomic studies? An assessment in patients with small-cell lung cancer, Anal. Chem. 83 (2011) 6689–6697.

[16] S. Cuhadar, A. Atay, M. Koseogl, A. Dirican, A. Hur, Stability studies of common biochemical analytes in serum separator tubes with or without gel barrier subjected to various storage conditions, Biochemia Medica 22(2) (2012) 202–14.

[17] R.A.R. Bowen, G. L. Hortin, G. Csako, O.H. Otañez, A.T. Remaley, Impact of blood collection devices on clinical chemistry assays, Clin. Biochem. 43 (2010) 4–25.

[18] M. Breier, S. Wahl, C. Prehn, M. Fugman, U. Ferrari, M. Weise, F. Banning, J. Seissler, H. Grallert, J. A.A. Lechner, Targeted metabolomics identifies reliable and stable metabolites in human serum and plasma samples, PLoS ONE 9 (2014) 2:e89728.

[19] B. Jørgenrud, S. Jäntti, I. Mattila, P. Pöhö, K. S. Rønningen, H. Yki-Järvinen, M. Orešič, T. Hyötyläinen, The influence of sample collection methodology and sample preprocessing on the blood metabolic profile, Bioanalysis 7 (2015) 991–1006.

[20] J. Karppi, K.K. Akerman, M. Parviainen, Suitability of collection tubes with separator gels for collecting and storing blood samples for therapeutic drug monitoring (TDM), Clin. Chem. Lab. Med. 38 (2000) 313–20.

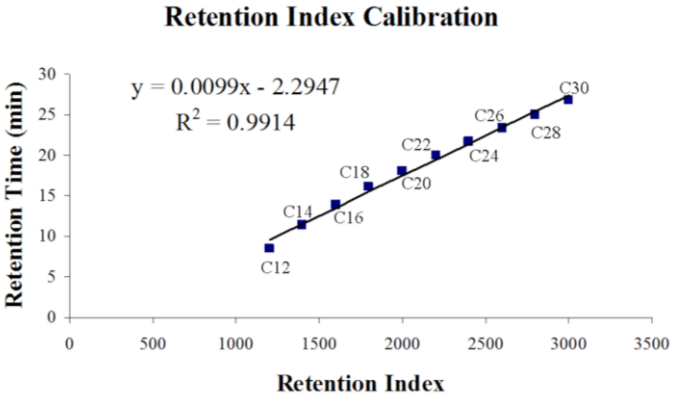
[21] Off. J. World Med. Assoc. 50 (2004) 50–51.

[22] A. Jiye, J. Trygg, J. Gullberg, A.I. Johansson, P. Jonsson, H. Antti, S.L. Marklund, T. Moritz, Extraction and GC/MS analysis of the human blood plasma metabolome, Anal. Chem. 77 (2005) 8086–8094.

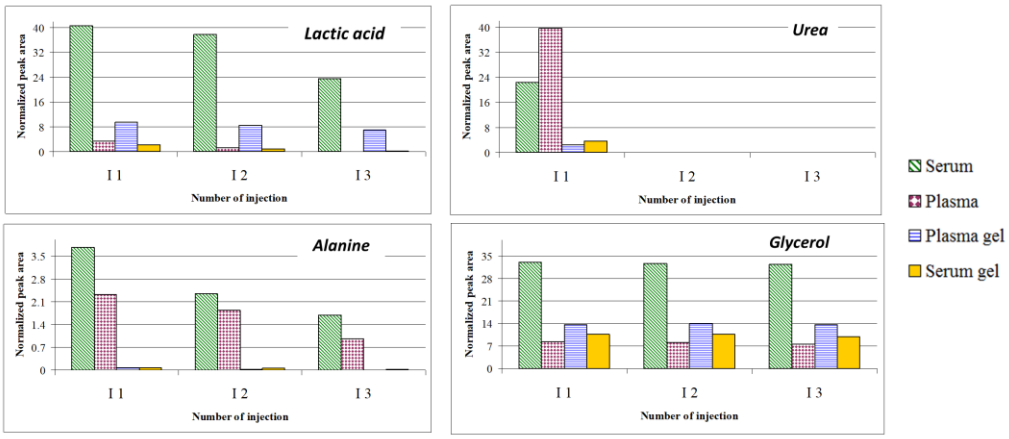
[23] M.P. O'Keane, S.K. Cunningham, Evaluation of three different specimen types (serum, plasma lithium heparin and serum gel separator) for analysis of certain analytes: clinical significance of differences in results and efficiency in use, Clin. Chem. Lab. Med. 44 (2006) 662–668.

[24] R. Xue1, Z. Lin, C. Deng, L. Dong, T. Liu, J. Wang, X. Shen, A serum metabolomic investigation on hepatocellular carcinoma patients by chemical derivatization followed by gas chromatography/mass spectrometry, Rapid Commun. Mass Spectrom. 22 (2008) 3061–3068.

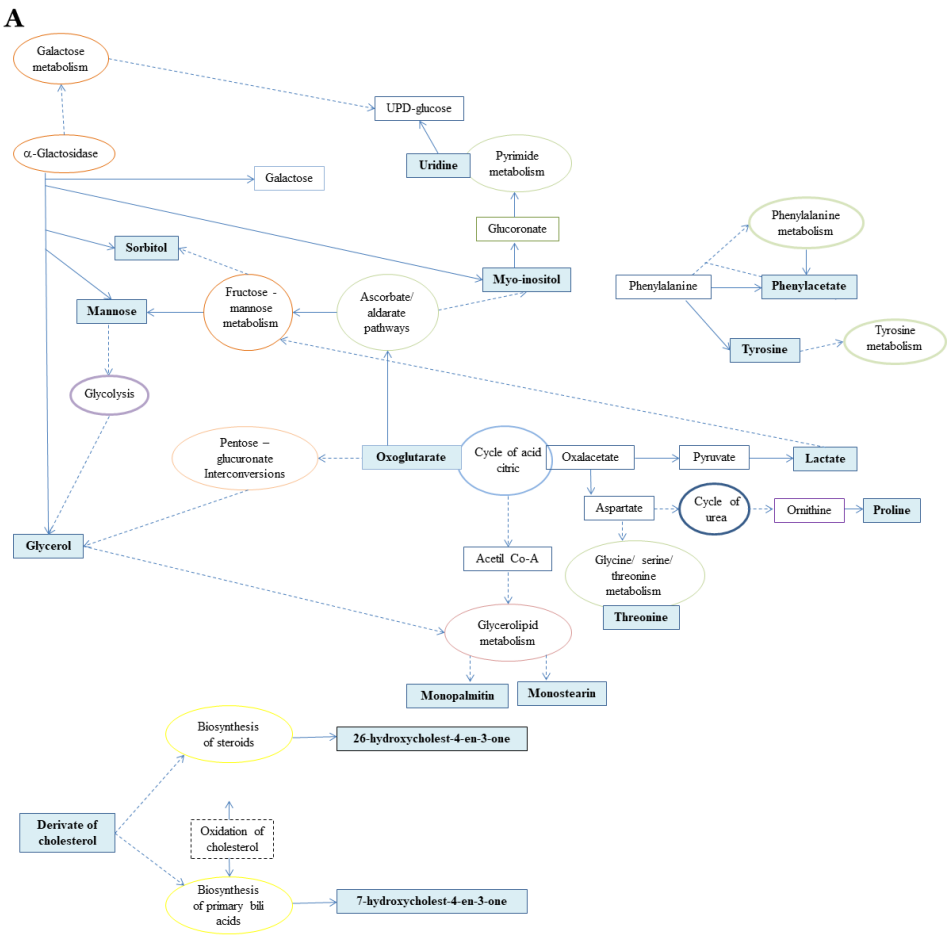
Supplementary Information



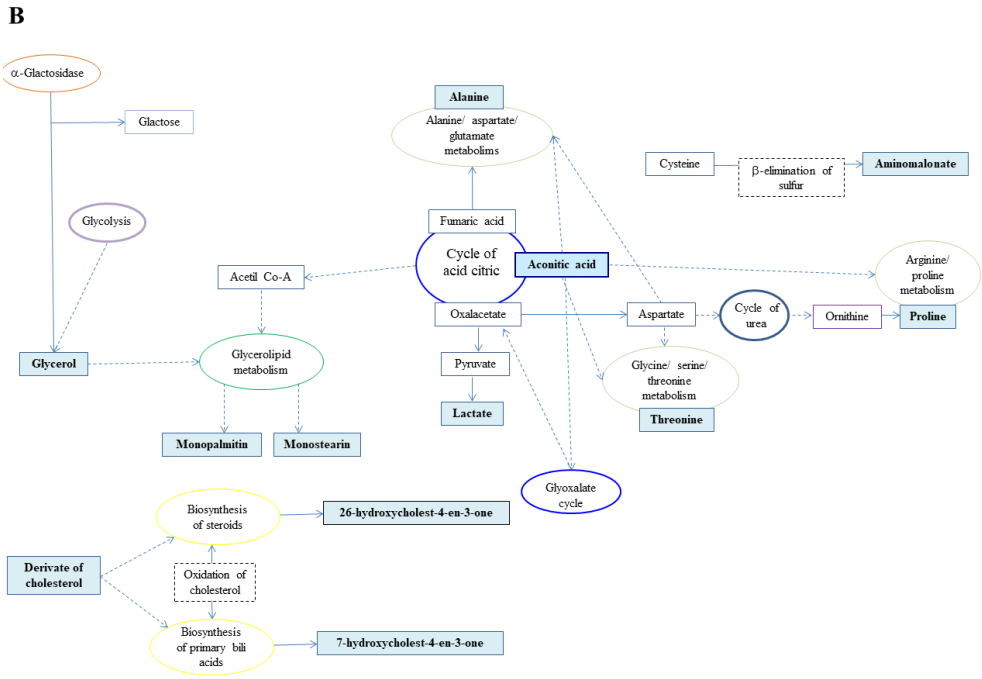
Supplementary Figure 1. RI calibration graph and equation fitting this model.



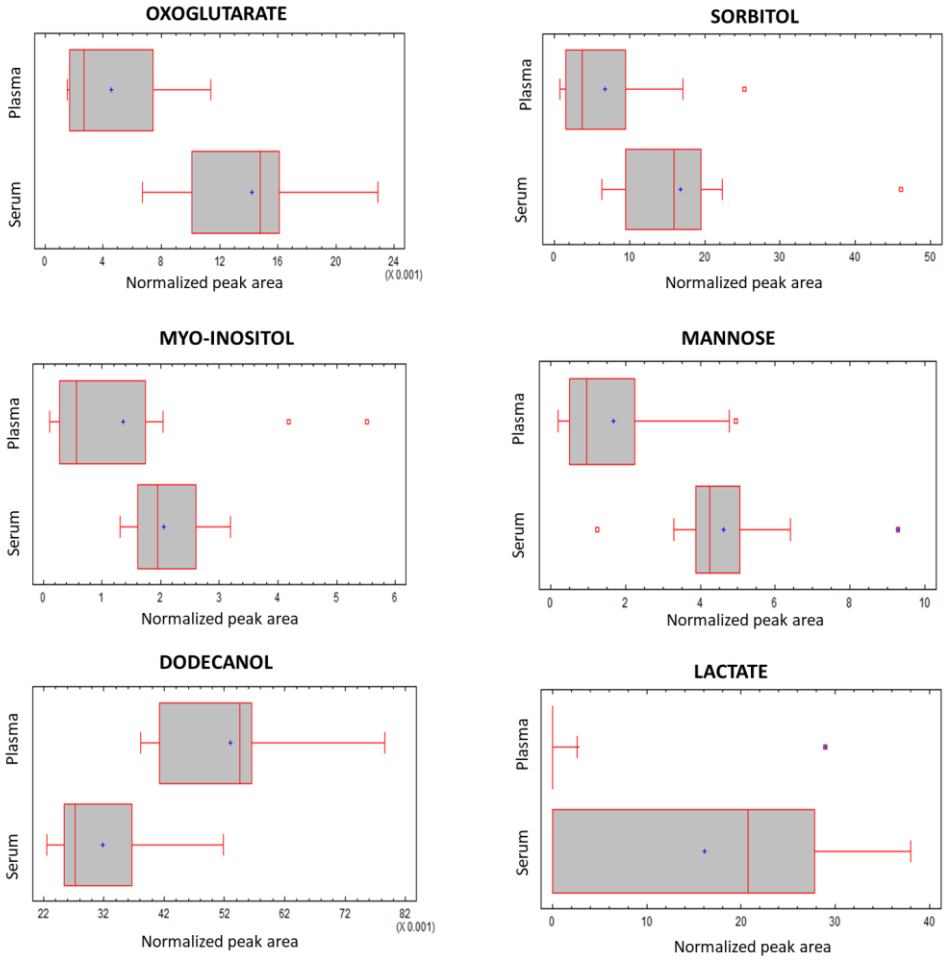
Supplementary Figure 2. Quantitative response of lactic acid, urea, alanine and glycerol in three consecutive injections of one sample of serum, plasma, serum-gel and plasma gel.



Supplementary Figure 3A. Scheme of the connections among the most important pathways including metabolites discriminating significantly: serum and plasma. Significant metabolites are in bold. The solid lines indicate metabolites conversions, and the dotted lines indicate metabolites involved in a particular pathway.



Supplementary Figure 3B. Scheme of the connections among the most important pathways including metabolites discriminating significantly: serum and serum-gel. Significant metabolites are in bold. The solid lines indicate metabolites conversions, and the dotted lines indicate metabolites involved in a particular pathway.



Supplementary Figure 4. Box-and-whisker plots provided by the comparison between serum *vs*. plasma of the normalized concentration of six identified metabolites.

Supplementary Table 1. Name, chemical class, HMDB code, retention time, derivatized formula and derivatized fragments for each identified metabolite in the four types of samples (serum, serum-gel, plasma and plasma-gel).

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Monopalmitin	Acylglycerol	HMDB31074	23.09	$C_{25}H_{54}O_4Si_2$	459.3341 – $[C_{24}H_{51}O_4Si_2]^+$
					371.2978 – $[C_{21}H_{43}O_3Si]^+$
					313.2553 – $[C_{18}H_{37}O_2Si]^+$
Monostearin	Acylglycerol	HMDB31075	24.26	$C_{27}H_{58}O_4Si_2$	487.3626 – $[C_{26}H_{55}O_4Si_2]^+$
					399.3283 – $[C_{23}H_{47}O_3Si]^+$
					341.2852 – $[C_{20}H_{41}O_2Si]^+$
Leucine	Amino acids	HMDB00687	7.83	$C_{12}H_{29}NO_2Si_2$	260.1489 – $[C_{11}H_{26}NO_2Si_2]^+$
					232.1535 – $[C_{10}H_{26}NOSi_2]^+$
					158.1357 – $[C_8H_{20}NSi]^+$
Proline	Amino acids	HMDB00162	9.89	$C_{11}H_{25}NO_2Si_2$	216.1224 – $[C_9H_{22}NOSi_2]^+$
					170.0620 – $[C_7H_{12}NO_2Si]^+$
					142.1043 – $[C_7H_{16}NSi]^+$
Cysteine	Amino acids	HMDB00574	13.23	$C_{12}H_{31}NO_2SSi_3$	340.1679 - $[C_{12}H_{31}NO_2SSi_3]^+$
					220.0999 - $[C_8H_{20}NO_2Si_2]^+$
					100.0555 - $[C_4H_{10}NSi]^+$
Aminobutyric acid	Amino acids	HMDB00650	8.06	$C_{10}H_{25}NO_2Si_2$	218.1012 – $[C_8H_{20}NO_2Si_2]^+$
					142.0681 – $[C_6H_{12}NOSi]^+$
					114.0715 – $[C_5H_{12}NSi]^+$
Glutamic acid	Amino acids	HMDB00148	13.99	$C_{14}H_{33}NO_4Si_3$	348.1465 –$[C_{13}H_{30}NO_4Si_3]^+$
					246.1333 – $[C_{10}H_{24}NO_2Si_2]^+$
					230.1019 – $[C_9H_{20}NO_2Si_2]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Glycine	Amino acids	HMDB00123	9.99	C11H29NO2Si3	276.1248 – $[C_{10}H_{26}NO_2Si_3]^+$
					248.1303 – $[C_9H_{26}NOSi_3]^+$
					174.1125 – $[C_7H_{20}NSi_2]^+$
Alanine	Amino acids	HMDB00161	7	C9H23NO2Si2	218.1013 – $[C_8H_{20}NO_2Si_2]^+$
					190.1067 – $[C_7H_{20}NOSi_2]^+$
					116.0885 – $[C_5H_{14}Nsi]^+$
Pyroglutamic acid	Amino acids	HMDB00267	12.52	C8H15NO3Si	200.0768 – $[C_8H_{14}NO_3Si]^+$
					186.0553 – $[C_7H_{12}NO_3Si]^+$
					103.0213 – $[C_3H_7O_2Si]^+$
Tyrosine	Amino acids	HMDB00158	17.44	C18H35NO3Si3	382.1682 – $[C_{17}H_{32}NO_3Si_3]^+$
					354.1731 – $[C_{16}H_{32}NO_2Si_3]^+$
					280.1538 – $[C_{14}H_{26}NOSi_2]^+$
Valine	Amino acids	HMDB00883	6.79	C11H27NO2Si2	246.1324 – $[C_{10}H_{24}NO_2Si_2]^+$
					218.1110 – $[C_9H_{22}O_2Si_2]^+$
					203.0809 – $[C_7H_{17}NO_2Si_2]^+$
Serine	Amino acids	HMDB00187	10.73	C12H31NO3Si3	218.1033 – $[C_8H_{20}NO_2Si_2]^+$
					204.1283 – $[C_8H_{22}NOSi_2]^+$
					188.0924 – $[C_7H_{18}NOSi_2]^+$
Threonine	Amino acids	HMDB00167	9.83	C10H25NO3Si2	264.1464 – $[C_{10}H_{25}NO_3Si_2]+$
					219.1090 – $[C_8H_{21}NO_2Si_2]+$
					130.0671 – $[C_5H_{12}NOSi]+$
Phenylalanine	Amino acids	HMDB00159	13.17	C15H27NO2Si2	294.1339 – $[C_{14}H_{24}NO_2Si_2]^+$
					218.1021 – $[C_8H_{20}NO_2Si_2]^+$
					192.1194 – $[C_{11}H_{18}Nsi]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Phenylacetic acid	Amino acids derivatives	HMDB00209	14.07	C16H26O2Si	278.1652 – $[C_{16}H_{26}O_2Si]^+$
					263.1483 - $[C_{15}H_{23}O_2Si]^+$
					204.0982 - $[C_{12}H_{16}OSi]^+$
Benzoic acid	Benzenoids	HMDB01870	9.22	C10H14O2Si	179.0519 – $[C_9H_{11}O_2Si]^+$
					122.0353 – $[C_7H_6O_2]^+$
					105.0330 – $[C_7H_5O]^+$
Sorbitol	Carbohydrates	HMDB00247	16.55	C18H44O5Si4	453.2335– $[C_{18}H_{44}O_5Si_4]^+$
					349.1665 – $[C_{18}H_{29}O_3Si_2]^+$
					217.1084 – $[C_9H_{21}O_2Si_2]^+$
Erythritol	Carbohydrates	HMDB02994	12.63	C16H42O4Si4	366.1292– $[C_{16}H_{30}O_2Si_4]^+$
					321.1533 – $[C_{12}H_{33}O_2Si_4]^+$
					277.1476 – $[C_{11}H_{29}O_2Si_3]^+$
Erythronic acid	Carbohydrates	HMDB00498	13.83	C16H40O5Si4	379.1639– $[C_{14}H_{35}O_4Si_4]^+$
					293.1419 – $[C_{11}H_{29}O_3Si_3]^+$
					217.1084 – $[C_9H_{21}O_2Si_2]^+$
Mannitol	Carbohydrates	HMDB00765	17.3	C24H62O6Si6	520.1592 - $[C_{23}H_{36}O_6Si_4]^+$
					319.1569 – $[C_{13}H_{31}O_3Si_3]^+$
					205.1043 – $[C_{12}H_{17}OSi]+$
Glucose	Carbohydrates	HMDB00122	16.95	C22H55NO6Si5	319.1563 – $[C_{13}H_{31}O_3Si_3]^+$
					205.1065 – $[C_8H_{21}O_2Si_2]^+$
					160.0784 – $[C_6H_{14}NO_2Si]^+$
			17.8	C21H52O6Si5	435.1960 – $[C_{21}H_{39}O_2Si_4]^+$
					305.1449 – $[C_{11}H_{33}Si_5]^+$
					204.1752 – $[C_{10}H_{28}Si_2]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Glycerol	Carbohydrates	HMDB00131	9.54	C12H32O3Si3	293.1401 – $[C_{11}H_{29}O_3Si_3]^+$
					218.1127 – $[C_9H_{22}O_2Si_2]^+$
					205.1071 – $[C_8H_{21}O_2Si_2]^+$
Threonic acid	Carbohydrates	HMDB00943	13.06	C16H40O5Si4	410.1814 – $[C_{15}H_{38}O_5Si_4]^+$
					378.1146 – $[C_{13}H_{30}O_5Si_4]^+$
					319.1201 – $[C_{12}H_{27}O_4Si_3]^+$
Myo-Inositol	Carbohydrates	HMDB00211	18.9	C24H60O6Si6	612.2926 – $[C_{24}H_{60}O_6Si_6]^+$
					305.1403 – $[C_{12}H_{29}O_3Si_3]^+$
					217.1059 – $[C_9H_{21}O_2Si_2]^+$
Glyceric acid	Carbohydrates	HMDB00139	10.11	C12H30O4Si3	307.1189 – $[C_{11}H_{27}O_4Si_3]^+$
					292.1329 – $[C_{11}H_{28}O_3Si_3]^+$
					189.0760 – $[C_7H_{17}O_2Si_2]^+$
Xylitol	Carbohydrates	HMDB02917	15.13	C20H52O5Si5	430.1843– $[C_{18}H_{38}O_4Si_4]^+$
					307.1571 – $[C_{12}H_{31}O_3Si_3]^+$
					217.1072 – $[C_9H_{21}O_2Si_2]^+$
Mannose	Carbohydrates	HMDB00169	16.04	C21H52O6Si5	540.2631- $[C_{21}H_{52}O_6Si_5]+$
					437.2020 – $[C_{17}H_{41}O_5Si_4]^+$
					265.0939 – $[C_{14}H_{26}O_6Si]^+$
			17.07		435.1877– $[C_{21}H_{35}O_4Si_3]^+$
					204.1003 – $[C_8H_{20}O_2Si_2]^+$
					147.0656 – $[C_5H_{15}Osi_2]^+$
Aminomalonic acid	Carboxylic acids	HMDB01147	12.16	C12H29NO4Si3	320.1163 – $[C_{11}H_{26}NO_4Si_3]^+$
					292.1216 – $[C_{10}H_{26}NO_3Si_3]^+$
					218.1028 – $[C_8H_{20}NO_2Si_2]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Aconitic acid	Carboxylic acids	HMDB31193	16.19	C18H40O7Si4	465.1590 – $[C_{17}H_{37}O_7Si_4]^+$
					375.1100 – $[C_{14}H_{27}O_6Si_3]^+$
					273.0971 – $[C_{11}H_{21}O_4Si_2]^+$
Acetic acid	Carboxylic acids	HMDB00042	6.56	C8H20O3Si2	205.0698 – $[C_7H_{17}O_3Si_2]^+$
					177.0749 – $[C_6H_{17}O_2Si_2]^+$
					147.0650 – $[C_6H_{17}OSi_2]^+$
Succinic acid	Carboxylic acids	HMDB00254	10.1	C10H22O4Si2	262.1103 – $[C_{10}H_{22}O_4Si_2]^+$
					172.0564 – $[C_7H_{12}O_3Si]^+$
					147.0643 – $[C_9H_{11}Si]^+$
3-Hydroxybutyric acid	Fatty acids	HMDB00011	7.88	C10H24O3Si2	249.1321 – $[C_{10}H_{25}O_3Si_2]^+$
					233.1009 – $[C_9H_{21}O_3Si_2]^+$
					191.0922 – $[C_7H_{19}O_2Si_2]^+$
Arachidonic acid	Fatty acids	HMDB01043	21.37	C23H40O2Si	327.0277 – $[C_{23}H_7Osi]^+$
					262.1653 – $[C_{20}H_{22}]^+$
					91.0535 – $[C_7H_7]^+$
Decanoic acid	Fatty acids	HMDB00511	12.03	C13H28O2Si	229.1606 – $[C_{12}H_{25}O_2Si]^+$
					185.0990 – $[C_9H_{17}O_2Si]^+$
					131.0519 – $[C_5H_{11}O_2Si]^+$
Dodecanoic acid	Fatty acids	HMDB00638	14.41	C15H32O2Si	272.2159 – $[C_{15}H_{32}O_2Si]^+$
					257.1921 – $[C_{14}H_{29}O_2Si]^+$
					145.0665 – $[C_6H_{13}O_2Si]^+$
Heptadecanoic acid	Fatty acids	HMDB02259	19.42	C20H42O2Si	342.2974 – $[C_{20}H_{42}O_2Si]^+$
					299.2385 – $[C_{17}H_{35}O_2Si]^+$
					117.0378 – $[C_4H_9O_2Si]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Palmitic acid	Fatty acids	HMDB00220	18.5	C19H40O2Si	328.2762 – $[C_{19}H_{40}O_2Si]^+$
					313.2541 – $[C_{18}H_{37}O_2Si]^+$
					285.2229 – $[C_{18}H_{37}O_2Si]^+$
Hexanoic acid	Fatty acids	HMDB00535	6.53	C9H20O2Si	173.0991 – $[C_8H_{17}O_2Si]^+$
					159.0832 – $[C_7H_{15}O_2Si]^+$
					131.0521 – $[C_5H_{11}O_2Si]^+$
Myristic acid	Fatty acids	HMDB00806	21.57	C23H50O4Si2	408.2930– $[C_{23}H_{44}O_2Si_2]^+$
					258.0952 – $[C_{14}H_{18}OSi_2]^+$
					147.0650 – $[C_5H_{15}Osi_2]^+$
Nonanoic acid	Fatty acids	HMDB00847	10.84	C12H26O2Si	215.1465 – $[C_{11}H_{23}O_2Si]^+$
					145.0670 – $[C_6H_{13}O_2Si]^+$
					117.0370 – $[C_4H_9O_2Si]^+$
Octadecanoic acid	Fatty acids	HMDB00827	20.3	C21H44O2Si	356.3080 – $[C_{21}H_{44}O_2Si]^+$
					341.2854 – $[C_{20}H_{41}O_2Si]^+$
					313.2534 – $[C_{18}H_{37}O_2Si]^+$
Octanoic acid	Fatty acids	HMDB31291	9.42	C11H24O2Si	215.1458 – $[C_{11}H_{23}O_2Si]^+$
					157.0679 – $[C_7H_{13}O_2Si]^+$
					117.0367 – $[C_4H_9O_2Si]^+$
Oleic acid	Fatty acids	HMDB00207	20.07	C21H42O2Si	354.2942– $[C_{21}H_{42}O_2Si]^+$
					339.2713– $[C_{20}H_{39}O_2Si]^+$
					185.0990– $[C_9H_{17}O_2Si]^+$
9,12-Octadecadienoic acid	Fatty acids	HMDB00673	20.02	C21H40O2Si	353.2832– $[C_{21}H_{40}O_2Si]^+$
					337.2557 – $[C_{20}H_{37}O_2Si]^+$
					309.2208 – $[C_{18}H_{33}O_2Si]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Hydroxybutyric acid	Fatty acids	HMDB00008	7.23	C10H24O3Si2	245.0974 - [C10H21O3Si2]+
					205.1071 - [C8H21O2Si2]+
					131.0880 - [C6H15OSi]+
Dodecanol	Fatty alcohols	HMDB11626	13.43	C15H34OSi	243.2138 – $[C_{14}H_{31}OSi]^+$
					183.1733 – $[C_{12}H_{23}O]^+$
					55.0544 – $[C_4H_7]^+$
Phosphoric acid	Homogenous non-metal	HMDB02142	8.11	C9H27O4PSi3	314.0950 – $[C_9H_{27}O_4PSi_3]^+$
					299.0722 – $[C_8H_{24}O_4PSi_3]^+$
					283.0427 – $[C_7H_{20}O_4PSi_3]^+$
2,3-Dihydroxybutanoic acid	Hydroxy acids	HMDB00498	10.5	C13H32O4Si3	321.1365 – $[C_{12}H_{29}O_4Si_3]^+$
					203.0927 – $[C_8H_{19}O_2Si_2]^+$
					117.0739 – $[C_5H_{13}OSi]^+$
2-Ethyl-3-hydroxypropionic acid	Hydroxy acids	HMDB00396	8.87	C10H24O3Si2	231.1064 – $[C_9H_{19}O_3Si_2]^+$
					149.0692 – $[C_5H_{13}O_3Si_2]^+$
					93.9885 – $[C_4H_2OSi]^+$
Lactic acid	Hydroxy acids	HMDB00190	6.31	C9H22O3Si2	233.1064 – $[C_9H_{21}O_3Si_2]^+$
					219.0923 – $[C_8H_{19}O_3Si_2]^+$
					203.0604 – $[C_7H_{15}O_3Si_2]^+$
11-Octadecenoic acid	Hydroxyacil acids	HMDB03231	20.13	C21H42O2Si	355.2340 - [C21H42O2Si]+
					339.2734 – $[C_{20}H_{39}O_2Si]^+$
					311.2393 – $[C_{18}H_{35}O_2Si]^+$
3-Methyl-2-oxovaleric acid	Keto acids	HMDB00491	7.13	C9H18O3Si	193.0299 – $[C_9H_9O_3Si]^+$
					174.1045 – $[C_8H_{18}O_2Si]^+$
					143.0876 – $[C_7H_{15}OSi]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Oxoglutarate	Keto acids	HMDB00208	13.83	C14H30O5Si3	347.1164 – $[C_{13}H_{27}O_5Si_3]^+$
					291.1262 – $[C_{11}H_{27}O_3Si_3]^+$
					113.0431 – $[C_5H_9OSi]^+$
2-Ketoisocaproic acid	Keto acids	HMDB00695	7.18	C9H18O3Si	169.0699 – $[C_8H_{13}O_2Si]^+$
					157.0689 – $[C_7H_{13}O_2Si]^+$
					143.0882 – $[C_7H_{15}Osi]^+$
Acetoacetic acid	Keto acids	HMDB00060	8.56	C10H22O3Si2	246.1093 – $[C_{10}H_{22}O_3Si_2]^+$
					231.0901 – $[C_9H_{19}O_3Si_2]^+$
					157.0695 – $[C_7H_{13}O_2Si]^+$
Urea	Nitrogen compounds	HMDB00294	9.03	C7H20N2Osi2	204.1110 – $[C_7H_{20}N_2Osi_2]^+$
					189.0885 – $[C_6H_{17}N_2Osi_2]^+$
					171.0781 – $[C_6H_{15}N_2Si_2]^+$
Uridine	Nucleosides	HMDB00296	21.99	C18H36N2O6Si3	445.1653 – $[C_{17}H_{33}N_2O_6Si_3]^+$
					341.0185 – $[C_{18}H_9N_2O_2Si_2]^+$
					243.0850 – $[C_{14}H_{15}O_2Si]+$
Dehydroabietic acid	Prenol lipids	HMDB61925	21.6	C23H36O2Si	372.2485 – $[C_{23}H_{36}O_2Si]^+$
					357.2254 – $[C_{22}H_{33}O_2Si]^+$
					239.1805 – $[C_{18}H_{23}]^+$
Limonene	Prenol lipids	HMDB03375	5.95	C10H16	136.1227 – $[C_{10}H_{16}]^+$
					121.0992 – $[C_9H_{13}]^+$
					79.0524 – $[C_6H_7]^+$
Squalene	Prenol lipids	HMDB00256	24.8	C30H50	367.3350 – $[C_{27}H_{43}]^+$
					81.0691 – $[C_6H_9]^+$
					69.0691 – $[C_5H_9]^+$

Continuation Supplementary Table 1

Compound name	Chemical class	HMDB code	Retention time (min)	Derivatized formula	Derivatized fragments/ions
Cholesterol derivate: 26-hydroxycholest-4-en-one	Steroids	HMDB12459	25.5	C27H44O3	368.3433 - $[C_{23}H_{28}O_4]^+$ 131.0666 - $[C_6H_{11}O_3]^+$ 81.0690 - $[C_9H_9]^+$
Cholesterol derivate: 7-hidroxycholest-4-en-3-one	Steroids	HMDB12458		C27H42O4	
Ascorbic acid	Vitamins	HMDB00044	17.53	C18H40O6Si4	464.1910 – $[C_{18}H_{40}O_6Si_4]^+$
					332.1301 - $[C_{13}H_{28}O_4Si_3]^+$
					117.0362 - $[C_4H_9OSi]^+$

CHAPTER IV

Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue

Talanta

177 (2018) 86–93

Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue

María Asunción López-Bascón[1,2,3,4], Mónica Calderón-Santiago[1,2,3,4], Julia Sánchez-Ceinos[3,5], Alejandro Fernández-Vega[3,5], Rocío Guzmán-Ruiz[3,5] José López-Miranda[2,3,6], María del Mar Malagon[2,3,5], Feliciano Priego-Capote[1,2,3,4]*

[1]*Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain*

[2]*CeiA3 Agroalimentary Excellence Campus, University of Córdoba, Córdoba, Spain*

[3]*Maimónides Institute for Biomedical Research (IMIBIC)/University of Córdoba/Reina Sofía University Hospital, Córdoba, Spain*

[4]*CIBER Fragilidad y Envejecimiento Saludable (CIBERfes), Instituto de Salud Carlos III, Spain*

[5]*Department of Cell Biology, Physiology and Immunology, University of Córdoba, Córdoba, Spain.*

[6]*Lipids and Atherosclerosis Unit, Department of Medicine, University of Córdoba, Reina Sofía Hospital, Córdoba, Spain*

Influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue

María Asunción López-Bascón, Mónica Calderón-Santiago, Julia Sánchez-Ceinos, Alejandro Fernández-Vega, Rocío Guzmán-Ruiz, José López-Miranda, María del Mar Malagón, Feliciano Priego-Capote

ABSTRACT

The main limitations of lipidomics analysis are the chemical complexity of the lipids, the range of concentrations at which they exist, and the variety of samples usually analyzed. These limitations particularly affect the characterization of polar lipids owing to the interference of neutral lipids, essentially acylglycerides, which are at high concentration and suppress ionization of low concentrated lipids in mass spectrometry detection. The influence of sample preparation on lipidomics analysis of polar lipids in adipose tissue by LC–MS/MS was the aim of this research. Two common extractants used for lipids isolation, methanol:chloroform ($MeOH:CHCl_3$) and methyl *tert*-butyl ether (MTBE), were qualitatively and quantitatively compared for the extraction of the main families of lipids. The obtained results showed that each family of lipids is influenced differently by the extractant used. However, in general, the use of MTBE as extractant led to higher extraction efficiency for unsaturated fatty acids, glycerophospholipids and ceramides, while $MeOH:CHCl_3$ favored the isolation of saturated fatty acids and plasmalogens. The implementation of a solid-phase extraction (SPE) step for selective isolation of glycerophospholipids prior to LC–MS/MS analysis was assayed to evaluate its influence on lipids detection coverage as compared to direct analysis. This step was critical to enhance the detection coverage of glycerophospholipids by removal of ionization suppression effects caused by acylglycerides.

Keywords: Adipose tissue, lipidomics, sample preparation, liquid chromatography, mass spectrometry, extraction

1. Introduction

Lipids constitute a wide variety of biological molecules involved in essential functions due to their role as structural components of cell membranes, energy storage, and intermediates in signaling pathways [1–4]. For this reason, lipidomics, a branch of metabolomics devoted to the qualitative/quantitative analysis of the lipidome, has experienced a drastic expansion [3]. Because of their biological importance, lipids are under tight homeostatic control and exhibit spatial and dynamic complexity at multiple levels [5,6]. Thus, it is not surprising that altered lipid metabolism plays a key role in the pathogenesis of common diseases.

Adipose tissue is a complex, essential, and highly active metabolic and endocrine organ [2,7,8] whose lipid composition can be classified into two main groups: the neutral or non-polar lipids and the polar lipids. Neutral lipids, composed by acylglycerides (tri-, di- and monoacylglycerides), cholesteryl esters and cholesterol, constitute the most abundant group, while polar lipids with specific functional groups encompass essentially free fatty acids, the different families of glycerophospholipids and ceramides [8]. Adiposity, which represents the fraction of total body mass composed by neutral lipids, is closely related with key physiological parameters such as blood pressure, systemic insulin sensitivity, and concentration of molecules such as serum triglycerides and leptin [2,9,10]. Thereby, the excess of adipose tissue or obesity, particularly in the visceral compartment, is associated with insulin resistance, hyperglycemia, dyslipidemia, hypertension, and prothrombotic and proinflammatory states [2,7,11]. On the other hand, polar lipids are the main components of biological membranes, also in adipose tissue, and regulate cellular signalling to facilitate the transmission of biological information across them [12,13]. Additionally, polar lipids act as bioactive mediators that have been recognized as endogenous regulators of key cellular processes. Most of these bioactive mediators originate from the cleavage of lipid constituents of cellular membranes under the activity of phospholipases. Adipose tissue function, which has a crucial role in the development of obesity-related comorbidities including insulin resistance and non-alcoholic fatty liver disease, is dysregulated in obese individuals with a key function of polar lipids [14]. According to these

associations, it is important to study the composition of adipose tissue under different biological circumstances or conditions.

Main lipid species in human serum, liver and adipose tissue have already been characterized in several studies [15]. For this purpose, LC–MS/MS has proved a noticeable detection capability for analysis of the different families of lipids [8,9,16–20], except for free fatty acids that are better characterized by GC–MS after derivatization [21–24]. Nevertheless, LC–MS/MS is able to offer quantitative response for the most concentrated fatty acids found in biological samples. Concerning sample preparation, this critical step determines the type and concentration of isolated lipids and, for this reason, several preparation methods have been applied to biological samples with the goal of improving overall lipid coverage. The methods have been based on sample preparation techniques such liquid–liquid extraction, precipitation using organic solvents, and solid-phase extraction (SPE). Folch *et al.* [25] and Bligh–Dyer [26] proposed methods for general extraction of lipids based on methanol (MeOH) and chloroform ($CHCl_3$) mixtures as extractants. These two methods differ in the extractants ratio —concretely (1:2, v/v) or (2:1, v/v) MeOH:$CHCl_3$ for the Folch and Bligh–Dyer methods, respectively—, and the extractant–sample ratio. Specifically, the Folch method employs a higher extractant–sample ratio (roughly a 20-fold excess of extractant) as compared to the Bligh–Dyer protocol. According to the literature, the use of a MeOH:$CHCl_3$ mixture for lipids extraction ensures the isolation of all major lipid classes, which are mostly enriched in the chloroform phase [27,28]. Later, Matyash *et al.* introduced a novel extraction procedure for isolation of lipids using as extractant a MeOH and methyl *tert*-butyl ether (MTBE) mixture in a 1.5:5 (v/v*)* ratio, then adding water to improve the separation step. This methodology avoids the use of a toxic and carcinogenic extractant as chloroform, thus reducing the environmental burden as well as the health risks for the exposed personnel [29]. This protocol, tested in four biological matrices, provided recoveries similar to or even better than either the Folch or Bligh–Dyer methods for major lipid classes [29]. In any case, these overall extraction protocols are characterized by a dual limitation that affect in a different way to the two main lipid families in adipose tissue. Thus, neutral lipids are not quantitatively extracted with these protocols in complex samples such as adipose tissue, thus limitating the

identification of neutral lipids to the most concentrated species. On the other hand, the detection of polar lipids, less concentrated, is critically affected by the presence of neutral lipids that considerably exert ionization suppression effects. For this reason, Baker *et al.* proposed a dual extraction step for independent extraction of neutral lipids and polar lipids [8]. The first family was extracted by an isooctane:ethyl acetate solution using a protocol based on the work from Hutching *et al.*, which allows separating approximately 90% of neutral lipids from polar lipids [30]. The latter were isolated by a modified Bligh-Dyer method with further clean-up of co-extracted neutral lipids by SPE with silica gel as sorbent [30]. A limitation of this protocol is associated to the concentration of acetic acid used to favor the isolation of polar lipids. This concentration could promote degradation of certain families of minor polar lipids such as plasmalogens [8].

Considering the limitations in the analysis of the different families of lipids in adipose tissue, which specially affect polar lipids, this research was planned to evaluate the influence of sample preparation on the determination of polar lipids in visceral adipose tissue. Two different extractants, MeOH:$CHCl_3$ and MTBE, were tested to compare their efficiency for the extraction of polar lipids but also their inefficiency for extraction of acylglycerides: the main interferents in the detection of polar lipids. Additionally, the implementation of an SPE step with a selective sorbent for retention of glycerophospholipids was assessed to check its influence on the subsequent detection of this family of lipids.

2. Materials and methods

2.1. Chemicals and reagents

Chromatographic mobile phase B was prepared using LC–MS grade acetonitrile (ACN) and 2-propanol (IPA) from Sigma–Aldrich (Madrid, Spain). MS-grade formic acid from Scharlab (Barcelona, Spain) and ammonium acetate from Sigma–Aldrich were used as ionization agents for LC–MS/MS analysis. Deionized water (18 MΩ·cm) supplied by a Milli-Q water purification system from Millipore (Bedford, MA, USA) was

used to prepare the chromatographic aqueous phase (phase A). Chromatographic grade $CHCl_3$, MeOH, and MTBE from Scharlab were used for sample preparation. The buffer employed in this research for tissue homogenization was Dulbecco's phosphate buffered saline (D-PBS) from Lonza (Basilea, Switzerland).

For identification of free fatty acids (FAs) the injection of commercial standards of these compounds in MeOH was necessary owing to the scant MS/MS information generated by fragmentation. The following commercial standards of FAs were acquired from Fluka Analytical (Buches, Switzerland): lauric (C12:0), myristic (C14:0), palmitic (C16:0), palmitoleic (C16:1), hepatadecanoic (C17:0), stearic (C18:0), oleic (C18:1), linoleic acid (C18:2), linolenic (C18:3), eicosanoic (C20:0), eicosadienoic (C20:2n6), eicosatrienoic (C20:3n6), eicosatetraenoic (C20:4), arachidonic (C20:4), behenic (C22:0), docosatetraenoic (C22:4), docosapentanoic (C22:5), docosahexaenoic (C22:6) and tetracosenoic (C24:0) acids.

2.2. Adipose tissue samples

100 mg of visceral adipose tissue were obtained from a pool of individuals with morbid obesity undergoing bariatric surgery and recruited by the Lipids and Atherosclerosis Unit of the Reina Sofia University Hospital (Cordoba, Spain). The subjects who participated in this study gave written informed consent. The study was conducted according to the Declaration of Helsinki [31] and was approved by the Reina Sofia Hospital Biomedical Research and University of Cordoba Biomedical Experimentation Ethical Committees.

2.3. Apparatus and instruments

A Teflon homogenizer (Kimble Kontes Vineland, NJ, USA) was used for homogenization of adipose tissue samples. A Visiprep™ SPE vacuum manifold (Supelco, PA, USA) with disposable liners (Supelco, PA, USA) was used for enrichment of glycerophospholipids isolated from adipose tissue. A Concentrator Plus speed-vac from Eppendorf (Hamburg, Germany) was used to evaporate the MeOH phase after SPE elution to concentrate the sample, and a vortex shaker from IKA (Wilmington, NC, USA) was used for sample agitation.

An Agilent 1200 Series LC system coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer (Santa Clara, CA, USA) was used for analysis of the extracts. The QTOF detector was equipped with a dual electrospray ionization source for simultaneous spraying of chromatographic eluate and a reference solution to calibrate continuously the detected *m/z* ratios. Chromatographic eluates were monitored by tandem mass spectrometry in high resolution mode. The Agilent MassHunter Workstation software was used to control the instrument and acquire the data.

2.4. Extraction of the lipid fraction

100-mg aliquots of visceral adipose tissue were taken for homogenization with 250 µL of D-PBS, as can be seen in Supplementary Figure 1. Tissue homogenization was performed on ice with a Teflon homogenizer (2 pulses of 25 s) and later, the samples were sonicated for 8 cycles of 30 s on ice for tissue disaggregation. Subsequently, the analytical sample was heated at 35 °C for 15 min to promote solubilization of the fatty material in the buffer and then chilled on ice for 30 min. Finally, the sample was centrifuged at 16000 × *g* for 15 min at 4 °C and the upper phase, which contains the dissolved lipids, was collected. This phase was divided into two fractions to test the two different methodologies for extraction of polar lipids.

2.4.1. Extraction with 1:1 (v/v) MeOH:$CHCl_3$

This protocol was carried out by addition of 2.5 mL MeOH, 0.5 mL of 0.3% formic acid and 2.5 mL cold chloroform to each of the three aliquots of the analytical sample from the previous step. The mixture was vortexed for 2 min to favor lipids extraction. The mixture was then centrifuged at 1000 × *g* for 15 min at 4 °C. Then, the upper phase was discarded, the mid-phase saved for re-extraction, and the lower phase saved and mixed with the lower phase resultant from the re-extraction step. The extract was dried under N_2 and kept at −20 °C until analysis.

2.4.2. Extraction with MTBE

This second approach involved the addition of 2.5 mL of MTBE containing 0.1% formic acid to the homogenized tissue (*n*=3). Then, the mixture was agitated for 1 h at

room temperature. To favor phases separation, 625 µL of distilled water was added and the system stirred for 10 min at room temperature; and was centrifuged at 1000 × *g* for 10 min at 4 °C. Subsequently, the upper phase was saved and the lower phase used for re-extraction. The upper phases obtained from both the first extraction and the re-extraction were mixed. Similarly, the extract obtained was dried under N_2 and kept at –20 °C until analysis.

2.5. Solid-phase extraction for isolation of glycerophospholipids

The dried samples were reconstituted in 210 µL of the extractant (1:1 v/v MeOH:$CHCl_3$ or MTBE) and agitated for 5 min. Then, 100 µL of this fraction was put in a vial for direct analysis by LC–QTOF MS/MS. A second aliquot was used for analysis of glycerophospholipids that were isolated using 30 mg HybridSPE® cartridges from Supelco (PA, USA) with the following protocol [32]. A 100 µL aliquot of each lipid sample was mixed with 500 µL of ACN acidified with 1% formic acid (v/v) in the cartridge and left for 2 min for deproteination. Then, vacuum was applied to make the mixture to flow through the sorbent cartridge, which was then washed with 1 mL of ACN. Phospholipids were eluted by pH change using 1 mL of MeOH with 5% (v/v) ammonium hydroxide. The eluate was evaporated and the resulting residue was reconstituted with 40 µL of the same solvent used for lipid extraction, shaken for 1 min and then injected into the LC–QTOF MS/MS equipment.

2.6. LC–QTOF MS/MS analysis

Chromatographic separation was performed by using a Luna Phenomenex C8 column (100 mm × 0.46 mm i.d., 2.5 µm particle size) which was thermostated at 25 °C and protected using a C8 precolumn from Phenomenex. The mobile phases were water (phase A) and 5:2 ACN:IPA (phase B), both containing 0.1% (v/v) formic acid and 5 mM ammonium acetate as ionization agents. The LC pump was programmed at a flow rate of 0.25 mL min^{-1} and the elution gradient was as follows: from min 0 to 30, the percentage of phase B was modified from 50% to 100%, and then, the final mobile phase was hold for 20 min. A post-time of 11 min was used to regain the initial conditions for the next analysis. Thus, the total analysis time per sample was 61 min (including postprocessing). The injected volume was 5 µL, and the injector needle was washed 10

times between injections with 80% MeOH. Also, the needle seat back was flushed with 80% MeOH at 4 mL min^{-1} for 12 s to avoid cross contamination. The autosampler was kept at 4 °C to increase sample stability. The settings of the electrospray ionization source, which was operated in the negative and positive ionization modes, were as follows: capillary voltage ±3.5 kV, Q1 voltage 130 V, N_2 pressure in the nebulizer 35 psi; N_2 flow rate and temperature as drying gas 12 L min^{-1} and 325 °C, respectively. MS/MS data were acquired in both polarities using the centroid mode at a rate of 2.5 spectra s^{-1} in the extended dynamic range mode (2 GHz). Accurate mass spectra in MS scan were acquired in the *m/z* range 40–1200, and in MS/MS mode in the *m/z* range 30–1200. The instrument gave typical resolution 15,000 FWHM at *m/z* 118.0862 and 30,000 FWHM at *m/z* 922.0098. The instrument was calibrated and tuned as recommended by the manufacturer. To assure the desired mass resolution, continuous internal calibration was performed during analyses by using the signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonatedhexakis(1H,1H,3H-tetrafluoropropoxy) phosphazine or HP-921] in the positive ion mode; while in the negative ion mode, ions with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formate adduct) were used. The collision energy was set at 20 V for the whole run. The analytical samples were analyzed in data-dependent acquisition (DDA) to obtain information from fragmentation of the target compounds. The maximum number of precursors selected per cycle was set at 2, with an exclusion window of 0.1 min after 2 consecutive selections of the same precursor.

2.7. Data pretreatment

MassHunter Workstation software (version B7.00 Qualitative Analysis, Agilent Technologies, Santa Clara, CA, USA) was used to process all data obtained by LC–QTOF in auto MS/MS mode. Treatment of raw data files started by extraction of potential molecular features (MFs) with the suited algorithm included in the software. For this purpose, the extraction algorithm considered all ions with single charge state exceeding 500 counts for both polarities. These cut-off values were established taking into account the chromatographic background noise. Additionally, the isotopic distribution to consider MFs as valid should be defined by two or more ions (with a peak spacing

tolerance of *m/z* 0.0025, plus 7.0 ppm in mass accuracy). Ions and adducts formation in the positive (+H, +Na, $+NH_4$) and negative ionization (–H, +HCOO, +Cl) modes, as well as neutral loss by dehydration were included to identify features corresponding to the same potential metabolite.

Identification of the fatty acids and acylglycerides was supported on MS and MS/MS information and search in the METLIN MS and MS/MS databases (http://metlin.scripps.edu), the Human Metabolome Database (HMDB, 3.6 version) and the LIPID MAPS website ((http://www.lipidmaps.org), using in all cases the MFs obtained in the previous step. Additionally, fatty acids identification was verified by injection of a multistandard solution containing the most common fatty acids at 10 mg mL^{-1}. On the other hand, identification of glycerophospholipids was carried out by a search algorithm including characteristic product ions and neutral losses of each phospholipid class and ions ascribed to the alkanoyl chains. For all lipids, identification was validated with the aid of the METLIN MS and MS/MS databases (http://metlin.scripps.edu), the Human Metabolome Database (HMDB, version 3.5) and the Lipid Maps online tool for glycerophospholipids product ion calculation (http://www.lipidmaps.org).

In the next step, a database with all identified metabolites was used to perform a targeted compound extraction analysis using a tolerance window of 0.8 min and 10 ppm mass accuracy. This step was performed by MassHunter Workstation software (version B6.00 Profinder Analysis, Agilent Technologies, Santa Clara, CA, USA). A table with the peak area of all identified compounds in the different samples was obtained as a result.

3. Results and discussion

3.1. Tentative identification of polar lipids in adipose tissue samples

Lipid extracts obtained from human adipose tissue were analyzed by LC–QTOF in MS/MS acquisition mode both directly and after the SPE step. The inclusion of this step was proposed to improve the detection of minor glycerophospholipids that are not detected owing to the presence of other major lipids, particularly neutral lipids, which

suppress ionization of minor concentrated compounds. Both positive and negative ionization modes were tested to maximize the identification coverage. A total of 94 compounds pertaining to 13 lipid classes were tentatively identified by MS/MS considering all the different experiments tested in this research (schemed in Supplementary Figure 1). The classes were fatty acids (FAs), ceramides (CEs), glycerophosphatidylcholines (PCs), glycerophosphatidylinositols (PIs), glycerol-phosphatidylethanolamines (PEs), glycerophosphatidic acids (PAs), glycerol-phosphatidylglycerols (PGs), glycerophosphatidylserines (PSs), plasmalogens (PLs, including O-alkyl and O-alkenyl glycerophospholipids), sphingolipids (sphingomyelins and the sphingoid precursors, SLs), monoacylgycerides (MGs), diacylglycerides (DGs) and triacylglycerides (TGs). Identified lipids encompassed 81 polar lipids and 13 neutral lipids, the latter representing the most concentrated species.

Figure 1 shows the distribution of identified compounds belonging to each family. For glycerophospholipids it is worth emphasizing that *lyso* forms were also identified for two specific families, phosphatidylcholines and phosphatidylethanolamines. As can be seen, PEs, PIs and PLs were the two lipid families with a higher number of identified compounds: 13, 12 and 12% of the total identified lipids, respectively. The 13 identified families can be grouped in four main groups: fatty acids, ceramides, acylglycerides and glycerophospholipids. Supplementary Table 1 shows the chemical structure of the four classes, where fatty acids derivatives are also included. The complete list of identified lipids in adipose tissue appears in Supplementary Table 2, including the chromatographic retention time, precursor ion and the most characteristic product ions that supported the tentative identification of each lipid. As structural isomers may not be distinguished from the MS/MS spectra, no information about them has been included.

The identification of FAs was carried out in negative ionization mode, and the observed precursor ion was $[M–H]^-$ that fits the $R\text{-}COO^-$ ion. Furthermore, FAs identification was verified by injection of a multistandard solution of the main FAs from C12:0 to C24:0. Nine fatty acids were identified in adipose tissue: C16:0, C16:1, C18:0, C18:1, C18:2, C18:3, C20:0, C20:4 and C22:0. Fatty acids represent 10% of the total of

identified lipids (Figure 1). The most concentrated FAs detected in the studied adipose tissue were saturated C16:0 and C18:0.

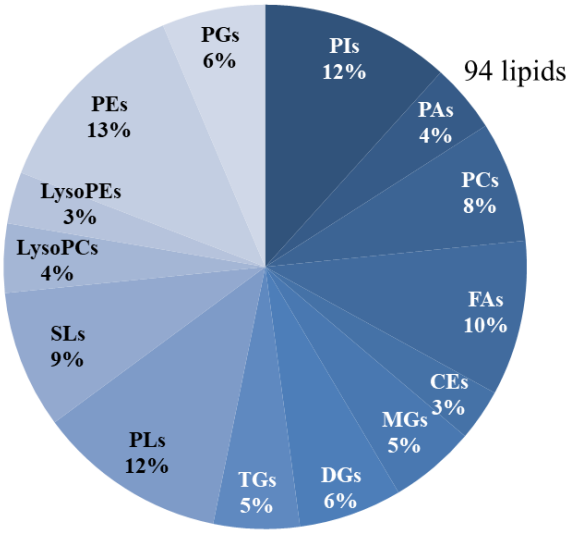


Figure 1. Distribution of lipids identified in adipose tissue by the approach used in this research. FAs: fatty acids; PCs: glycerophosphatidylcholines; PAs: glycerophosphatidic acids; PIs: glycerophosphatidylinositols; PGs: glycerophosphatidylglycerols; PEs: glycerophosphatidylethanolamines; LysoPEs: lysoglycerophosphatidylethanolamines; LysoPCs: lysoglycerophosphatidylcholines; SLs: sphingolipids; PLs: plasmalogens; TGs: triacylglycerides; DGs: diacylglycerides; MGs: monoacylglycerides; and CEs: ceramides.

The identified ceramides corresponded to 3% of total identified lipids, the family with the lowest number of detected compounds. They were better detected in positive ionization mode by the $[M+H]^+$ and $[M+H-H_2O]^+$ precursor ions. Identification was supported on the detection of characteristic product ions at *m/z* 282.3578 and 264.2156 formed by collision-induced dissociation (CID) of ceramides in the Q2 cell (Supplementary Table 3). These fragments were obtained by cleavage of the N-linked fatty acid moiety and loss of one or two molecules of water [31,33–35].

A strategy similar to that used in a previous study by Calderón-Santiago *et al.* [24] was used for the identification of glycerophospholipids. The well stablished and known fragmentation pattern for glycerophospholipid families was used to identify them in

adipose tissue (Supplementary Table 3) [32]. Thus, PCs and sphingomyelins (SMs) have the same fragmentation pattern in positive ionization, with the product ion *m/z* 184.0730 fitting the most characteristic fragment for both families. This ion corresponds to the phosphorylcholine moiety. Glycerophosphatidylethanolamines (PEs) were properly identified in both ionization modes by detection of representative neutral losses with *m/z* 141.0205 in positive ionization mode. Other less abundant families of glycerophospholipids give also a known fragmentation pattern. PIs were identified by a representative neutral loss with *m/z* 260.0651 in positive mode and a fragment of *m/z* 241.0135 in negative mode. Fragmentation of PAs was characterized by the product ion at *m/z* 152.9975 in negative ionization mode. These identifications were combined with detection of the product ions associated to alkanoyl chains ([R-COO$^-$]) in negative ionization mode. This combination allowed elucidating the complete structure of glycerophospholipids.

Apart from polar lipids, neutral lipids were identified in the complete set of experiments developed in this research. Thus, acylglycerides represented 15% of total identified lipids. The detection of acylglycerides was performed in the positive mode as $[M+NH_4]^+$ and $[M+Na]^+$ precursor ions. The primary fragmentation of acylgycerides occurs at the ester bond, yielding product ions easily differentiated by the chain length. These fragments, together with intense signals from the precursor ions, allowed their tentative identification [32,36]. Thus, in the case of MGs the main precursor ion fit the dehydrated $[M+H]^+$, although the $[M+NH_4]^+$ adduct was also observed for some MGs. DGs and TGs precursor ions fit the $[M+H]^+$ and $[M+NH_4]^+$ ions, and product ions formed by loss of the fatty acids conjugated with the acylglyceride were also detected in MS/MS.

3.2. Influence of the tested extractants on the analysis of polar lipids in adipose tissue

This study was targeted at comparing the extraction efficiency of polar lipids from adipose tissue by testing MeOH:$CHCl_3$ and MTBE as extractants. Therefore, comparative analysis considered only identified polar lipids, while the objective for acylglycerides was to minimize their co-extraction. Firstly, a qualitative comparison was

carried out with the number of polar lipids detected after application of all experiments planned in this research (Supplementary Figure 1). This study revealed that the number of compounds identified after using the two extractants was quite similar, 89 and 94 for MeOH:$CHCl_3$ and MTBE, respectively. A Venn-diagram comparing the entities detected with each extractant is shown in Supplementary Figure 2. Five polar lipids were exclusively extracted with MTBE as compared to MeOH:$CHCl_3$, which corresponded to three ceramides, lysoPC(18:1) and PC(34:1), all them low concentrated species. All polar lipids identified in the chloroform phase were also detected in the MTBE extract. Despite the protocols seem to be qualitatively similar in terms of identified lipids, a semiquantitative approach was necessary to clarify the efficiency of each protocol. With this purpose, each family of identified metabolites (fatty acids, ceramides, glycerophospholipids, and acylglycerides) was independently studied.

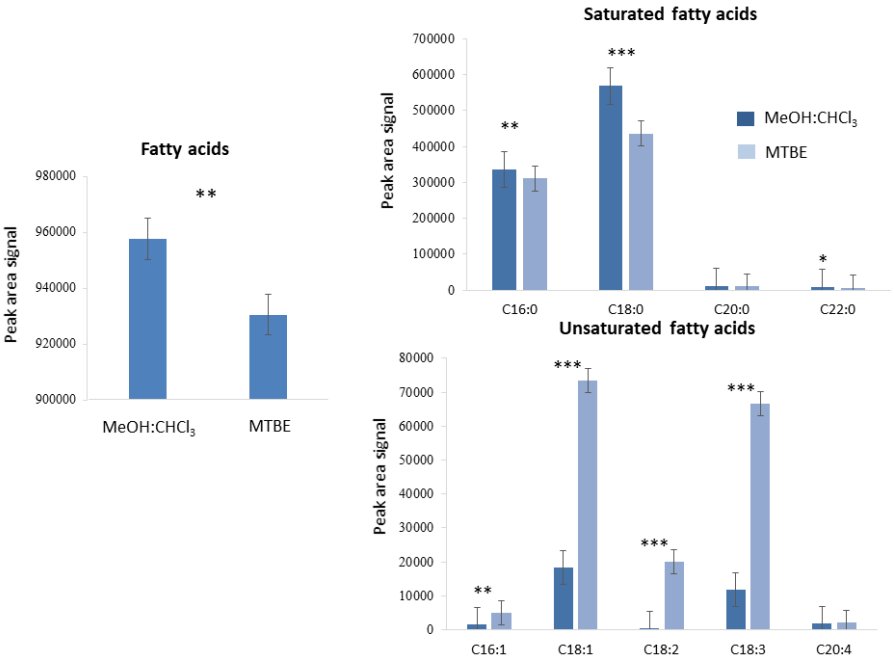


Figure 2. Bars diagrams comparing the sum of quantitative responses (peak area signals) of unsaturated and saturated fatty acids identified in adipose tissue after isolation by the two tested extractants. Statistical analysis was carried out by *t*-test analysis (**p*-value between 0.05 and 0.01, ***p*-value between 0.01 and 0.001, and ****p*-value < 0.001).

From the two extracts, the family of acylglycerides reported the signals with the highest intensity, which clearly dominate the base peak chromatogram (BPC). This is the main limitation in the analysis of polar lipids owing to acylglycerides typically provide wide chromatographic peaks interfering the ionization of minor concentrated lipids. On the contrary, ceramides provided the lowest intensity signals, which complicated their identification.

Fatty acids were extracted in overall terms with a similar efficiency by both extractants, as shows Figure 2. However, some differences were found when saturated and unsaturated fatty acids were independently monitored. Thus, saturated fatty acids —particularly palmitic and stearic acids— more concentrated in adipose tissue were quantitatively extracted by either MeOH:$CHCl_3$ or MTBE. Only a relative difference was found for stearic acid that was more concentrated in the extract obtained by MeOH:$CHCl_3$. Nevertheless, this difference did not affect the detection of stearic acid in adipose tissue since this fatty acid was the most concentrated in these samples. Unsaturated fatty acids were characterized by a behavior that depended on the extractant. Thus, MTBE clearly provided a higher extraction efficiency of unsaturated fatty acids. This situation was especially critical for unsaturated C18 fatty acids such as oleic, linoleic and linolenic acids, as Figure. 2 shows. As can be seen, the ratios of peak area intensity between saturated and unsaturated fatty acids were around 10, which gives an added value to MTBE as extractant for isolation of unsaturated fatty acids in the presence of more concentrated saturated fatty acids. Glycerophospholipids represented 71% of total identified lipids (Figure 1), being the most important group in terms of variability. The structural variability of glycerophospholipids justifies the independent assessment of the detected families PAs, PCs, PEs, PGs, PIs, PLs and SLs, as well as the *lyso* forms of PCs and PEs. Specifically, the *lyso* forms, which encompassed several lysoPCs and lysoPE(20:4), experienced a common behavior since they were better extracted with MTBE. This is particularly significant for lysoPE(20:4) since the signal corresponding to this glycerophospholipid was the most intense among those provided by the detected glycerophospholipids in the MTBE extract.

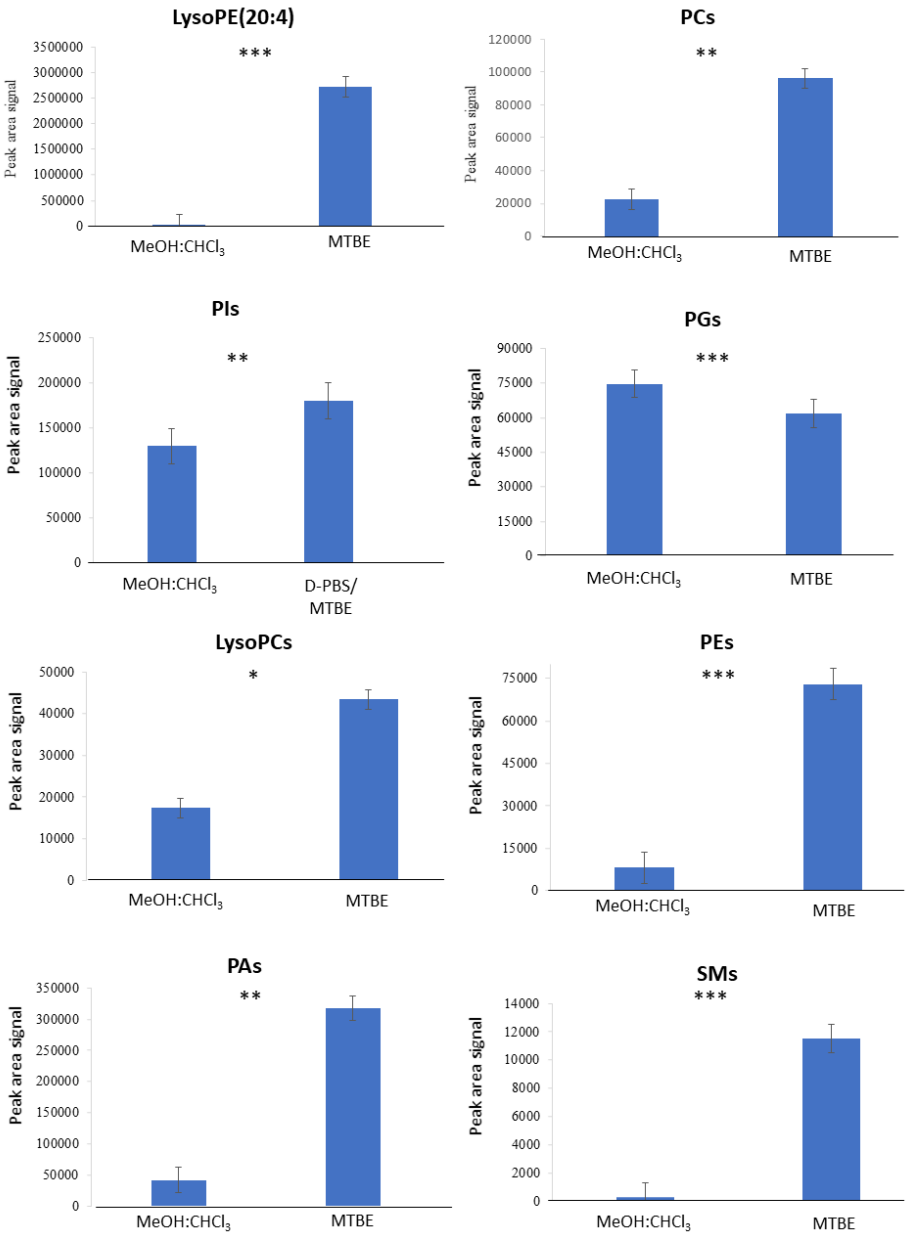


Figure 3. Bars diagrams comparing the sum of quantitative responses (peak area signals) of glycerophospholipids identified in adipose tissue after isolation by the two tested extractants. Statistical analysis was carried out by *t*-test analysis (**p*-value between 0.05 and 0.01, ***p*-value between 0.01 and 0.001, and ****p*-value < 0.001).

Figure 3 shows the variability observed for the other glycerophospholipid families. It is worth mentioning that PCs, PEs, PIs, PAs and SLs reported similar results, since their extraction was favored by the use of MTBE as compared to MeOH:$CHCl_3$. The only glycerophospholipid family that showed an opposite behaviour was that of PGs, since these compounds were better extracted by MeOH:$CHCl_3$. Therefore, considering the main families of glycerophospholipids, it can be deduced that the extraction performance of MTBE is better than that of MeOH:$CHCl_3$.

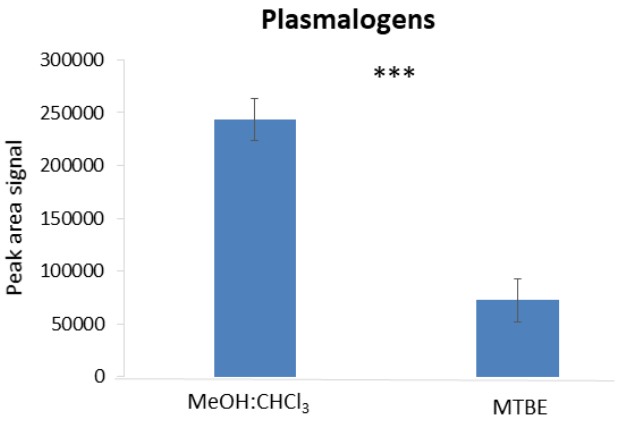


Figure 4. Bar diagram comparing the sum of quantitative responses (peak area signals) of plasmalogens identified in adipose tissue after isolation by the two tested extractants. Statistical analysis was carried out by *t*-test analysis (***p-value < 0.001).

Special attention was paid to plasmalogens that included O-1-alkyl and O-1-alkenyl glycerophospholipids. The extraction performance of the two extractants for this particular family of lipids, shown in Figure 4, revealed that this minor family of glycerophospholipids was better extracted by MeOH:$CHCl_3$. Finally, ceramides, the class giving the lowest intensity of chromatographic signals, were only detected in the extract obtained by MTBE.

3.3. Co-extraction of acylglycerides as interferents in the analysis of polar lipids in adipose tissue

Acylglycerides, the lipid family more concentrated in adipose tissue, represent a relevant group of interferents in the analysis of polar lipids by LC–MS/MS. Acylglycerides are co-extracted with polar lipids by using generic protocols such as the Bligh-Dyer or Folch methods. There are selective protocols for isolation of neutral lipids, but generic protocols for extraction of polar lipids are not selective at all. This effect is visualized in Supplementary Figure 3 that illustrates the extraction efficiency of total identified acylglycerides as well as the extraction efficiency of TGs, DGs and MGs. As can be seen, acylgycerides were similarly extracted by the two extractants, which certifies the lack of selectivity of generic extraction protocols. In general terms, almost all acylgycerides were better extracted with MeOH:$CHCl_3$ in comparison to MTBE. However, the most concentrated acylglyceride, TG(16:0/18:1/20:4), showed a preferred transfer from the homogenate of adipose tissue to the MTBE extractant. This effect compensates the lower extraction efficiency of MTBE observed for the rest of acylglycerides.

With these premises, it is worth noting that the detection of polar lipids demands for an additional sample preparation step to increase selectivity by minimizing the contribution of acylgycerides. This step was based on SPE for selective retention of glycerophospholipids, subsequently eluted after removal of acylglycerides and other interferents.

3.4. Enhancement of glycerophospholipids detection after SPE clean-up

The capability of SPE as sample preparation step prior to LC–MS/MS analysis was tested to enhance the detection capability in lipidomics analysis of adipose tissue. For this purpose, an SPE sorbent especially suited for selective retention of glycerophospholipids was selected. The sorbent, zirconia coated silica, allows a selective interaction between the phosphate moiety of glycerophospholipids (Lewis base) and Zr atoms acting as a Lewis acid by accepting electrons in *d*-orbitals. The selectivity of this treatment enables removal of other lipids such as acylglycerides or proteins that could exert ionization suppression effects. The implementation of SPE clean-up allowed

significant improvement of the detection capability of lipids isolated from adipose tissue, as Figure 5 shows. Supplementary Figure 4 compares the BPCs obtained by direct analysis with those after the SPE protocol. Direct analysis of the extract from adipose tissue allowed identification of 65 lipids belonging to different families, whereas the implementation of the SPE protocol increased the identification coverage up to 94 lipids (Figure 5). Therefore, after clean-up of the extract 29 glycerophospholipids no detected by direct analysis were identified. These additional glycerophospholipids pertained to different families such as *lyso* forms, PAs, PCs, PEs, PGs, PIs, and SLs. It should be emphasized that all glycerophospholipids identified by direct analysis (31 compounds) were also identified after the SPE step, and they provided higher signals in the latter case. This behavior can be visualized in Supplementary Figure 5, which illustrates differences in sensitivity for five representative glycerophospholipids. The sensitivity improvement was supported on the clean-up effect (since no other families of lipids are detected after this preparation), but also on a concentration effect since the SPE protocol starts with 100 mL and, after evaporation, the residue is reconstituted in 40 mL (concentration factor 2.5). It is evident that the treatment of the adipose tissue extract improves the detection capability by isolation of minor glycerophospholipids.

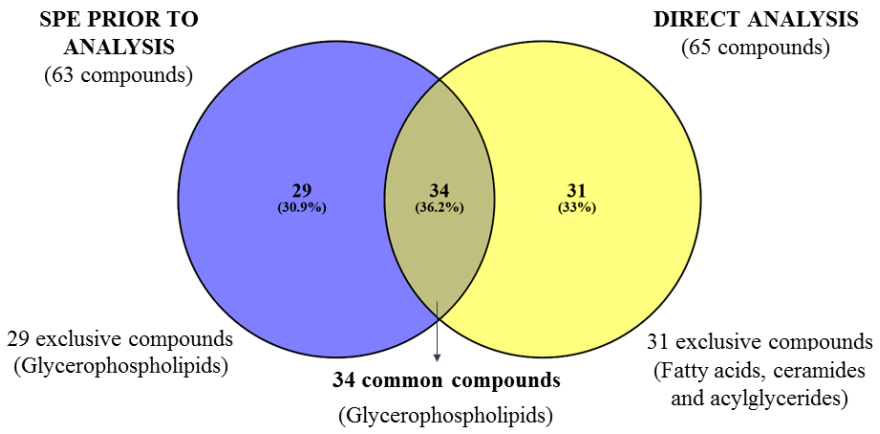


Figure 5. Venn diagram comparing the identification coverage obtained by direct analysis of the MTBE extract *vs*. the additional implementation of an SPE protocol for selective isolation of glycerophospholipids.

4. Conclusions

The influence of sample preparation on lipidomic analysis of polar lipids in adipose tissue was studied. Two different extractants for lipids isolation were assayed: MeOH:$CHCl_3$ and MTBE. The conclusions provided by this research can be summarized as follows: (i) sample preparation is a key step for the identification of different families of lipids in adipose tissue; (ii) the use of MTBE as extractant led to higher extraction efficiency for unsaturated fatty acids, glycerophospholipids and ceramides, while MeOH:$CHCl_3$ favored the isolation of saturated fatty acids and plasmalogens, aspect of especial importance taking into account that certain families are present at low concentrations; and (iii) the implementation of a clean-up step based on SPE improved the detection of minor glycerophospholipids as compared to direct analysis of the extract, increasing by 50% the number of detected glycerophospholipids. According to these results, the recommended sample preparation for analysis of polar lipids in adipose tissue would be liquid–liquid extraction combined with an SPE step to enhance detection of glycerophospholipids. Concerning the extractant, MTBE favored the detection of less abundant lipids such as ceramides and unsaturated fatty acids and, therefore, it would be suggested for untargeted analysis of polar lipids.

Acknowledgments

This research has been financed jointly by the Spanish Ministerio de Economía y Competitividad through the projects CTQ2015-68813-R, BFU2013-44229-R, and BFU2015-70454-REDT, Instituto de Salud Carlos III through the project PIE14/00005/1 (FLEXI-MET), and Junta de Andalucía/FEDER through the project FQM-1602. The three projects are co-funded by European Regional Development Fund/European Social Fund ("Investing in your future"). M.A. López-Bascón thanks the Ministerio de Educación, Cultura y Deporte (MECD) for an FPU scholarship (FPU15/03068). Professor M.D. Luque de Castro is also thanked for the careful revision of manuscript. CIBER de Fragilidad y Envejecimiento Saludable (CIBERfes) and CIBER Fisiopatología de la Obesidad y Nutrición (CIBERobn) are initiatives of ISCIII, Spain.

References

[1] M.M. Yore, I. Syed, P.M. Moraes-Vieira, T. Zhang, M.A. Herman, E.A. Homan, R.T. Patel, J. Lee, S. Chen, O.D. Peroni, A.S. Dhaneshwar, A. Hammarstedt, U. Smith, T.E. McGraw, A. Saghatelian, B.B. Kahn, Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects, Cell 159 (2014) 318–332.

[2] B.M. Spiegelman, J.S. Flier, Adipogenesis and obesity: rounding out the big picture, Cell 87 (1996) 377–389.

[3] M. Ståhlman, C.S. Ejsing, K. Tarasov, J. Perman, J. Borén, K. Ekroos, High-throughput shotgun lipidomics by quadrupole time-of-flight mass spectrometry, J. Chromatogr. B 877 (2009) 2664–2672.

[4] E. Fahy, S. Subramaniam, H.A. Brown, C.K. Glass, A.H. Merrill, R.C. Murphy, C.R.H. Raetz, D.W. Russell, Y. Seyama, W. Shaw, T. Shimizu, F. Spener, G. van Meer, M.S. VanNieuwenhze, S.H. White, J.L. Witztum, E. a Dennis, A comprehensive classification system for lipids, J. Lipid Res. 46 (2005) 839–861.

[5] T. Hyötyläinen, M. Orešič, Systems biology strategies to study lipidomes in health and disease, Prog. Lipid Res. 55 (2014) 43–60.

[6] E. Muro, G.E. Atilla-Gokcumen, U.S. Eggert, Lipids in cell biology: how can we understand them better?, Mol. Biol. Cell 25 (2014) 1819–23.

[7] E.E. Kershaw, J.S. Flier, Adipose tissue as an endocrine organ, J. Clin. Endocrinol. Metab. 89 (2004) 2548–2556.

[8] R.C. Baker, Y. Nikitina, A.R. Subauste, *Analysis of adipose tissue lipid using mass spectrometry*, in: Methods Enzymol., Elsevier, 2014: pp. 89–105.

[9] T. Cajka, O. Fiehn, Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry, Trends Anal. Chem. 61 (2014) 192–206.

[10] S.P. Weisberg, D. McCann, M. Desai, M. Rosenbaum, R.L. Leibel, A.W. Ferrante, Obesity is associated with macrophage accumulation in adipose tissue, J. Clin. Invest. 112 (2003) 1796–1808.

[11] N. Musi, R. Guardado-Mendoza, Adipose tissue as an endocrine organ, Cell. Endocrinol. Heal. Dis. 89 (2014) 229–237.

[12] R.W. Gross, X. Han, Lipidomics at the interface of structure and function in systems biology., Chem. Biol. 18 (2011) 284–91.

[13] M.R. Wenk, The emerging field of lipidomics, Nat. Rev. Drug. Discov. 4 (2005) 594–610.

[14] A. Lopategi, C. López-Vicario, J. Alcaraz-Quiles, V. García-Alonso, B. Rius, E. Titos, J. Clària, Role of bioactive lipid mediators in obese adipose tissue inflammation and endocrine dysfunction, Mol. Cell. Endocrinol. 419 (2016) 44–59.

[15] M. Orešič, Metabolomics, a novel tool for studies of nutrition, metabolism and lipid dysfunction, Nutr. Metab. Cardiovasc. Dis. 19 (2009) 816–824.

[16] J. Godzien, M. Ciborowski, P. Samczuk, A. Kretowski, C. Barbas, Rapid and reliable identification of phospholipids for untargeted metabolomics with LC–ESI–QTOF–MS/MS, J. Proteome Res. 14 (2015) 3204–32016.

[17] H. Li, X. Xia, X. Li, G. Naren, Q. Fu, Y. Wang, C. Wu, S. Ding, S. Zhang, H. Jiang, J. Li, J. Shen, Untargeted metabolomic profiling of amphenicol-resistant campylobacter jejuni by ultra-high-performance liquid chromatography – mass spectrometry, J. Proteome Res. 14 (2015) 1060–1068.

[18] H. Nygren, T. Seppänen-laakso, T.H. Castillo, Sandra, Matej Oreši˘, Liquid chromatography-mass spectrometry (LC-MS)-based lipidomics for studies of body fluids and tissues, Methods Mol. Biol. 708 (2011) 247–257.

[19] H. Ogiso, T. Suzuki, R. Taguchi, Development of a reverse-phase liquid chromatography electrospray ionization mass spectrometry method for lipidomics, improving detection of phosphatidic acid and phosphatidylserine, Anal. Biochem. 375 (2008) 124–131.

[20] L.D. Roberts, J.A. West, A. Vidal-Puig, J.L. Griffin, *Methods for performing lipidomics in white adipose tissue*, in: Methods Enzymol., Elsevier, 2014: pp. 211–231.

[21] A. Kotronen, T. Seppänen-Laakso, J. Westerbacka, T. Kiviluoto, J. Arola, A. Ruskeepää, H. Yki-Järvinen, M. Oresic, Comparison of lipid and fatty acid composition of the liver, subcutaneous and intra-abdominal adipose tissue, and serum, Obesity 18 (2010) 937–944.

[22] M. Ebadi, V.E. Baracos, O.F. Bathe, L.E. Robinson, V.C. Mazurak, Loss of visceral adipose tissue precedes subcutaneous adipose tissue and associates with n-6 fatty acid content, Clin. Nutr. 35 (2016) 1347–1353.

[23] L. Hodson, M. Neville, M.F.-F. Chong, I. Rogers, S.S. Huda, D.J. Freeman, K.N. Frayn, B.A. Fielding, Micro-techniques for analysis of human adipose tissue fatty acid composition in dietary studies, Nutr. Metab. Cardiovasc. Dis. 23 (2013) 1128–33.

[24] L. Hodson, C.M. Skeaff, B.A. Fielding, Fatty acid composition of adipose tissue and blood in humans and its use as a biomarker of dietary intake, Prog. Lipid Res. 47 (2008) 348–380.

[25] J. Folch, M. Lees, G.H.S. Stanley, A simple method for the isolation and purification of total lipides from animal tissues, J. Biol. Chem. (1957) 497–509.

[26] E.G. Bligh, W.J. Dyer, A rapid method of total lipid extraction and purification, Can. J. Biochem. Physiol. 37 (1959).

[27] M. Pulfer, R.C. Murphy, Electrospray mass spectrometry of phospholipids, Mass Spectrom. Rev. 22 (2003) 332–364.

[28] M. Jové, J.M. Moreno-Navarrete, R. Pamplona, W. Ricart, M. Portero-Otín, J.M. Fernández-Real, Human omental and subcutaneous adipose tissue exhibit specific lipidomic signatures, FASEB J. 28 (2014) 1071–1081.

[29] V. Matyash, G. Liebisch, T. V Kurzchalia, A. Shevchenko, D. Schwudke, Lipid extraction by methyl-*tert*-butyl ether for high-throughput lipidomics, J. Lipid Res. 49 (2008) 1137–

46.

[30] P.M. Hutchins, R.M. Barkley, R.C. Murphy, Separation of cellular nonpolar neutral lipids by normal-phase chromatography and analysis by electrospray ionization mass spectrometry, J. Lipid Res. 49 (2008) 804–813.

[31] WMA, Declaration of Helsinki., in: WORLD Med. Assoc. Declar. HELSINKI, 2004: pp. 1418–1419.

[32] M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M.D. Luque de Castro, Analysis of serum phospholipid profiles by liquid chromatography-tandem mass spectrometry in high resolution mode for evaluation of atherosclerotic patients, J. Chromatogr. A 1371 (2014) 154–62.

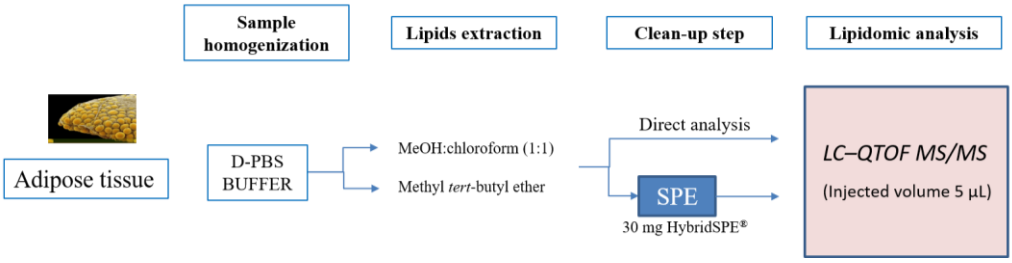
[33] M. Fillet, J.C. Van Heugen, A.C. Servais, J. De Graeve, J. Crommen, Separation, identification and quantitation of ceramides in human cancer cells by liquid chromatography-electrospray ionization tandem mass spectrometry, J. Chromatogr. A 949 (2002) 225–233.

[34] T. Kasumov, H. Huang, Y. Chung, R. Zhang, A.J. Mccullough, J.P. Kirwan, Quantification of ceramide species in biological samples by liquid chromatography electrospray ionization tandem mass spectrometry, Anal. Biochem. 401 (2010) 154–161.

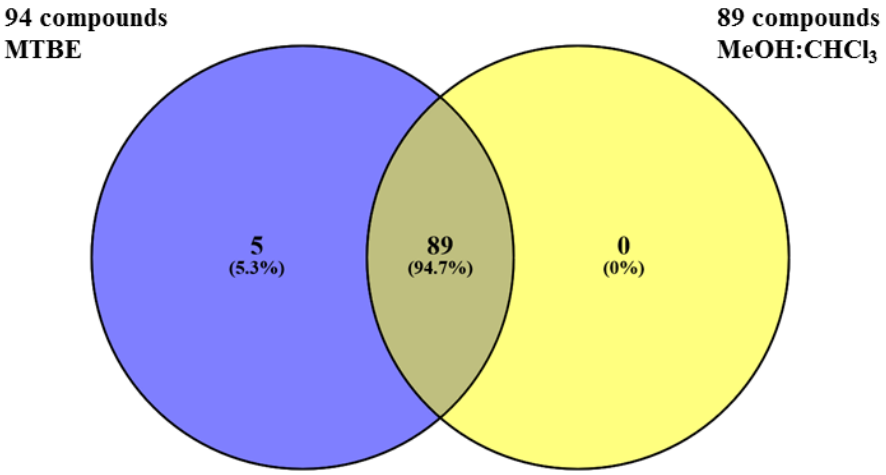
[35] M. Gu, J.L. Kerwin, J.D. Watts, R. Aebersold, Ceramide profiling of complex lipid mixtures by electrospray ionization mass spectrometry, Anal. Biochem. 244 (1997) 347–356.

[36] P.T. Ivanova, S.B. Milne, M.O. Byrne, Y. Xiang, H.A. Brown, *Glycerophospholipid identification and quantitation by electrospray ionization mass spectrometry*, in: Methods Enzymol., 2007: pp. 21–57.

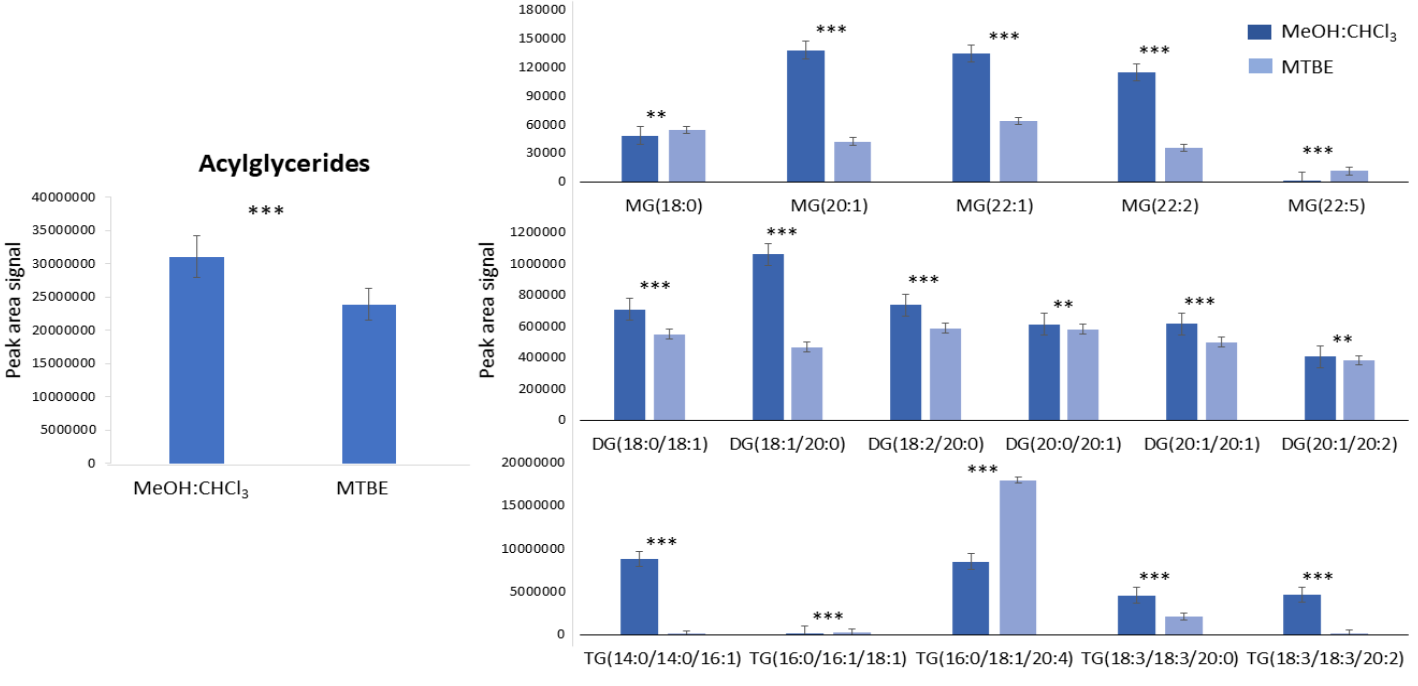
Supplementary Information



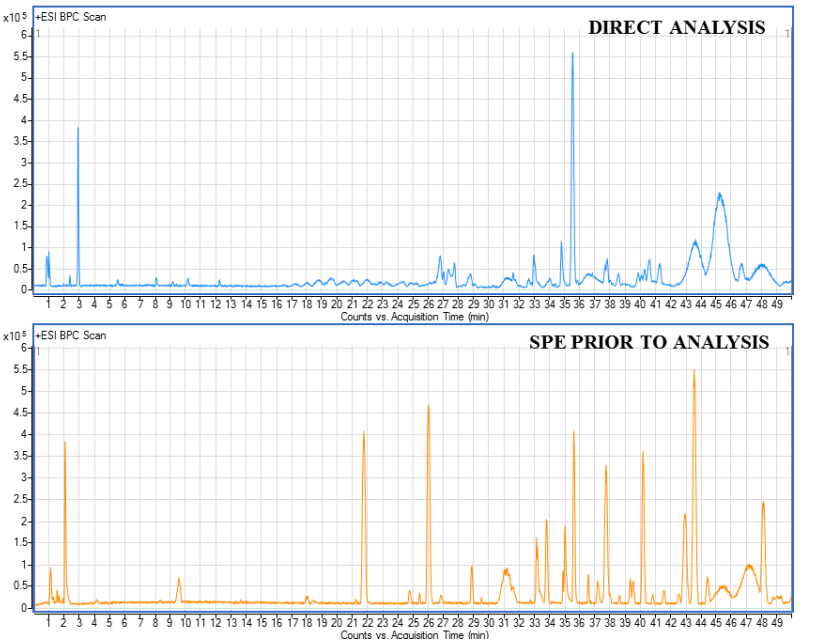
Supplementary Figure 1. Experiments designed to study the influence of sample preparation on the lipidomics analysis of human adipose tissue.



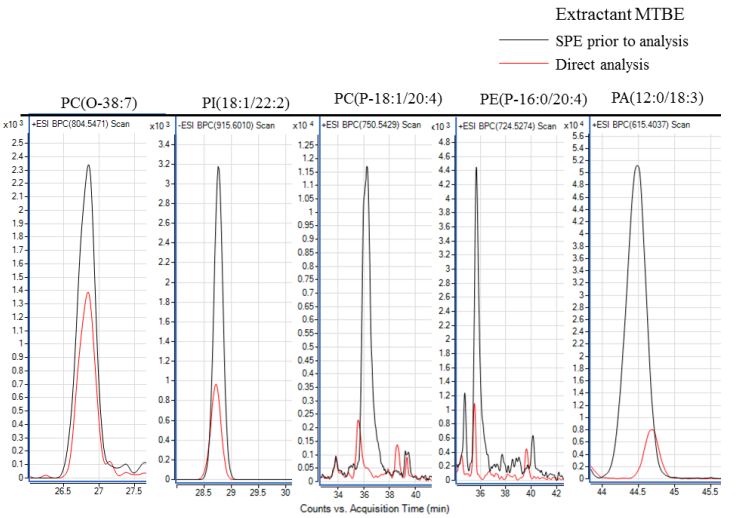
Supplementary Figure 2. Venn diagram for comparison of the lipids identified in human adipose tissue after isolation by the two tested extractants.



Supplementary Figure 3. Bars diagrams comparing the sum of quantitative responses (peak area signals) of mono-, di- and triglycerides identified in adipose tissue after isolation with the two tested extractants. Statistical analysis was carried out by *t*-test analysis (**p*-value between 0.05 and 0.01, ***p*-value between 0.01 and 0.001, and ****p*-value < 0.001).



Supplementary Figure 4. Base peak chromatograms obtained by direct analysis of human adipose tissue subjected to extraction with MTBE and after SPE clean-up/preconcentration of the extract.



Supplementary Figure 5. Base peak chromatograms illustrate differences in sensitivity between direct analysis of the MTBE extract (red line) and analysis after SPE clean-up/preconcentration (black line) for five representative glycerophospholipids.

CHAPTER V

Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage

Talanta

199 (2019) 303–309

Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage

María Asunción López-Bascón[a,b,c,d], Mónica Calderón-Santiago[a,b,c,d], Héctor Argüello[e], Luis Morera[e], Juan José Garrido[e], Feliciano Priego-Capote[a,b,c,d]*

[a]*Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain*

[b]*CeiA3 Agroalimentary Excellence Campus, University of Córdoba, Córdoba, Spain*

[c]*Maimónides Institute for Biomedical Research (IMIBIC)/University of Córdoba/Reina Sofía University Hospital, Córdoba, Spain*

[d]*CIBER Fragilidad y Envejecimiento Saludable (CIBERfes), Instituto de Salud Carlos III, Spain*

[e]*Department of Genetic, University of Córdoba, Córdoba, Spain*

Comprehensive analysis of pig feces metabolome by chromatographic techniques coupled to mass spectrometry in high resolution mode: Influence of sample preparation on the identification coverage

María Asunción López-Bascón, Mónica Calderón-Santiago, Héctor Argüello, Luis Morera, Juan José Garrido, Feliciano Priego-Capote

ABSTRACT

Pig feces is an interesting biological sample to be implemented in metabolomics experiments by virtue of the information that can be deduced from the interaction between host and microbiome. However, pig fecal samples have received scant attention, especially in untargeted metabolomic studies. In this research, an analytical strategy was planned to maximize the identification coverage of metabolites found in pig fecal samples. For this purpose, two complementary platforms such as LC–QTOF MS/MS and GC–TOF/MS were used. Concerning sample preparation, six extractant solvents with different polarity grade were tested to evaluate the extraction performance and, in the particular case of GC–MS, two derivatization protocols were compared. A total number of 303 compounds by combination of all the extractants and analytical platforms were tentatively identified. The main identified families were amino acids, fatty acids and derivatives, carbohydrates and carboxylic acids. For GC–TOF/MS analysis, the recommended extractant is methanol, while methoxymation was required in the derivatization protocol since this step allows detecting the -keto acids, which are direct markers of the microbiome status. Concerning LC–QTOF MS/MS analysis, a dual extraction approach with methanol (MeOH) or MeOH/water and ethyl acetate is proposed to enhance the detection of polar and non-polar metabolites.

Keywords: Feces samples, LC–QTOF MS/MS, GC–TOF/MS, metabolites, microbiome, pig

1. Introduction

Intestinal pathologies caused by zoonotic pathogens such as *Salmonella spp* are considered one of the most important risk factors affecting pig farms [1]. Pigs are susceptible to intestinal diseases, which can also be transmitted to humans. Therefore, effective veterinary control of pigs aims firstly to the protection of consumers and, secondly, to the maintenance of animal productivity [1]. The pig is also extensively used as preferred animal model for analysis of a wide range of physiological functions and diseases.

The majority (>90%) of the bacteria in the pig intestinal microbiome are classified into two main *phyla*: *Firmicutes* and *Bacteroidetes*. Additionally, the ileum has a high percentage of *Proteobacterium phylum* bacteria (up to 40%) [2]. The microbiota and the genes that comprise the microbiome play a key role in the health of organisms. The gut microbiota is involved in important functions such as providing specific dietary energy to the host with the generation of digestible carbohydrates and short chain fatty acids (SCFAs), or synthetizing compounds with antibiotic properties to protect against infectious diseases [3,4]. Therefore, dysbiosis of gut bacteria contributes to the occurrence of diseases and pathogenic invasion [5].

Metabolomics, thanks to the technology used nowadays, enables to measure the concentration of hundreds of small molecules in a biological sample, which may aid to elucidate biological mechanisms and facilitate early diagnosis of diseases. In this sense, untargeted metabolomic analysis is being increasingly employed since it allows the comprehensive determination of metabolic profiles from which novel biomarkers could be proposed. Despite sample preparation is critical for this type of analysis, this step has received scant attention, being the limiting factors both the matrix diversity and sample physiological variation.

Urine and blood (serum and plasma) are the most commonly used biofluids for exploring systematic alteration of metabolites because they are reasonably easy to obtain and are collected in a relatively non-invasive way [5]. However, there are other biological samples of great interest for metabolomic analysis, which have been scarcely

considered. This is the case of feces samples that can reflect the microbial-mammalian interaction as an essential element in the study of mammalian metabolome [6,7].

The objective of this research was to maximize the identification coverage of metabolites found in pig fecal samples. For this purpose, two hyphenated detection techniques such as LC–QTOF MS/MS and GC–TOF/MS were combined to evaluate their additivity in terms of identification. Concerning sample preparation, six extractant solvents with different polarity were tested to evaluate the extraction performance and, in the particular case of GC–MS, two derivatization protocols were compared to check the influence of this step on the detection capability of metabolomics methods.

2. Materials and methods

2.1. Samples, chemicals and reagents

Feces samples were aseptically collected from five pigs and poured in plastic tubes that were immediately frozen in liquid nitrogen. Frozen samples were kept at −80 °C for one week before being processed. A pool of pig feces was prepared by homogenous mixing of samples from the five animals. All procedures involving animals were performed in accordance with the European regulations regarding the protection of animals used for experimental and other scientific purposes, under the supervision of the Ethical and Animal Welfare Committee of the University of León (Spain).

Mass spectrometry grade (MS-grade) methanol (MeOH), dichloromethane, *n*-hexane and ethyl acetate from Sigma–Aldrich (Madrid, Spain) were used as extractants. Deionized water (18 MΩ·cm) supplied by a Milli-Q water purification system from Millipore (Bedford, MA, USA) was used to prepare the chromatographic aqueous phase (phase A) and also as extractant solvent. MS-grade acetonitrile (ACN) from Sigma–Aldrich was used to prepare chromatographic mobile phase B. MS-grade formic acid and ammonium formate from Sigma–Aldrich were used as ionization agents for LC–MS/MS analysis. Bis-(trimethylsilyl) fluoroacetamide (BSTFA), trimethylchlorosilane (TMCS) and methoxyamine hydrochloride from Sigma–Aldrich were used as silylation

and methoxymation agents in the derivatization step. Pyridine from Merck (Darmstadt, Germany) was used as solvent in the derivatization step. Mass spectrometry reference solution for ESI and MS grade perfluorotri-*n*-butylamine (PFTBA) from Agilent Technologies (Santa Clara, CA, USA) were used for daily mass calibration of LC–QTOF MS/MS and GC–TOF/MS instrumentation, respectively. An alkane standard mixture (from C10 to C40) from Sigma–Aldrich was used to establish the retention index (RI)–retention time correlation in GC–TOF/MS.

2.2. Apparatus and instruments

A Teflon homogenizer from Kimble Kontes Vineland (NJ, USA) was used for feces sample–extractant homogenization and a block heater from Stuart Equipment (Staffordshire, ST15 OSA, UK) was used in the derivatization step. A concentrator Plus speed-vac from Eppendorf (Hamburg, Germany) was used to evaporate the solvents of extraction step and a vortex shaker from IKA (Wilmington, NC, USA) was used for sample agitation.

An Agilent 1200 Series LC system coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer was used for LC–MS/MS analysis of the extracts in both ionization modes. The QTOF detector was equipped with a Jet Stream Technology electrospray ion source for simultaneous spraying of chromatographic eluate and a reference solution to calibrate continuously the detected *m/z* ratios. Chromatographic eluates were monitored by tandem mass spectrometry in high resolution mode. For analysis by GC–MS an Agilent 7890A Series GC system coupled to an Agilent 7200 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with an electron impact (EI) source was used. The analytical sample was thus monitored in high resolution mode. The Agilent MassHunter Workstation software (version B.06, Agilent Technologies) was used to control the instrument and acquire the data in both instruments.

2.3. Extraction of metabolites from feces samples

Supplementary Figure 1 shows a scheme of sample preparation protocols. Eighteen 160-mg aliquots of feces pool sample were weighted in glass tubes for

extraction of metabolites with 4 mL of extractant. The tested extractants were deionized water, MeOH, 1:1 (v/v) deionized water:MeOH, ethyl acetate, hexane and dichloromethane, and each extractant was tested with three aliquots. The heterogeneous system thus formed in each tube was kept under agitation for 5 min in a Teflon homogenizer. Then, each tube was centrifuged at 12500 × *g* for 10 min at 4 °C, the supernatant was filtered through 0.2 μm filter and divided into three aliquots of 100 μL that were put in three glass tubes and evaporated to dryness using a speed-vac. One aliquot was used for LC–QTOF MS/MS analysis in both ionization modes, while the other two aliquots were prepared for GC–TOF/MS. For this purpose, the resulting solid residues were reconstituted with the adequate solvent, which is specified in the protocol for each instrument.

2.4. LC–QTOF MS/MS analysis

For LC–QTOF MS/MS analysis, the resulting residues were reconstituted with 100 μL of ACN and shaken in a vortex for 30 s (Supplementary Figure 1). All samples were analyzed in triplicate. Chromatographic separation was performed by using a Poroshell 120 EC-C18 column (50 mm × 2.1 mm i.d., 2.7 μm particle size) which was thermostated at 25 °C and protected using an EC-C18 precolumn (4.5 × 5 mm i.d., 2.7 μm) from Agilent Technologies. The mobile phases were 95:5 water:ACN (phase A) and 95:5 ACN:water (phase B), both containing 0.1% (v/v) formic acid and 5 mM ammonium formate as ionization agents. The LC pump was programmed at a flow rate of 0.4 mL min^{-1} and the elution gradient was as follows: 100% phase A as initial mobile phase was kept constant for 2 min, from min 2 to 11, the percentage of phase B was modified from 0% to 100% and then, the final percentage was hold for 10 min. A post-time of 7 min was used to equilibrate the system to initial conditions for the next analysis. Thus, the total analysis time per sample was 29 min (including postprocessing). The injected volume was 5 μL, and the injector needle was washed 10 times between injections with 80% MeOH. Also, the needle seat back was flushed with 80% MeOH at 4 mL min^{-1} for 12 s to avoid cross contamination. The autosampler was kept at 4 °C to increase sample stability. The settings of the electrospray ion source, which was operated in the negative and positive ionization modes, were as follows:

capillary voltage ±3.5 kV, Q1 voltage 130 V, N_2 pressure in the nebulizer 40 psi; N_2 flow rate and temperature as drying gas 10 L min^{-1} and 325 °C, respectively, and sheath gas flow and temperature were set at 12 mL min^{-1} and 350 °C. MS/MS data were acquired in both polarities, using the centroid mode at a rate of 2.5 spectra s^{-1} in the extended dynamic range mode (2 GHz).

The instrument gave typical resolution 18,000 Full Width at Half Maximum (FWHM) at *m/z* 118.0862 and 35,000 FWHM at *m/z* 922.0098. The instrument was calibrated and tuned as recommended by the manufacturer. To assure the desired mass resolution, continuous internal calibration was performed during analyses by using the signals at *m/z* 121.0509 (protonated purine) and *m/z* 922.0098 [protonatedhexakis-(1H,1H,3H-tetrafluoropropoxy) phosphazine or HP-921] in the positive ion mode; while in the negative ion mode, ions with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formate adduct) were used. The analytical samples were injected in auto MS/MS acquisition mode to obtain information from fragmentation of the target compounds. The collision energy was set at 20 V for the whole run. The maximum number of precursors selected per cycle was set at 2, with an exclusion window of 0.1 min after 2 consecutive selections of the same precursor. Accurate mass spectra in MS scan were acquired in the *m/z* range 40–1200, and in MS/MS mode in the *m/z* range 30–1200.

2.5. GC–TOF/MS analysis

Two different derivatization protocols were tested for GC–TOF/MS analysis, as can be seen in Fig. S1. For the first derivatization protocol, the residues resulting from the sample preparation step were reconstituted with 20 µL of methoxyamine in pyridine (20 mg mL^{-1}) and maintained at 30 °C for 90 min. Later, 180 µL of a 98:2 (v/v) BSTFA–TMCS mixture was added to the reconstituted analytical sample, shaken for 30 s and maintained at 37 °C for 60 min. In the second derivatization protocol the methoxymation step was omitted and the residues were reconstituted in pyridine, and then, the mixture BSTFA–TMCS was added. All samples were analyzed in triplicate. GC–TOF/MS analyses were performed by EI ionization mode at 70 eV. Chromatographic separation was carried out with a fused silica DB-5MS-UI (30 m ×

0.25 mm i.d, × 0.25 μm) film thickness capillary column from Agilent Technologies. The GC oven temperature program started at 60 °C (1 min held), followed by a temperature ramp of 10 °C min^{-1} to final 300 °C (2 min held). Post-run time was programmed for 4 min up to 310 °C to assure complete elution of the injected sample. Pulsed splitless injections of sample (1 μL) were carried out at 250 °C and ultrapure grade helium was used as carrier gas at 1.0 mL min^{-1} flow rate. The interface, ion source and quadrupole temperatures were set at 280, 300 and 200 °C, respectively. A solvent delay of 5.5 min was used to prevent damage in the ion source filament. The TOF detector was operated at 5 spectra s^{-1} in the mass range *m/z* 50–550 and the instrument gave a resolution of 8500 full width half maximum (FWHM) at *m/z* 501.9706. A mass calibration between samples was performed with PFTBA, as recommended by the manufacturer.

2.6. Data processing, identification of metabolites and statistical analysis

MassHunter Workstation software (version B7.00 Qualitative Analysis, Agilent Technologies) was used to process all data obtained by LC–QTOF in data-dependent acquisition MSMS mode. Treatment of raw data files started by extraction of potential molecular features (MFs) with the suited algorithm included in the software. For this purpose, the extraction algorithm considered all ions exceeding 1500 counts for both polarities with a single charge state. This cut-off value was established taking into account the chromatographic background noise. Additionally, the algorithm considered that an MF should have a valid isotopic distribution defined by two or more ions (with a peak spacing tolerance of *m/z* 0.0025, plus 7.0 ppm in mass accuracy). Ions and adducts formation in positive (+H, +Na, +K, +NH4) and negative ionization (–H, +HCOO, +Cl) modes, as well as neutral loss by dehydration were included to identify features corresponding to the same potential metabolite. Identification of metabolites was supported on MS and MS/MS information that was searched in the METLIN MS and MS/MS databases (http://metlin.scripps.edu), the Human Metabolome Database (HMDB, 3.6 version) and the LIPID MAPS website (http://www.lipidmaps.org, using in all cases the MFs obtained from the previous step. A database with all identified metabolites was used to perform a targeted compound extraction analysis using a tolerance window of 0.8 min and 6 ppm mass accuracy. This step was performed with

MassHunter software. A table with the peak area of all compounds identified in the different samples injected was obtained as a result.

Unknown Analysis software (version 7.0, Agilent Technologies) was used to unzip all data files obtained by GC–TOF/MS in full scan mode. Then, MassHunter software was used to process GC–TOF/MS data files. Treatment of raw data files started by deconvolution of chromatograms to obtain a list of MFs considered as potential compounds defined by the *m/z* value of one representative ion for each chromatographic peak and its RT. For this purpose, the deconvolution algorithm was applied to each sample by considering all ions exceeding 1500 counts for the absolute height parameter, with an accuracy error of 5 ppm and a window size factor of 150 units. The list of MFs obtained for each analysis was exported as data files in compound exchange format (.cef files). Tentative identification of compounds was performed by searching each mass spectrum in the NIST database (v.11) using the RI or RT value, respectively. The identification was firstly carried out by searching MS spectra on the NIST database. Only identifications with a match factor and a reverse match factor higher than 700 were considered as valid. The RI values included in the NIST database were also taken into account to support identifications. An RI calibration model was built by plotting the retention times obtained by analysis of the alkane standard mixture (C10 to C40 with an even number of carbons) with the chromatographic method used in this research and the RI values provided for each alkane by the NIST database. Then, the RI value was experimentally estimated for each identified compound by using the retention time and the calibration equation. The requirement to accept NIST identifications was that the difference between the experimental RI and the theoretical value provided by the NIST for each target compound should be below 100 units. The NIST database does not contain high resolution MS information as provided by the TOF detector. For this reason, a third step was included to validate identification of each compound by using high resolution MS. Thus, the molecular formula for the $[M]^+$ ion and the most intense fragments for each MFs should fit the NIST 11 identification by setting a cut-off value in mass accuracy of 10 ppm. A table with the peak area of all compounds identified in the different samples injected was obtained as a result.

MarvinSketch (v. 18.3.0) software from ChemAxon (http://www.chemaxon.com) was used for characterizing chemical structures of each compound by calculating polarizability ($Å^3$).

3. Results and discussion

3.1. Identification of metabolites in pig feces samples

The aim of this research was to maximize the detection coverage in the analysis of pig feces as a mandatory step to enhance the identification of metabolites in this sample. Previous studies dealing with feces metabolomics analysis have used frozen samples [7–9]. For this reason, frozen samples (see Section 2.1.) were selected to avoid physico-chemical alterations, particularly the loss of volatile components such as SCFAs [8,9], which may be of interest as potential biomarkers. Previous studies dealing with metabolomics analysis of fresh feces employed a simple MeOH extraction protocol to obtain fast and informative fecal fingerprints [10,11]. With the experimental plan adopted in this research, it was possible to tentatively identify a total number of 303 compounds by combination of all the extractants and analytical platforms employed, namely, LC–MS/MS and GC–MS. The complete list of identified metabolites in fecal samples is listed in Supplementary Tables 1 and 2, which present the compounds classified in chemical families and includes parameters supporting the identification. Figure 1 shows the distribution of identified compounds belonging to the main chemical families. As can be seen, fatty acids, conjugates and derivatives, amino acids and analogues, carboxylic acids and derivatives, bile acids, bilirubins and derivatives, and carbohydrates and derivatives were the families with a high number of identified compounds: 59, 31, 25, 21 and 19 metabolites, respectively. Among identified compounds it is also worth mentioning the presence of exogenous metabolites that are ascribed to the animal feed. These encompassed amino acids (ornithine, betaine), alkaloids (trigonelline, quinoline, dioscoretine), terpenic compounds (saponin, farnesol), vitamins (α-tocotrienol), prenol lipids (α-carotenal) and phenolic compounds (hydrocaffeic acid, apigenin, diosmetin, trihydroxyflavone).

The results of chemical families identified in this research agreed with other studies about fecal samples from human or animals and involving similar or complementary detection techniques [10,12–14]. Nevertheless, it is worth mentioning that there are few references for untargeted metabolomic analysis in pig fecal samples and most of them were focused on finding significant differences in individuals subjected to different feeds. Sun *et al.* (2015) reported that bile acids, lipids such as fatty acids and glycerophospholipids, amino acids, and flavonoid conjugates are the final dietary metabolites in fecal samples from pigs subjected to different feeds [12]. The authors used UPLC–MS/MS in high resolution mode to identify 90 significantly altered metabolites [12]. Sun *et al.* (2016) detected 92 metabolites involved in multiple biochemical processes in the hindgut of pigs [15] by GC–MS analysis.

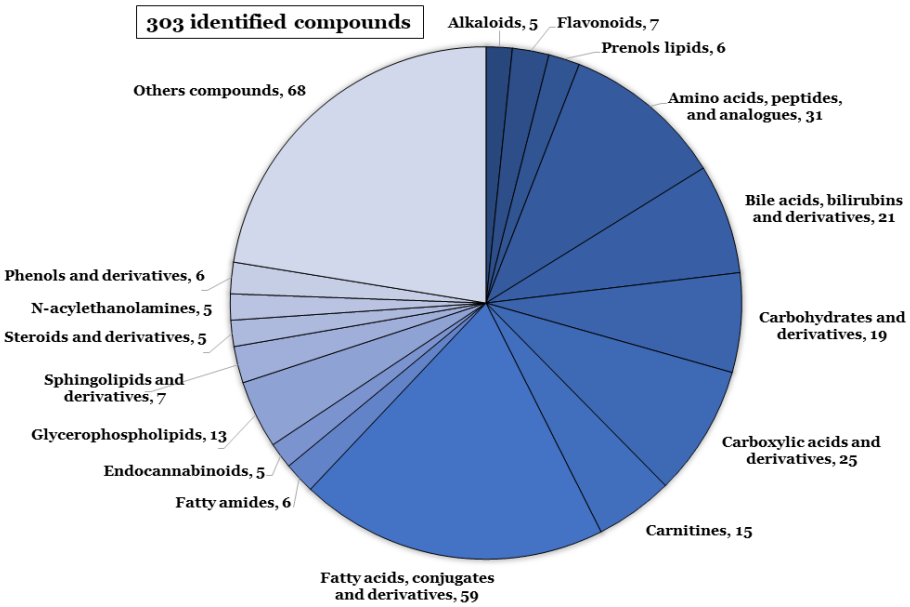


Figure 1. Chemical classification of all metabolites tentatively identified in pig fecal samples by all the experiments performed in this research.

These compounds encompassed organic acids, amino acids, fatty acids, saccharides, amines and lipids. Concerning human fecal samples, most of the existing studies have used a unique detection technique. For this reason, identification of

compounds has been frequently limited to amino acids, carboxylic acids, carbohydrates and phenols [8,15,16].

3.2. Influence of extractants on detection coverage for untargeted analysis by LC–MS/MS

A total of 189 metabolites were tentatively identified when combining all the extracts from feces samples analyzed by LC–QTOF MS/MS (See Supplementary Tables 1 and 3). Most compounds identified in feces samples pertain to chemical families previously described in the literature [8,9,11] such as fatty acids and derivatives, carnitines, bile acids and amino acids (Supplementary Table 1). It is also worth mentioning the presence of exogenous metabolites, particularly alkaloids such as quinoline or flavonoids such as apigenin.

The polar extracts, obtained with water, MeOH and 1:1 water/MeOH provided a similar detection pattern with tentative identification of 123, 146 and 146 compounds, respectively, as Figure 2A shows. The use of MeOH and water/MeOH as extractants evidently improved the detection of mid-polar and non-polar families of metabolites, which were scarcely detected or non-detected in the water extract. On the other hand, subtle differences were observed in the comparison between MeOH and water/MeOH, mainly affecting to bilirubin, glycosylated bilirubin, tyrosine, glutamic acid, propionylcarnitine, dodecanoylcarnitine, flavidin and dimethylPGD$_2$ (Supplementary Table 3).

Concerning the non-polar extractants, 168 and 158 compounds were tentatively identified in ethyl acetate and dichloromethane, respectively, whereas 113 compounds were found in hexane extracts. As Figure 2B reveals, ethyl acetate and dichloromethane offered a complete view of the non-polar fraction. However, ethyl acetate allowed detecting 12 exclusive compounds, which encompassed phospholipids —LysoPC(14:0), LysoPC(22:6) and LysoPE(16:0)—, and flavonoids such as apigenin, daidzein and genistein (Supplementary Table 3). On the other hand, two compounds, 5-methylcytidine and N(1-deoxy-1-fructosyl)leucine, were only found in the dichloromethane extract.

As previously mentioned, the most frequently used extractant for analysis of feces samples is MeOH [8]. In this research, a complementarity effect was observed by comparing the identification coverage of polar and non-polar extractants. Supplementary Figure 2 reports a heat map that shows the relative abundances of detected compounds ordered by retention time. As expected, polar extractants led to a more sensitive detection of metabolites eluting in the first part of the chromatogram, while non-polar extractants improved the detection of less polar compounds.

Figure 2. Venn diagrams comparing the compounds tentatively identified in the extracts from pig fecal samples by LC–QTOF MS/MS: (A) extracts from polar solvents, (B) extracts from non-polar solvents and (C) comparison between the polar and non-polar extractants that offered the best performance.

According to these results, a dual extraction with ethyl acetate and MeOH or MeOH/water should be recommended to fraction the metabolome of feces samples as an approach to increase the number of identified compounds. Comparing MeOH and

ethyl acetate, most compounds, 126, were commonly detected with both extractants (Figure 2C). However, 42 and 20 compounds were only detected in ethyl acetate and MeOH extracts, respectively. Among compounds exclusively identified in ethyl acetate extracts it is worth mentioning the presence of particular families of lipids such as fatty acids and derivatives, prenol lipids, sphingolipids and bile acids (Supplementary Table 3).

3.3. Influence of extractants and derivatization protocols on detection coverage for feces untargeted analysis by GC–MS

A total of 126 compounds were tentatively identified in feces samples by GC–TOF/MS analysis combining the 6 extractants and two derivatization protocols. Supplementary Table 4 lists the metabolites detected with each protocol. As can be seen in Supplementary Table 2, these compounds pertain mainly to the following families: carbohydrates (26 metabolites), amino acids (23), fatty acids (18), carboxylic acids (16) and SCFAs (6). As Figure 3 illustrates, all polar extractants were characterized by a very similar profile of compounds, with slight differences. Water/MeOH extractant allowed detecting all metabolites as compared with water and MeOH when they were used independently.

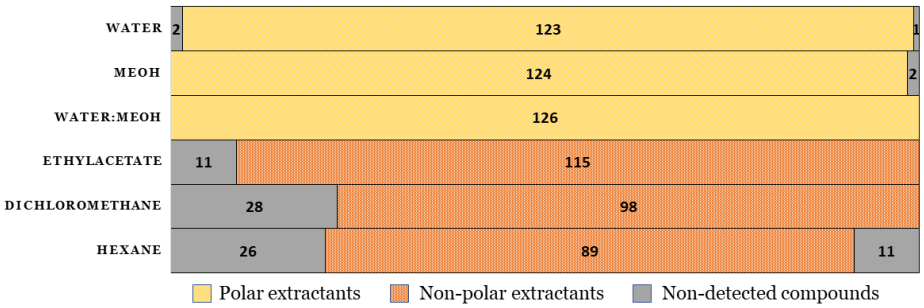


Figure 3. Bar diagram comparing the compounds tentatively identified by GC–TOF/MS in the six extracts obtained from pig fecal samples. The horizontal axis represents the compounds tentatively identified so that when the line corresponding to each extractant is colored (orange or yellow for non-polar or polar extractants, respectively) when the compound was detected in the extractant and grey when not. The number refers to the number of metabolites included in each color block.

Concerning non-polar extractants, ethyl acetate presented a wider coverage as compared to dichloromethane and hexane, since the former allowed the detection of some families of polar metabolites such as amino acids (homoserine, phenylalanine and leucine), carbohydrates (allofuranose, allopyranose, myoinositol, etc.) and carboxylic acids (2-aminobutyric acid). Nevertheless, ethyl acetate did not improve detection coverage as compared to the hydroalcoholic extractant. In fact, all compounds identified in the ethyl acetate extract were also found in the hydroalcoholic extract. On the other hand, amino acids such as methionine, histidine and proline were exclusively found in the latter (Supplementary Table 4). Therefore, the extraction with non-polar solvents does not contribute to expand the detection coverage by GC–MS analysis in contrast to LC–MS/MS.

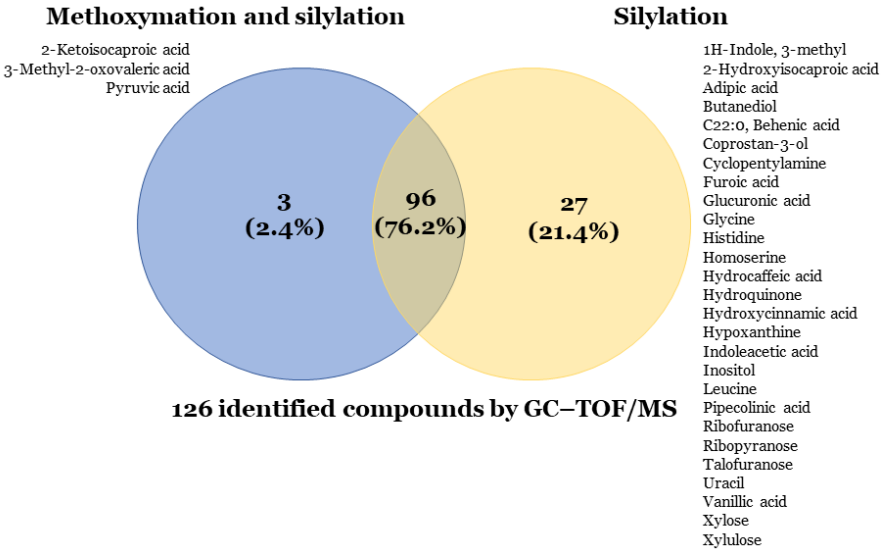


Figure 4. Venn diagram comparing the number of compounds identified in pig fecal extracts after derivatization with and without methoxymation prior to silylation.

Derivatization is undoubtedly a crucial step in analytical protocols dealing with GC–MS. The most common derivatization protocol implemented in untargeted metabolomic analysis by GC–MS consists of a methoxymation followed by silylation [15,18,19]. Methoxymation protects ketones and aldehydes by conversion to

methoxyamine groups [20]. However, not all aldehyde or ketone groups are protectable; in fact, carbonyl groups adjacent to heteroatoms, such as nitrogen or oxygen, are not electropositive enough for nucleophilic attack by the methoxyamine reagent. Therefore, the methoxymation step does not alter some kind of metabolites, thus being optional or unnecessary [21,22]. The removal of this step not only reduces the derivatization time, but also gives more stability to the final products. Thus, two derivatization protocols were considered in this study: silylation with and without methoxymation.

Figure 4 shows the Venn diagram comparing the compounds tentatively identified by both derivatization protocols, which are listed in Supplementary Table 2. As can be seen, 96 metabolites were commonly detected with both protocols, but some specific differences were found. Among the main families commonly detected with both derivatization processes, it is worth mentioning amino acids, carbohydrates, fatty acids and carboxylic acids. These results are in agreement with other studies previously reported in human fecal samples [17]. The inclusion of a methoxymation step only provided the tentative identification of three analytes that were not found when this step was omitted. These were three α-keto acids, particularly, 2-ketoisocaproic acid, 3-methyl-2-oxovaleric acid and pyruvic acid, which require protection of ketones to make possible their detection. On the other hand, 27 metabolites were only identified when methoxymation was not carried out. Therefore, it is evident that the information level attained without methoxymation is higher than that achieved when this step is implemented. Nevertheless, it is worth mentioning that α-keto acids are important metabolites related to microbiome status and, in some cases, their detection could be crucial. For this reason, depending on the objective of the research, one or two derivatization protocols may be required to cover the metabolites of interest.

3.4. Comparison between feces analysis by LC–QTOF MS/MS and GC–TOF/MS

As commented above, 303 compounds were tentatively identified in extracts from pig fecal samples by using two analytical platforms, LC–MS/MS and GC–MS, both in high resolution mode. As Figure 5A shows, the LC–QTOF MS/MS platform allowed identifying a higher number of metabolites as compared to the GC–TOF/MS approach, 189 *vs.* 126 compounds, respectively. However, the Venn diagram proved that both

platforms are complementary for untargeted analysis of pig feces. In fact, only 12 compounds were commonly identified with both platforms. These metabolites were essentially free fatty acids, lactic acid and some amino acids, including aromatic amino acids such as phenylalanine and tyrosine. The common fatty acids (palmitic acid, stearic acid, linoleic acid, oleic acid and behenic acid) were also the most concentrated fatty acids in fecal samples from pigs. These common metabolites were also identified by Sun *et al.* (2016) in fecal samples from pigs [15].

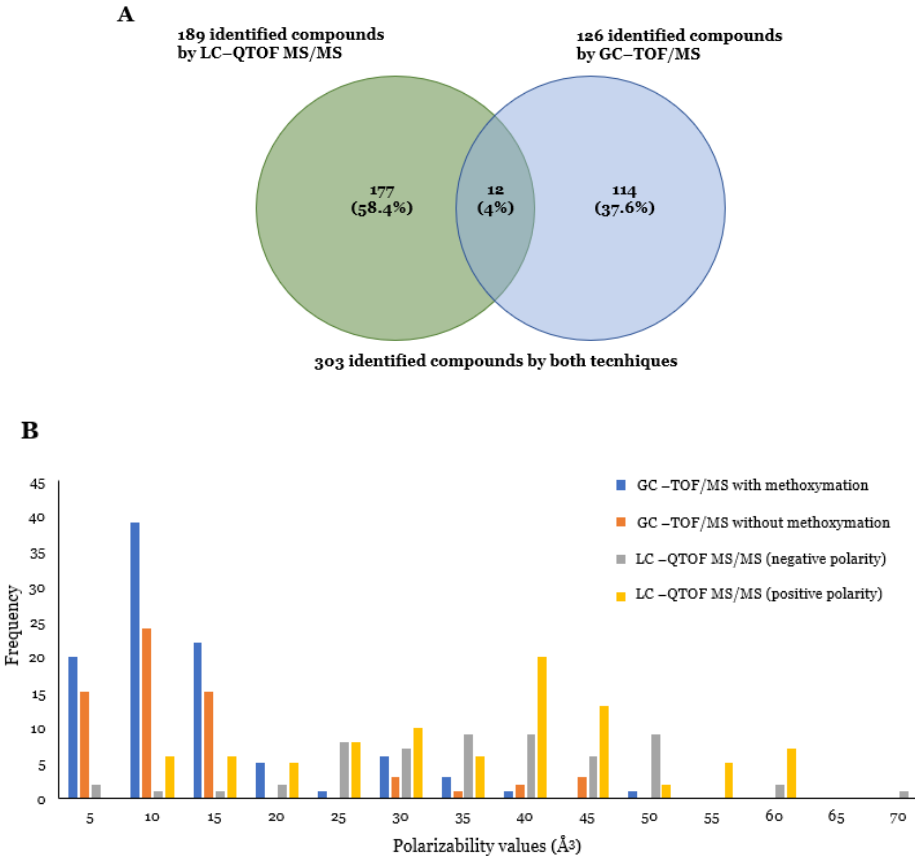


Figure 5. (A) Venn diagram comparing the number of compounds identified in fecal extracts by LC–QTOF MS/MS and GC–TOF/MS. (B) Histogram representing the number of compounds tentatively identified by LC–QTOF MS/MS in the two ionization modes and GC–TOF/MS with the two derivatization protocols grouped by polarizability values (Å3).

On the other hand, numerous chemical families were exclusively detected with one of the platforms. Thus, carbohydrates were only identified by GC–TOF/MS, whereas carnitines, bile acids, bilirubin, alkaloids and prenol lipids were only found by LC–QTOF MS/MS. The complementarity of LC–MS/MS and GC–MS, both in high resolution mode, is also revealed in Figure 5B, which plots the number of identified compounds *vs.* polarizability (Å^3) with each platform, namely LC–MS/MS in the two ionization modes and GC–MS with the two derivatization protocols. The polarizability profile enabled to discriminate two complementary areas, one for low polarizability compounds (between 5 and 15 Å^3), with preferred detection of metabolites by GC–MS, and a second area with high polarizability (between 25 and 60 Å^3), which included metabolites preferentially identified by LC–MS/MS. With these premises, it would be recommended to combine the two platforms to improve the identification coverage of metabolites in pig fecal samples.

4. Conclusions

The influence of sample preparation on the identification coverage for pig fecal samples analysis by LC–MS/MS and GC–MS has been studied in this research. A total number of 303 compounds by combination of all the extractants and analytical platforms were tentatively identified, being the main families amino acids, fatty acids and derivatives, carbohydrates and carboxylic acids. According to the results obtained, it should be recommended the utilization of MeOH/water as extractant for GC–MS analysis, but for LC–MS/MS analysis the extracts obtained with MeOH or MeOH/water and ethyl acetate can lead to a significant increase of identified. Concerning the derivatization step, the implementation of methoximation prior to silylation provided the identification of three α-keto acids that are not detected with the other tested strategies. Thus, since these are important metabolites related to microbiome status, the implementation of a double derivatization strategy could be of interest for some studies. Concerning the complementarity of the two analytical platforms evaluated in this research, LC–MS/MS and GC–MS, their combined use allowed the identification of 303 metabolites with only 12 common metabolites. Thus, it is obvious that both

techniques should be complemented to obtain a comprehensive view of the pig feces metabolome.

Acknowledgments

This research has been financed jointly by the Spanish Ministerio de Economía y Competitividad (through the CTQ2015-68813-R and AGL2014-54089-R projects), and the Instituto de Salud Carlos III (through the PIE14/00005/1, FLEXI-MET project). The two projects are co-funded by European Regional Development Fund/European Social Fund ("Investing in your future"). M.A. López-Bascón thanks the Ministerio de Educación, Cultura, Deporte (MECD) for an FPU scholarship (FPU15/02373). H. Argüello is granted by the Spanish Ministerio de Economía y Competitividad (IJCI-2016-30795).

References

[1] G. Evangelopoulou, S. Kritas, G. Christodoulopoulos, A.R. Burriel, The commercial impact of pig salmonella spp. infections in border-free markets during an economic recession, Vet. World. 8 (2015) 257–272.

[2] R. Isaacson, H.B. Kim, The intestinal microbiome of the pig, Anim. Heal. Res. Rev. 13 (2012) 100–109.

[3] S. Devaraj, P. Hemarajata, J. Versalovic, The human gut microbiome and body metabolism: implications for obesity and diabetes, Clin. Chem. 59 (2014) 617–628.

[4] W.-L. Hao, Y.-K. Lee, Microflora of the gastrointestinal tract: a review, Methods Mol. Biol. 268 (2004) 491–502.

[5] A.K. Kosmides, K. Kamisoglu, S.E. Calvano, S.A. Corbett, I.P. Androulakis, Metabolomic fingerprinting: challenges and opportunities, Crit. Rev. Biomed. Eng. 41 (2013) 205–21.

[6] H. Cao, H. Huang, W. Xu, D. Chen, J. Yu, J. Li, L. Li, Fecal metabolome profiling of liver cirrhosis and hepatocellular carcinoma patients by ultra performance liquid chromatography – mass spectrometry, Anal. Chim. Acta 691 (2011) 68–75.

[7] S. Matysik, C.I. Le Roy, G. Liebisch, S.P. Claus, Metabolomics of fecal samples: a practical consideration, Trends Food Sci. Technol. 57 (2016) 244–255.

[8] O. Deda, H.G. Gika, I.D. Wilson, G.A. Theodoridis, An overview of fecal sample

preparation for global metabolic profiling, J. Pharm. Biomed. Anal. 113 (2015) 137–150.

[9] S. Lamichhane, C.C. Yde, M.S. Schmedes, H.M. Jensen, S. Meier, H.C. Bertram, Strategy for nuclear-magnetic-resonance-based metabolomics of human feces, Anal. Chem. 87 (2015) 5930–5937.

[10] N. Cesbron, A.L. Royer, Y. Guitton, A. Sydor, B. Le Bizec, G. Dervilly-Pinel, Optimization of fecal sample preparation for untargeted LC-HRMS based metabolomics, Metabolomics 13 (2017) 99.

[11] J. Su, Y. Ng, U. Ryan, R.D. Trengove, G.L. Maker, Development of an untargeted metabolomics method for the analysis of human faecal samples using Cryptosporidium-infected samples, Mol. Biochem. Parasitol. 185 (2012) 145–150.

[12] J. Sun, M. Monagas, S. Jang, A. Molokin, J.M. Harnly, J.F. Urban, G. Solano-aguilar, P. Chen, A high fat, high cholesterol diet leads to changes in metabolite patterns in pigs – A metabolomic study, Food Chem. 173 (2015) 171–178.

[13] O. Deda, A. Chrysovalantou, S. Fasoula, D. Palachanis, N. Raikos, G.A. Theodoridis, H.G. Gika, Sample preparation optimization in fecal metabolic profiling, J. Chromatogr. B 1047 (2017) 115–123.

[14] A. V. Melnik, R.R. Da Silva, E.R. Hyde, A.A. Aksenov, F. Vargas, A. Bouslimani, I. Protsyuk, A.K. Jarmusch, A. Tripathi, T. Alexandrov, R. Knight, P.C. Dorrestein, Coupling targeted and untargeted mass spectrometry for metabolome-microbiome-wide association studies of human fecal samples, Anal. Chem. 89 (2017) 7549–7559.

[15] Y. Sun, Y. Su, W. Zhu, Microbiome-metabolome responses in the cecum and colon of pig to a high resistant starch diet, Front. Microbiol. 7 (2016) 1–10.

[16] L. Cheng Phua, P. Koon Koh, P. Yean Cheah, H. Kiat Ho, E. Chun Yong Chan, Global gas chromatography/time-of-flight mass spectrometry (GC/TOFMS)-based metabonomic profiling of lyophilized human feces, J. Chromatogr. B 937 (2013) 103–113.

[17] X. Gao, E. Pujos-guillot, J. Sébédio, Development of a quantitative metabolomic approach to study clinical human fecal water metabolome based on trimethylsilylation derivatization and GC/MS analysis, Anal. Chem. 82 (2010) 6447–6456.

[18] K. Ponnusamy, J.N. Choi, J. Kim, S.Y. Lee, C.H. Lee, Microbial community and metabolomic comparison of irritable bowel syndrome faeces, J. Med. Microbiol. 60 (2011) 817–827.

[19] S. Mao, W. Huo, W. Zhu, Microbiome – metabolome analysis reveals unhealthy alterations in the composition and metabolism of ruminal microbiota with increasing dietary grain, Environ. Microbiol. 18 (2016) 525–541.

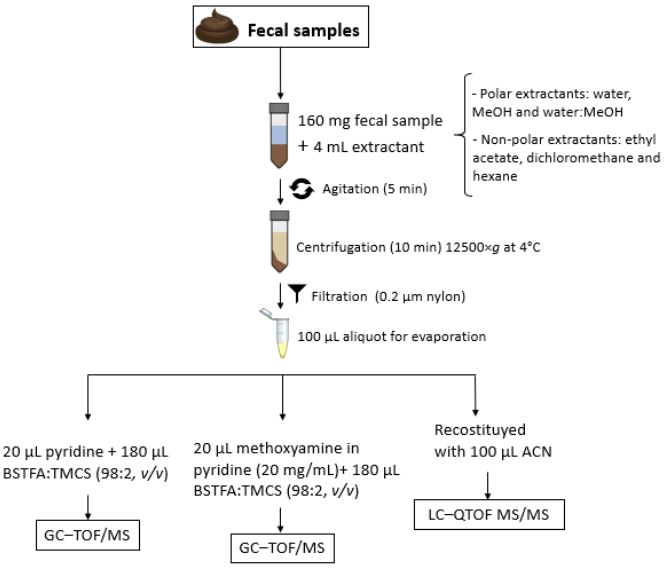
[20] O. Fiehn, Fiehn GC/MS metabolomics RTL library, User Guide Agilent Technologies (2013) 221.

[21] M. del M. Delgado-Povedano, M. Calderón-Santiago, F. Priego-capote, M.D Luque de Castro, Development of a method for enhancing metabolomics coverage of human sweat by gas chromatography e mass spectrometry in high resolution mode, Anal. Chim. Acta

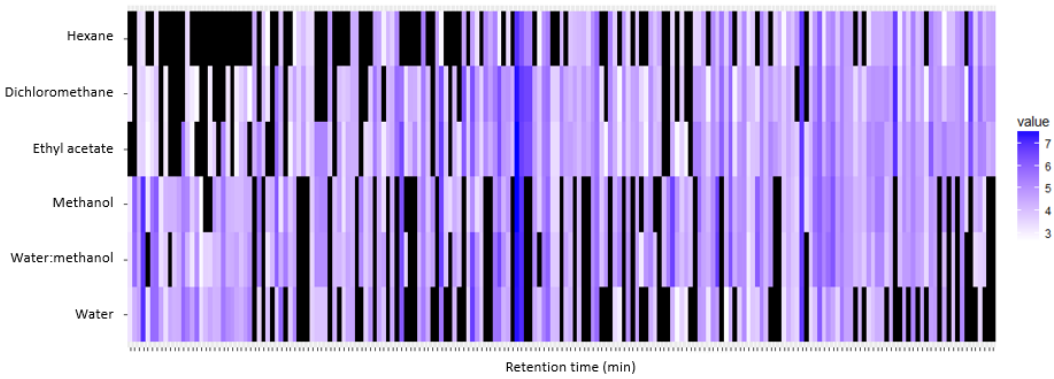
905 (2016)115–125.

[22] X. Gao, E. Pujos-Guillot, J.F. Martin, P. Galan, C. Juste, W. Jia, J.L. Sebedio, Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization, Anal. Biochem. 393 (2009) 163–175.

Supplementary information



Supplementary Figure 1. Experimental set-up designed to study the influence of sample preparation on the metabolomics analysis of pig feces samples.



Supplementary Figure 2. Heat map comparing relative concentrations of compounds tentatively identified in water, (1:1) water:methanol, methanol, ethyl acetate, dichloromethane and hexane extracts from fecal samples by LC–QTOF MS/MS. Color code indicates differences in normalized concentration for each metabolite.

Supplementary Table 1. List of metabolites tentatively identified in fecal samples by LC–QTOF MS/MS. Information about family, formula, chromatographic retention time, monoisotopic molecular weight, theorical and experimental *m/z* values for precursor ion as well as the characteristic product ions supporting the identification are included.

Compound	Familiy	Exogenous/ Endogenous	Formula	RT (min)	Monoisotopic molecular weight	Adduct	Precursor ion theoretical (m/z)	Precursor ion experimental (m/z)	Error (ppm)	Product ion (m/z)
2-Apo-beta-carotenal	Prenol lipids	Exogenous	C27H36O3	12.61	508.3705	$[M+H]^+$	509.3778	509.3772	-1.15	465.3531, 213.1301
11-Tetradecen-1-ol	Alkaloids	Exogenous	C14H28O	9.83	212.2140	$[M+H]^+$	213.2213	213.2217	1.99	197.2273, 71.0820
12-amino-dodecanoic acid	Fatty amides	Endogenous	C12H25NO2	8.29	215.1890	$[M+NH_4]^+$	233.2228	233.222	-3.53	198.1857, 170.1905
12-Ketodeoxycholic acid	Bile acids and derivatives	Endogenous	C24H34O4	12.70	386.2460	$[M+H]^-$	385.2387	385.2386	-0.32	327.2631, 111.0815
13-Docosenamide	Fatty amides	Endogenous	C22H43NO	11.63	337.3340	$[M+H]^+$	338.3413	338.3414	0.37	322.3451
13-Hydroxy-docosanoic acid	Hydroxy fatty acids	Endogenous	C22H44O3	11.73	356.3280	$[M+H]^+$	357.3353	357.3368	4.26	281.3201
1-Dodecene	Hydrocarbons and derivatives	Endogenous	C12H24	9.88	168.1880	$[M+NH_4]^+$	186.2218	186.2213	-2.68	85.1002, 57.0701
22-Oxo-docosanoate	Keto fatty acids	Endogenous	C22H42O3	11.73	354.3130	$[M+H]^+$	355.3203	355.3209	1.76	319.2987, 57.0332
24-keto-1,25-Dihydroxyvitamin D3	Vitamins and derivatives	Exogenous	C27H42O4	7.20	430.3070	$[M+H]^-$	429.2997	429.3002	1.11	351.2354
2-Aminoethylphosphocholate	Glycerophospholipids	Endogenous	C26H46NO7P	9.50	515.3000	$[M+H]^+$	516.3073	516.3069	-0.73	331.2649, 126.0319
3-(2-Hydroxyphenyl) propionic acid	Phenylpropanoic acids	Endogenous	C9H10O3	6.67	166.0640	$[M+H]^-$	165.0567	165.0564	-1.96	121.0671, 106.0426
3-Hydroxy-eicosanoic acid	Hydroxy fatty acids	Endogenous	C20H40O3	10.75	328.2970	$[M+H]^+$	329.3043	329.3032	-3.27	283.2989, 85.1063
3-tert-Butyl-5-methylcatechol	Phenols and derivatives	Exogenous	C11H16O2	7.76	180.1140	$[M+NH_4]^+$	198.1478	198.1474	-2.13	135.1172, 107.0857
5alpha,17alpha-Pregn-2-en-20-yn-17-ol acetate	Steroids and steroid derivatives	Endogenous	C23H32O2	13.03	340.2400	$[M+H]^+$	341.2473	341.2479	1.83	279.2031, 121.1017
5-Chola-8(14)11-dien-24-oic acid	Bile acids and derivatives	Endogenous	C24H36O2	9.69	356.2720	$[M+H]^+$	357.2793	357.2793	0.07	300.2071, 85.0297
5-Methylcytidine	Nucleosides, nucleotides, and analogues	Endogenous	C10H15N3O5	0.49	257.1030	$[M+H]^+$	258.1103	258.1099	-1.46	126.0678
8,9-EET	Hydroxy fatty acids	Endogenous	C20H32O3	11.40	320.2330	$[M+H]^-$	319.2257	319.2269	3.68	275.2403, 95.0821

Continuation Supplementary Table 1

C16:4, Hexadecatetraenoic acid	Fatty acids and conjugates	Endogenous	C16H24O2	9.59	248.1780	[M+H]-	247.1707	247.1700	-2.93	247.17
C17:1, Heptadecenoic acid	Fatty acids and conjugates	Endogenous	C17H32O2	10.09	268.2400	[M+H]-	267.2327	267.2322	-1.96	267.2322
C18 Sphinganine	Sphingolipids and derivatives	Endogenous	C18H39NO2	10.71	301.2980	[M+H]+	302.3053	302.3056	1.07	284.2924, 71.0857
C18:0, Stearic acid	Fatty acids and conjugates	Endogenous	C18H36O2	15.37	284.2710	[M+H]-	283.2637	283.2639	0.62	283.2639
C18:1, Oleic acid	Fatty acids and conjugates	Endogenous	C18H34O2	12.86	282.2559	[M+H]-	281.2486	281.2492	2.12	281.2492
C18:2, Linoleic acid	Fatty acids and conjugates	Endogenous	C18H32O2	13.35	280.2400	[M+H]-	279.2327	279.2335	2.78	279.2335, 261.9984
C20:1, Eicosenoic acid	Fatty acids and conjugates	Endogenous	C20H38O2	14.16	310.2870	[M+H]-	309.2797	309.2785	-3.96	309.2785
C20:4, Eicosatetraenoic acid	Fatty acids and conjugates	Endogenous	C20H32O2	13.20	304.2400	[M+H]-	303.2327	303.2331	1.24	259.2487, 177.1643
C22:0, Behenic acid	Fatty acids and conjugates	Endogenous	C22H44O2	12.79	340.3340	[M+H]-	339.326724	339.3257	-3.02	339.3257
C22:3, Docosatrienoic acid	Fatty acids and conjugates	Endogenous	C22H38O2	13.13	334.2870	[M+H]-	333.2797	333.2789	-2.47	333.2789
C22:4, Docosatetraenoic acid	Fatty acids and conjugates	Endogenous	C22H36O2	11.89	332.2710	[M+H]-	331.2637	331.2634	-0.98	331.2634
C22:5, Docosapentaenoic acid	Fatty acids and conjugates	Endogenous	C22H34O2	13.59	330.2560	[M+H]-	329.2487	329.2485	-0.68	329.2485
C22:6, Docosahexaenoic acid	Fatty acids and conjugates	Endogenous	C22H32O2	13.03	328.2400	[M+H]-	327.2327	327.2340	3.90	283.2461, 59.0134
C24:5, Tetracosapentaenoic acid	Fatty acids and conjugates	Endogenous	C24H38O2	13.88	358.2850	[M+H]-	357.2777	357.2769	-2.31	357.2769
C4:0, Butanoic acid	Short chain fatty acids and derivatives	Endogenous	C4H8O2	0.55	88.0530	[M+H]-	87.0457	87.0454	-3.45	87.0454
Carnitine	Carnitines	Endogenous	C7H15NO3	0.50	161.1050	[M+H]+	162.1123	162.1119	-2.32	103.0389, 60.0812
Cholestane-tetrol	Cholesterols and derivatives	Endogenous	C27H48O4	10.37	436.3550	[M+H]+	437.3623	437.3628	1.20	403.3215, 361.3106
Cholesterol sulfate	Cholesterols and derivatives	Endogenous	C27H46O4S	13.96	466.3120	[M+H]-	465.3047	465.3051	0.81	96.9601
Cholic acid	Bile acids and derivatives	Endogenous	C24H40O5	9.10	408.2880	[M+H]-	407.2807	407.2809	0.43	343.2619
Choline	Cholines	Endogenous	C5H13NO	0.50	103.1000	[M+H]+	104.1073	104.1074	1.19	60.0828, 45.0332
Corticosterone	Steroids and steroid derivatives	Endogenous	C21H30O4	10.88	346.2140	[M+H]+	347.2213	347.2205	-2.23	311.2022, 59.0135
Cortisone	Steroids and steroid derivatives	Endogenous	C21H28O5	8.93	360.1910	[M+H]-	359.1837	359.1835	-0.62	301.1822, 203.1075

Continuation Supplementary Table 1

Creatine	Amino acids, peptides, and analogues	Endogenous	C4H9N3O2	0.52	131.0700	[M+H]+	132.0773	132.0777	3.21	72.0801, 44.0503
Daidzein	Flavonoids	Exogenous	C15H10O4	7.94	254.0580	[M+H]+	255.0653	255.0654	0.49	227.0709, 199.075
Decanoylcarnitine	Carnitines	Endogenous	C17H33NO4	11.48	315.2410	[M+H]+	316.2483	316.2474	-2.77	144.1023, 85.0297
Decanoylglycine	Carboxylic acids and derivatives	Endogenous	C12H23NO3	8.23	229.1670	[M+H]+	230.1743	230.1751	3.58	230.1751, 59.0131
Decenedioic acid	Dicarboxylic acids and derivatives	Endogenous	C10H16O4	8.47	200.1050	[M+H]-	199.0977	199.0982	2.39	183.1017, 73.0281
Dehydroepiandrosterone-glucuronide	Steroids and steroid derivatives	Endogenous	C25H36O8	9.56	464.241	[M+H]+	465.2483	465.2491	1.77	447.2360, 275.2418
Dehydrophytosphingosine	Sphingolipids and derivatives	Endogenous	C18H37NO3	10.25	315.2770	[M+H]+	316.2843	316.284	-0.87	298.2749
Deoxycholic acid	Bile acids and derivatives	Endogenous	C24H40O4	9.47	392.2930	[M+H]-	391.2857	391.2846	-2.87	345.2779
Dicyclohexylurea	Organic carbonic acids and derivatives	Endogenous	C13H24N2O	9.32	224.1890	[M+H]+	225.1963	225.1964	0.55	143.1189, 100.1101
DiHETE	Eicosanoids	Endogenous	C20H32O4	11.70	336.2330	[M+H]-	335.2257	335.2262	1.42	219.1523, 71.0123
DiHODE isomer 1	Hydroperoxy fatty acids	Endogenous	C18H32O4	9.33	312.2300	[M+H]-	311.2227	311.2219	-2.65	235.2068, 121.1007
DiHODE isomer 2	Hydroperoxy fatty acids	Endogenous	C18H32O3	10.50	312.2300	[M+H]-	311.2227	311.2238	3.46	235.2062, 111.1167
Dihomo-gamma-Linolenoyl ethanolamide	N-acylethanolamines	Endogenous	C22H39NO2	9.47	349.2970	[M+H]-	348.2897	348.2881	-4.66	348.2901, 318.2503
Dihydrobiliverdin	Bilirubins	Endogenous	C33H36N4O6	16.49	584.2630	[M+H]+	585.2703	585.2702	-0.13	299.1399
Dihydroxy-5b-pregnane-11,20-dione	Steroids and steroid derivatives	Endogenous	C21H32O4	8.22	348.2300	[M+H]+	349.2373	349.2370	-0.79	273.2215, 73.0278
Dihydroxystearic acid	Hydroxy fatty acids	Endogenous	C18H36O4	10.32	316.2614	[M+H]-	315.2541	315.2549	2.59	297.2441, 141.1268
Dimethyl prostaglandin D2	Eicosanoids and derivatives	Endogenous	C20H32O5	10.87	352.2280	[M+H]-	351.2207	351.2205	-0.64	300.2432, 158.0735
Dimethylnonanoyl carnitine	Carnitines	Endogenous	C18H35NO4	10.07	329.2560	[M+H]+	330.2633	330.2651	5.52	144.012
Dioscoretine	Alkaloids	Exogenous	C13H23NO3	10.42	241.1670	[M+H]-	240.1597	240.1602	1.98	196.1755, 87.0477
Diosmetin	Flavonoids	Exogenous	C16H12O6	8.52	300.0630	[M+H]-	299.0557	299.0550	-2.42	284.0304
DL-Benzylsuccinic acid	Phenylpropanoic acids	Endogenous	C11H12O4	6.12	208.0740	[M+H]-	207.0667	207.0660	-3.50	163.0744, 91.0597
Docosanamide	Alkaloids	Exogenous	C22H45NO	17.76	339.3500	[M+H]+	340.3573	340.3568	-1.40	261.976

Continuation Supplementary Table 1

Docosanedioic acid	Dicarboxylic acids and derivatives	Endogenous	C22H42O4	11.67	370.3080	$[M+H]^+$	371.3153	371.3162	2.49	327.3260, 87.0412
Dodecanoylcarnitine	Carnitines	Endogenous	C19H37NO4	9.84	343.2750	$[M+H]^+$	344.2823	344.2819	-1.09	183.1754, 144.1014
Eicosanoylethanolamine isomer 1	N-acylethanolamines	Endogenous	C22H45NO2	11.97	355.3460	$[M+H]^+$	356.3533	356.354	2.03	324.3261, 71.0852
Eicosanoylethanolamine isomer 2	N-acylethanolamines	Endogenous	C22H45NO2	13.39	355.3450	$[M+H]^+$	356.3523	356.3519	-1.06	309.3023, 71.0855
Eicosenoic acid 14-hydroxy	Hydroxy fatty acids	Endogenous	C20H38O3	12.66	326.2820	$[M+NH_4]^+$	344.3158	344.3149	-2.68	283.2831, 57.0711
Epoxy-9Z15Z-octadecatrienoic acid	Hydroperoxy fatty acids	Endogenous	C18H28O3	10.36	292.2020	$[M+NH_4]^+$	310.2358	310.2355	-1.04	275.2014, 105.0697
Farnesol	Prenol lipids	Exogenous	C15H26O	9.44	222.1990	$[M+NH_4]^+$	240.2328	240.2325	-1.34	223.2043, 95.0859
Flavidin	Flavonoids	Exogenous	C15H12O3	0.50	240.0770	$[M+H]^+$	241.0843	241.0844	0.51	194.0503
Ganoderiol	Prenol lipids	Exogenous	C31H50O5	12.43	502.3630	$[M+NH_4]^+$	520.3968	520.3948	-3.89	453.3366, 309.2210
Genistein	Flavonoids	Exogenous	C15H10O5	9.25	270.0530	$[M+H]^-$	269.0457	269.0464	2.51	224.0483, 133.0268
Gingerol	Phenols and derivatives	Exogenous	C17H26O4	9.58	294.1830	$[M+H]^+$	295.1903	295.1910	2.45	137.0602
Glutamic acid	Amino acids, peptides, and analogues	Endogenous	C5H9NO4	0.54	147.0540	$[M+H]^-$	146.0467	146.0462	-3.59	102.0561
Glycochenodeoxycholic acid	Bile acids and derivatives	Endogenous	C26H43NO5	8.73	449.3150	$[M+H]^-$	448.3077	448.3062	-3.40	430.3040, 74.0265
Glycocholic acid	Bile acids and derivatives	Endogenous	C26H43NO6	6.91	465.3070	$[M+H]^+$	466.3143	466.3156	2.84	430.3085
Glycodeoxycholic acid	Bile acids and derivatives	Endogenous	C26H43NO5	9.39	449.3141	$[M+H]^-$	448.3068	448.3051	-3.90	430.2915, 76.0383
Glycoursodeoxycholic acid	Bile acids and derivatives	Endogenous	C26H43NO5	8.47	449.3141	$[M+H]^-$	448.3068	448.3075	1.46	373.2731, 76.0405
HETE-G	Hydroxy fatty acids	Endogenous	C23H38O5	11.09	394.2690	$[M+H]^+$	395.2763	395.2762	-0.19	360.2654, 111.1161
Hexadec-2-enoyl carnitine	Carnitines	Endogenous	C23H43NO4	11.41	397.3190	$[M+NH_4]^+$	415.3528	415.3524	-1.02	144.1019
Hexadecanedioic acid mono-carnitine ester	Carnitines	Endogenous	C23H43NO6	7.75	429.3100	$[M+NH_4]^+$	447.3438	447.3442	0.84	144.1021
Hexadecylamine	Amine derivative	Endogenous	C16H35N	11.88	241.2770	$[M+H]^+$	242.2843	242.2851	3.40	85.1003, 71.0852
HODE	Hydroperoxy fatty acids	Endogenous	C18H32O3	10.50	296.2350	$[M+H]^-$	295.2277	295.2276	-0.42	277.2175, 171.1024

Continuation Supplementary Table 1

Homotyrosine	Amino acids, peptides, and analogues	Endogenous	$C_{10}H_{13}NO_3$	10.20	195.0900	$[M+H]^-$	194.0827	194.0822	-2.70	107.0495, 95.0500
HpODE	Hydroperoxy fatty acids	Endogenous	$C_{18}H_{32}O_4$	10.07	312.2300	$[M+NH_4]^+$	330.2638	330.2629	-2.79	221.1519, 71.0859
Hydroxy-21-nor-20-oxavitamin D_3	Vitamins and derivatives	Exogenous	$C_{25}H_{40}O_3$	14.66	388.2960	$[M+NH_4]^+$	406.3298	406.3294	-1.04	221.1862, 71.0859
Hydroxy-cholanoic acid	Bile acids and derivatives	Endogenous	$C_{24}H_{40}O_3$	11.72	376.2950	$[M+NH_4]^+$	394.3288	394.3278	-2.59	253.2073, 73.0287
Hydroxydodecanoic acid	Hydroxy acids and derivatives	Endogenous	$C_{12}H_{24}O_3$	7.94	216.1720	$[M+NH_4]^+$	234.2058	234.206	0.76	186.1868, 74.0594
Hydroxydodecanoyl carnitine	Carnitines	Endogenous	$C_{17}H_{33}NO_5$	7.03	331.2330	$[M+H]^+$	332.2403	332.2405	0.67	144.1025
Hydroxyhexadecanoic acid	Hydroxy fatty acids	Endogenous	$C_{16}H_{32}O_3$	12.71	272.2350	$[M+H]^-$	271.2277	271.2272	-1.93	225.2216
Hydroxymyristic acid	Hydroxy fatty acids	Endogenous	$C_{16}H_{26}O_3$	9.48	244.2038	$[M+H]^-$	243.1965	243.1972	2.78	193.1513
Hydroxymyristoyl carnitine	Carnitines	Endogenous	$C_{21}H_{41}NO_4$	11.11	371.3040	$[M+H]^+$	372.3113	372.3121	2.21	144.1027, 85.0301
Hydroxystearic acid	Hydroxy fatty acids	Endogenous	$C_{18}H_{36}O_3$	9.83	300.2660	$[M+H]^-$	299.2587	299.2594	2.26	281.2456
Hypoxanthine	Purines and purine derivatives	Endogenous	$C_5H_4N_4O$	0.55	136.0380	$[M+H]^+$	137.0453	137.0452	-0.55	119.0339, 94.0397
Indoleacetic acid	Indoles and derivatives	Endogenous	$C_{10}H_9NO_2$	7.52	175.0630	$[M+H]^+$	176.0703	176.0711	4.68	149.0388, 133.0275, 105.0331
Isokobusone	Alcohols and polyols	Endogenous	$C_{14}H_{22}O_2$	9.58	222.1620	$[M+H]^+$	223.1693	223.1689	-1.68	205.1594
ketocholesterol	Cholesterols and derivatives	Endogenous	$C_{27}H_{44}O_2$	18.19	400.3340	$[M+H]^-$	399.3267	399.3278	2.69	381.3142
ketosphingosine	Sphingolipids and derivatives	Endogenous	$C_{18}H_{35}NO_2$	11.03	299.2830	$[M+H]^+$	300.2903	300.2901	-0.59	282.2801, 262.2518
Ketostearic acid	Keto fatty acids	Endogenous	$C_{18}H_{34}O_3$	11.39	298.2510	$[M+NH_4]^+$	316.2848	316.284	-2.60	256.2642, 282.2579
Lactic acid	Hydroxy acids and derivatives	Endogenous	$C_3H_6O_3$	0.55	90.0320	$[M+H]^-$	89.0247	89.0251	3.66	45.00001, 43.0191
Leucine	Amino acids, peptides, and analogues	Endogenous	$C_6H_{13}NO_2$	0.51	131.0946	$[M+H]^+$	132.1019	132.102	0.72	86.0954, 44.0489
Linoelaidyl carnitine	Carnitines	Endogenous	$C_{25}H_{45}NO_4$	12.17	423.3390	$[M+H]^+$	424.3463	424.3472	2.18	144.1021
Linolenyl carnitine	Carnitines	Endogenous	$C_{25}H_{45}NO_3$	11.46	407.3390	$[M+H]^+$	408.3463	408.3461	-0.43	144.1025
Linoleoyl ethanolamide	N-acylethanolamines	Endogenous	$C_{20}H_{37}NO_2$	12.35	323.2830	$[M+H]^+$	324.2903	324.29	-0.85	306.2799, 62.0601

Continuation Supplementary Table 1

LysoPC(14:0)	Glycerophospholipids	Endogenous	C22H46NO7P	10.41	467.3010	$[M+H]^+$	468.3083	468.3085	0.48	184.0747, 104.1068
LysoPC(16:0)	Glycerophospholipids	Endogenous	C24H50NO7P	11.28	495.3330	$[M+H]^+$	496.3403	496.34	-0.56	184.0732, 104.1065
LysoPC(16:1)	Glycerophospholipids	Endogenous	C24H48NO7P	10.66	493.3170	$[M+H]^+$	494.3243	494.3249	1.26	184.0738, 104.1067
LysoPC(18:1)	Glycerophospholipids	Endogenous	C26H52NO7P	11.51	521.3490	$[M+H]^+$	522.3563	522.3552	-2.06	184.0747, 104.1068
LysoPC(18:2)	Glycerophospholipids	Endogenous	C26H50NO7P	10.94	519.3330	$[M+H]^+$	520.3403	520.3398	-0.91	184.0747, 104.1068
LysoPC(18:3)	Glycerophospholipids	Endogenous	C26H48NO7P	10.49	517.3170	$[M+H]^+$	518.3243	518.3231	-2.27	184.0741, 104.1068
LysoPC(20:5)	Glycerophospholipids	Endogenous	C28H48NO7P	10.93	541.3150	$[M+H]^+$	542.3223	542.3223	0.04	184.0737, 104.1069
LysoPC(22:5)	Glycerophospholipids	Endogenous	C30H52NO7P	8.81	569.3520	$[M+H]^+$	570.3593	570.3586	-1.19	184.0749, 104.1066
LysoPC(22:6)	Glycerophospholipids	Endogenous	C30H50NO7P	10.94	567.3320	$[M+H]^+$	568.3393	568.3382	-1.89	184.0742, 104.1063
LysoPE(16:0)	Glycerophospholipids	Endogenous	C21H44NO7P	10.96	453.2860	$[M+H]^-$	452.2787	452.2784	-0.72	255.233, 196.0417
LysoPE(18:0)	Glycerophospholipids	Endogenous	C23H48NO7P	10.99	481.3170	$[M+H]^-$	480.3097	480.3120	4.74	283.2644, 196.0423
Mesobilirubinogen	Bile acids and derivatives	Endogenous	C33H44N4O6	7.97	592.3260	[M+H]+	593.3333	593.3331	-0.30	575.3124, 468.2462
Methyl cinnamate	Cinnamics acids and derivatives	Endogenous	C10H10O2	8.21	162.0680	$[M+H]^+$	163.0753	163.0751	-1.08	91.0539, 77.0389
MG(18:1)	Acyglycerides	Endogenous	C21H40O4	11.17	356.2930	$[M+H]^+$	357.3003	357.2999	-1.05	339.2901, 75.0470
MG(20:1)	Acyglycerides	Endogenous	C23H44O4	12.19	384.3240	$[M+H]^+$	385.3313	385.3311	-0.46	275.2754, 75.0431
MG(20:2)	Acyglycerides	Endogenous	C23H42O4	12.45	382.3080	$[M+NH_4]^+$	400.3418	400.3416	-0.56	75.0465
N(1-Deoxy-1-fructosyl)leucine	Carboxylic acids and derivatives	Exogenous	C12H23NO7	0.56	293.1460	$[M+H]^+$	294.1533	294.1546	4.50	230.1389, 144.1035
N2-Fructopyranosaylarginine	Carboxylic acids and derivatives	Exogenous	C12H24N4O7	0.56	336.1660	$[M+H]^+$	337.1733	337.1741	2.44	232.1175, 57.0743
N-Docosahexaenoyl GABA	Amino acids, peptides, and analogues	Endogenous	C26H39NO3	8.87	413.2920	$[M+H]^+$	414.2993	414.3001	1.99	396.2893, 311.2367
N-Stearoyl GABA	Amino acids, peptides, and analogues	Endogenous	C22H43NO3	12.96	369.3240	$[M+H]^+$	370.3313	370.3318	1.41	352.2984, 227.1283
Nutriacholic acid	Bile acids and derivatives	Endogenous	C24H38O4	9.29	390.2770	$[M+H]^+$	391.2843	391.2845	0.57	313.2524, 95.0871

Continuation Supplementary Table 1

Oleamide	Fatty amides	Endogenous	C18H35NO	13.62	281.2720	$[M+H]^+$	282.2793	282.2792	-0.27	265.2553, 177.1636
Oleanolic acid	Triterpene	Exogenous	C30H48O3	14.71	456.3600	$[M+H]^-$	455.3527	455.3531	0.83	455.3565
Oleoyl ethyl amide	Fatty amides	Endogenous	C20H39NO	15.67	309.3030	$[M+H]^+$	310.3103	310.3099	-1.21	128.1070,72.0442
Oxohexadecanoic acid isomer 1	Keto fatty acids	Endogenous	C16H30O3	9.56	270.2200	$[M+NH_4]^+$	288.2538	288.2538	-0.08	271.2293, 85.1017
Oxohexadecanoic acid isomer 2	Keto fatty acids	Endogenous	C16H30O3	10.36	270.2200	$[M+NH_4]^+$	288.2538	288.2536	-0.77	271.2305, 85.1002
Oxotetradecanoic acid	Keto fatty acids	Endogenous	C14H26O3	6.41	242.1880	$[M+H]^-$	241.1807	241.1799	-3.42	195.1757, 59.0125
Palmitaldehyde	Fatty aldehydes	Endogenous	C16H32O	10.17	240.2450	$[M+H]^+$	258.2788	258.2789	0.30	240.2672, 95.0851
Palmitoyl dopamine	Fatty amides	Endogenous	C24H41NO3	9.69	391.3080	$[M+NH_4]^+$	409.3418	409.3421	0.68	392.3254
Palmitoyl ethanolamide	Carboximidic acids	Endogenous	C18H37NO2	12.86	299.2830	$[M+H]^+$	300.2903	300.2901	-0.59	282.2784, 252.2703
Palmitoylcarnitine	Carnitines	Endogenous	C23H45NO4	12.45	399.3350	$[M+H]^+$	400.3423	400.3428	1.31	144.1021
PGF2a acetate	Prostanoids	Endogenous	C22H36O6	10.94	396.2550	$[M+H]^-$	395.2477	395.2481	1.01	
Phenylacetaldehyde	Benzene and substituted derivatives	Endogenous	C8H8O	8.93	120.0580	$[M+H]^+$	121.0653	121.0651	-1.45	91.0549, 65.03852
Phenylalanine	Amino acids, peptides, and analogues	Endogenous	C9H11NO2	0.55	165.0790	$[M+H]^+$	166.0858	166.0861	1.81	120.0804, 93.0700
Physcion	Anthraquinone	Exogenous	C22H22O10	8.53	446.1210	$[M+H]^-$	445.1137	445.1129	-1.85	240.0425
Phytosphingosine	Sphingolipids and derivatives	Endogenous	C18H39NO3	9.83	317.2930	$[M+H]^+$	318.3003	318.3001	-0.55	300.2895, 282.2793
Pinane thromboxane A2	Eicosanoids and derivatives	Endogenous	C24H40O3	11.69	376.2980	$[M+H]^-$	375.2907	375.2894	-3.53	357.27
Propionyl carnitine	Carnitines	Endogenous	C10H19NO4	0.56	217.1320	$[M+H]^+$	218.1393	218.14	3.32	144.1025, 85.0281
Prunetin	Flavonoids	Exogenous	C16H12O5	8.03	284.0690	$[M+H]^+$	285.0762	285.0761	-0.35	253.0497
Quinoline	Alkaloids	Exogenous	C9H7N	7.52	129.0580	$[M+H]^+$	130.0653	130.065	-2.12	103.0512, 77.0377
Resolvin E1	Hydroxyeicosapentaenoic acids	Endogenous	C20H30O5	8.77	350.2093	$[M+H]^+$	351.2166	351.2178	3.42	245.1542
Ricinoleic acid	Hydroxy fatty acids	Endogenous	C18H34O3	11.40	298.2510	$[M+H]^-$	297.2437	297.2430	-2.44	279.2286, 183.1380

Continuation Supplementary Table 1

Saponin H	Prenol lipids	Exogenous	C36H58O10	9.77	650.3990	$[M+H]^+$	651.4063	651.4075	1.88	347.2576
Sarmentine	Pyrrolidines	Exogenous	C14H23NO	8.22	221.1780	$[M+H]^+$	222.1853	222.1853	0.11	178.1238, 72.081
Sphingosine 16:1	Sphingolipids and derivatives	Endogenous	C16H33NO2	10.03	271.2520	$[M+H]^+$	272.2593	272.2585	-2.85	282.2790, 252.2667
Sphingosine 18:1 isomer 1	Sphingolipids and derivatives	Endogenous	C18H37NO2	11.04	299.2830	$[M+H]^+$	300.2903	300.2901	-0.59	282.2782, 252.2701
Sphingosine 18:1 isomer 2	Sphingolipids and derivatives	Endogenous	C18H37NO2	11.76	299.2820	$[M+H]^+$	300.2893	300.2898	1.74	282.2788, 252.2751
S-Succinylglutathione	Enzyme	Endogenous	C14H21N3O9S	9.74	407.1020	$[M+NH_4]^+$	425.1358	425.1345	-3.11	346.1068, 76.0393
Stearamide	Fatty amides	Endogenous	C18H37NO	14.98	283.2880	$[M+H]^+$	284.2953	284.2947	-2.03	130.1219, 102.0911
Stearoyl carnitine	Carnitines	Endogenous	C25H49NO4	12.42	427.3660	$[M+H]^+$	428.3733	428.373	-0.64	309.2781, 144.1028
Stearoylethanolamide	N-acylethanolamines	Endogenous	C20H41NO2	14.17	327.3140	$[M+H]^+$	328.3213	328.3213	0.07	311.2897, 62.0601
Stercobilin	Bilirubins	Endogenous	C33H46N4O6	8.14	594.3420	$[M+H]^+$	595.3493	595.3474	-3.15	126.0915
Taurocholic acid	Bile acids and derivatives	Endogenous	C26H45NO7S	7.97	515.2930	$[M+H]^-$	514.2857	514.2877	3.84	514.2877
Taurodeoxycholic acid	Bile acids and derivatives	Endogenous	C26H45NO6S	8.16	499.2980	$[M+H]^+$	500.3053	500.3061	1.65	400.3275, 126.0234
Tetradecenoyl carnitine	Carnitines	Endogenous	C21H39NO4	12.34	369.2880	$[M+H]^+$	370.2953	370.2952	-0.21	165.1623, 144.1019
Triethylamine	Amine derivative	Exogenous	C6H15N	0.55	101.1200	$[M+H]^+$	102.1273	102.1276	3.17	74.0964, 58.0654, 46.0653
Trigonelline	Alkaloids	Exogenous	C7H7NO2	0.50	137.0470	$[M+NH_4]^+$	155.0808	155.0814	3.72	138.0542, 94.0660
Trihydroxyflavone	Flavonoids	Exogenous	C15H10O5	8.81	270.0530	$[M+H]^-$	269.0457	269.0451	-2.32	151.0538
Tyrosine	Amino acids, peptides, and analogues	Endogenous	C9H11NO3	0.56	181.0739	$[M+H]^+$	182.0812	182.0817	2.92	119.0501, 91.0547
Undecanoylglycine	Carboxylic acids and derivatives	Endogenous	C13H25NO3	8.69	243.1840	$[M+H]^+$	244.1913	244.1915	0.92	198.1822, 76.0389
Undecanoylglycine	Carboxylic acids and derivatives	Endogenous	C13H25NO3	8.85	243.1840	$[M+H]^+$	244.1913	244.1911	-0.72	198.1834, 76.0374
Urobilin	Bilirubins	Endogenous	C33H40N4O6	8.19	588.2940	$[M+H]^-$	587.2867	587.2885	3.02	450.2031, 124.0726
Ursodeoxycholic acid	Bile acids and derivatives	Endogenous	C24H38O5	9.29	392.2927	$[M+H]^+$	393.2999	393.2994	-1.36	357.2781
Vaccenyl carnitine	Carnitines	Endogenous	C25H47NO4	12.32	425.3500	$[M+H]^+$	426.3573	426.3576	0.76	307.263, 144.1022

Supplementary Table 2. List of metabolites tentatively identified in fecal samples by GC–TOF/MS. Information about family, chromatographic retention time, derivatization formula, monoisotopic molecular weight and formula and *m/z* values supporting the identification are included.

Compound	Protocol	Familiy	Exogenous/ Endogenous	RT (min)	Mono-isotopic molecu-lar weight	Formula (derivatization product)	Fragments (derivatization product)		
1H-Indole, 3-methyl	Without methoxymation	Indoles and derivatives	Endogenous	11.95	131.1745	C12H17NSi	202.1035 – [C12H16NSi]+	188.0873 – [C11H14NSi]+	130.06659 – [C9H8N]+
2-Aminobutyric acid	Both protocols	Carboxylic acids and derivatives	Endogenous	7.89	103.0633	C10H25NO2Si2	218.1012 – [C8H20NO2Si2]+	142.0681 – [C6H12NOSi]+	114.0715 – [C5H12NSi]+
2-Hydroxy-2-methylsuccinic acid	Both protocols	Hydroxy fatty acids	Endogenous	12.17	148.0372	C5H8O5	148.0372 – [C5H8O5]+	103.0381 – [C4H7O3]+	87.0076 – [C3H3O3]+
2-Hydroxyisocaproic acid	Without methoxymation	Hydroxy fatty acids	Endogenous	8.53	132.0786	C12H28O3Si2	159.1374 – [C9H19O2]+	103.0743 – [C5H11O2]+	
2-Ketoisocaproic acid	With methoxymation	Short-chain keto acids and derivatives	Endogenous	7.85	130.0630	C9H18O3Si	169.0699 – [C8H13O2Si]+	157.0689 – [C7H13O2Si]+	143.0882 – [C7H15OSi]+
2-Oxo-3-methylvaleric acid	Both protocols	Short-chain keto acids and derivatives	Endogenous	8.53	130.0630	C9H18O3Si	193.0301 – [C9H9O3Si]+	174.1042 – [C8H18O2Si]+	143.0878 – [C7H15OSi]+
3-Amino- 2-piperidone	Both protocols	Carboxylic acids and derivatives	Endogenous	12.08	114.0793	C11H26N2OSi2	258.1551 – [C11H26N2OSi2]	243.1327 – [C10H23N2OSi2]+	203.1135 – [C8H21NOSi2]+
3-Methyl-2-oxovaleric acid	With methoxymation	Short-chain keto acids and derivatives	Endogenous	7.15	130.0630	C9H18O3Si	193.0299 – [C9H9O3Si]+	174.1045 – [C8H18O2Si]+	143.0876 – [C7H15OSi]+
4-Hydroxybenzoic acid	Both protocols	Phenols and derivatives	Endogenous	14.08	138.0316	C13H22O3Si2	282.1081 – [C13H22O3Si2]+	267.0866 – [C12H19O3Si2]+	223.0962 – [C11H19OSi2]+
Acetoacetic acid	Both protocols	Short-chain keto acids and derivatives	Endogenous	8.53	102.0317	C10H22O3Si2	246.1095 – [C10H22O3Si2]+	231.0903 – [C9H19O3Si2]+	157.0696 – [C7H13O2Si]+
Aconitic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	15.40	174.0164	C18H40O7Si4	465.1590 – [C17H37O7Si4]+	375.1100 – [C14H27O6Si3]+	273.0971 – [C11H21O4Si2]+
Adipic acid	Without methoxymation	Dicarboxylic acids and derivatives	Endogenous	12.48	146.0579	C12H26O4Si2	185.0986 – [C9H17O2Si]+	111.1155 – [C8H15]+	83.0844 – [C6H11]+
Alanine	Both protocols	Amino acids, peptides, and analogues	Endogenous	6.84	89.0477	C9H23NO2Si2	218.1013 – [C8H20NO2Si2]+	190.1067 – [C7H20NOSi2]+	116.0885 – [C5H14NSi]+
Allofuranose	Both protocols	Carbohydrates and derivatives	Endogenous	17.70	180.0634	C21H52O6Si5	345.1369 - [C14H29O4Si3]+	217.1080 - [C9H21O2Si2]+	191.0923 - [C7H19O2Si2]+
Allopyranose	Both protocols	Carbohydrates and derivatives	Endogenous	17.05	180.0634	C21H52O6Si5	319.1552 – [C17H27O2Si2]+	217.1068 – [C9H21O2Si2]+	191.0913 – [C7H19O2Si2]+
Aminoadipic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	13.15	161.0688	C15H35NO4Si3	260.1497 – [C11H26NO2Si2]+	217.1318 – [C9H23NOSi2]+	73.0477 – [C3H7NO]+
Aminomalonic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	12.01	119.0219	C12H29NO4Si3	320.1163 – [C11H26NO4Si3]	292.1216 – [C10H26NO3Si3]+	218.1028 – [C8H20NO2Si2]+
Ascorbic acid	Both protocols	Vitamins and derivatives	Exogenous	17.59	176.0321	C18H40O6Si4	464.1910 – [C18H40O6Si4]+	332.1301 - [C13H28O4Si3]+	117.0362 - [C4H9OSi]+
Asparagine	Both protocols	Amino acids, peptides, and analogues	Endogenous	14.51	132.0535	C16H40N2O3Si4	420.2114 – [C16H40N2O3Si4]+	303.1374 – [C11H27N2O2Si3]+	290.1289 – [C10H26N2O2Si3]+

Continuation Supplementary Table 2

Aspartic acid	Both protocols	Amino acids, peptides, and analogues	Endogenous	12.32	133.0375	C10H23NO4Si2	262.0923 – [C9H20NO4Si2]+	172.0461 – [C6H10NO3Si2]+	160.0785 – [C6H14NO2Si]+
Azelaic acid	Both protocols	Dicarboxylic acids and derivatives	Endogenous	15.83	188.1049	C15H32O4Si2	317.1577 – [C14H29O4Si2]+	217.1070 – [C9H21O2Si2]+	204.1012 – [C8H20O2Si2]+
Benzoic acid	Both protocols	Phenols and derivatives	Exogenous	9.06	122.0368	C10H14O2Si	179.0519 – [C9H11O2Si]+	122.0353 – [C7H6O2]+	105.0330 – [C7H5O]+
Butanediol	Without methoxymation	Alcohols and polyols	Endogenous	7.64	90.0681	C4H10O2	90.0667 – [C4H10O2]+	71.0492 – [C4H7O]+	
Butanoic acid,-3-methyl	Both protocols	Short-chain keto acids and derivatives	Endogenous	7.15	102.0680	C5H10O2	87.0451 – [C4H7O2]+	60.0571 – [C3H8O]+	
C10:0, Decanoic acid	Both protocols	Fatty acids and conjugates	Endogenous	11.98	172.1463	C13H28O2Si	229.1606 – [C12H25O2Si]+	185.0990 – [C9H17O2Si]+	131.0519 – [C5H11O2Si]+
C12:0, Dodecanoic acid	Both protocols	Fatty acids and conjugates	Endogenous	14.40	200.1776	C15H32O2Si	272.2159 – [C15H32O2Si]+	257.1921 – [C14H29O2Si]+	145.0665 – [C6H13O2Si]+
C16:0, Palmitic acid	Both protocols	Fatty acids and conjugates	Endogenous	18.36	256.2402	C19H40O2Si	328.2762 – [C19H39O2Si]+	313.2541 – [C18H36O2Si]+	285.2229 – [C16H32O2Si]+
C16:1, Palmitoleic acid	Both protocols	Fatty acids and conjugates	Endogenous	18.36	254.2246	C19H38O2Si	313.2559 – [C18H37O2Si]+	129.0363 – [C9H5O]+	117.0365 – [C8H5O]+
C17:0, Heptadecanoic acid	Both protocols	Fatty acids and conjugates	Endogenous	19.27	270.2559	C20H42O2Si	342.2974 – [C20H42O2Si]+	299.2385 - [C17H35O2Si]+	117.0378 - [C4H9O2Si]+
C18:0, Stearic acid	Both protocols	Fatty acids and conjugates	Endogenous	20.15	284.2710	C21H44O2Si	356.3080 – [C21H44O2Si]+	341.2854 – [C20H41O2Si]+	313.2534 – [C18H37O2Si]+
C18:1, Oleic acid	Both protocols	Fatty acids and conjugates	Endogenous	19.90	282.2559	C21H42O2Si	354.2942 – [C21H42O2Si]+	339.2713 – [C20H39O2Si]+	185.0990 – [C9H17O2Si]+
C18:2, Linoleic acid	Both protocols	Fatty acids and conjugates	Endogenous	20.00	280.2400	C21H40O2Si	353.2832 – [C21H40O2Si]+	337.2557 – [C20H37O2Si]+	309.2208 – [C18H33O2Si]+
C22:0, Behenic acid	Both protocols	Fatty acids and conjugates	Endogenous	23.36	340.3341	C25H52O2Si	397.3492 – [C24H49O2Si]+	145.0652 - [C10H9O]+	129.0367 - [C9H5O]+
C4:0, Butanoic acid	Both protocols	Short chain fatty acids and derivatives	Endogenous	7.50	88.0530	C7H16O2Si	145.0679 – [C6H13O2Si]+	117.0374 – [C4H10O2Si]+	73.0653 – [C4H9O]+
C6:0, Caproic acid	Both protocols	Short chain fatty acids and derivatives	Endogenous	6.42	116.0837	C9H20O2Si	173.0991 – [C8H17O2Si]+	159.0832 – [C7H15O2Si]+	131.0521 – [C5H11O2Si]+
C8:0, Caprylic acid	Both protocols	Fatty acids and conjugates	Endogenous	9.33	144.1150	C11H24O2Si	216.1495 – [C11H24O2Si]+	201.1293 – [C10H21O2Si]+	129.0384 – [C5H9O2Si]+
C9:0, Nonanoic acid	Both protocols	Fatty acids and conjugates	Endogenous	10.61	158.1307	C12H26O2Si	215.1465 – [C11H23O2Si]+	145.0670 – [C6H13O2Si]+	117.0370 – [C4H9O2Si]+
Citric acid	Both protocols	Carboxylic acids and derivatives	Endogenous	16.22	192.0270	C18H40O7Si4	466.1321 – [C17H37O7Si4]+	347.1144 – [C13H27O5Si3]+	273.0951 – [C11H21O4Si2]+
Coprostan-3-ol	Without methoxymation	Cholesterols and derivatives	Exogenous	26.54	388.3705	C30H56OSi	403.0515 – [C29H11OSi]+	370.0810 – [C26H14OSi]+	215.0901 – [C13H15OSi]+
Cyclopentylamine	Without methoxymation	Amine derivative	Endogenous	10.92	85.0891	C8H19NSi	142.1048 – [C7H16NSi]+	128.0891 – [C6H14NSi]+	1020732 – [C4H12NSi]+
Cysteine	Both protocols	Amino acids, peptides, and analogues	Endogenous	15.21	121.0197	C12H31NO2SSi3	340.1679 – [C12H31NO2SSi3]+	220.0999 – [C8H20NO2Si2]+	100.0555 – [C4H10NSi]+

Continuation Supplementary Table 2

Dihydroxybutanoic acid	Both protocols	Hydroxy fatty acids	Endogenous	10.54	120.0423	C13H32O4Si3	321.1365 – [C12H29O4Si3]+	203.0927 – [C8H19O2Si2]+	117.0739 – [C5H13OSi]+
Erythritol	Both protocols	Sugar alcohols	Endogenous	12.62	122.0579	C16H4204Si4	366.1292 – [C16H30O2Si4]+	321.1533 – [C12H33O2Si4]+	277.1476 – [C11H29O2Si3]+
Erythronic acid	Both protocols	Lactones	Endogenous	10.88	118.0266	C16H42O4Si4	379.1639 – [C14H35O4Si4]+	293.1419 – [C11H29O3Si3]+	217.1084 – [C9H21O2Si2]+
Erythrulose	Both protocols	Carbohydrates and derivatives	Endogenous	14.57	120.0423	C16H41NO4Si4	320.1512 – [C16H26NO2Si2]	117.0641 – [C4H11NOSi]+	103.0472 – [C3H9NOSi]+
Fructose	Both protocols	Carbohydrates and derivatives	Endogenous	16.76	180.0634	C22H55NO6Si5	335.1492 – [C17H27O3Si2]+	217.1070 – [C10H17O5]+	191.0913 – [C8H15O5]+
Fumaric acid	Both protocols	Carboxylic acids and derivatives	Endogenous	10.45	116.0110	C10H20O4Si2	245.1014 – [C10H21O3Si2]+	155.0521 – [C7H11O2Si]+	143.0516 – [C6H11O2Si]+
Furoic acid	Without methoxymation	Furans and derivatives	Endogenous	7.37	112.0160	C8H12O3Si	169.0315 – [C7H9O3Si]+	125.0233 – [C6H5O3]+	95.0122 – [C5H3O2]+
Galactose	Both protocols	Carbohydrates and derivatives	Endogenous	16.84	180.0634	C21H52O6Si5	305.1386 – [C16H25O2Si2]+	265.1099 – [C10H21O6Si]+	204.0997 – [C8H20O2Si2]+
Glucofuranoside	Both protocols	Carbohydrates and derivatives	Endogenous	16.92	180.0634	C21H52O6Si5	305.1392 – [C16H25O2Si2]+	204.1001 - [C8H20O2Si2]+	191.0917 - [C7H19O2Si2]+
Gluconic acid	Both protocols	Sugars acids and derivatives	Endogenous	18.01	196.0583	C24H60O7Si6	434.1427 – [C16H34O6Si4]+	393.1586 – [C15H33O6Si3]+	
Glucopyranose	Both protocols	Carbohydrates and derivatives	Endogenous	17.59	180.0634	C21H52O6Si5	305.1395 – [C16H25O2Si2]+	204.0998 - [C8H20O2Si2]+	191.0913 - [C7H19O2Si2]+
Glucuronic acid	Without methoxymation	Carboxylic acids and derivatives	Endogenous	14.85	194.0427	C21H50O7Si5	449.2025 – [C18H41O5Si4]+	305.1408 – [C12H29O3Si3]+	204.1002 – [C8H20O2Si2]+
Glutamic acid	Both protocols	Amino acids, peptides, and analogues	Endogenous	13.82	147.0540	C14H33NO4Si3	348.1464 – [C13H30NO4Si3]	246.1338 – [C10H24NO2Si2]+	230.1019 – [C9H20NO2Si2]+
Glutamine	Both protocols	Amino acids, peptides, and analogues	Endogenous	15.55	146.0691	C14H34N2O3Si3	301.0706 – [C14H19NOSi3]+	156.0841 – [C7H14NOSi]+	130.0666 – [C5H12NOSi]+
Glyceric acid	Both protocols	Sugar acids and derivatives	Endogenous	10.27	106.0266	C12H30O4Si3	307.1189 - [C11H27O4Si3]+	292.13351 - [C11H28O3Si3]+	189.0767 - [C7H17O2Si2]+
Glycerol	Both protocols	Sugar alcohols	Endogenous	9.64	92.0473	C12H32O3Si3	293.1401 – [C11H29O3Si3]+	218.1127 – [C9H22O2Si2]+	205.1071 – [C8H21O2Si2]+
Glycine	Without methoxymation	Amino acids, peptides, and analogues	Endogenous	8.20	75.0320	C11H29NO2Si3	276.1248 – [C10H26NO2Si3]	248.1303 – [C9H26NOSi3]+	174.1125 – [C7H20NSi2]+
Glycolic acid	Both protocols	Hydroxy acids and derivatives	Endogenous	6.43	76.0160	C8H20O3Si2	205.0698 – [C7H17O3Si2]+	177.0744 – [C4H8O3Si2]+	161.0308 – [C6H17OSi]+
Heptanedioic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	13.69	160.0736	C13H28O4Si2	289.1287 – [C12H25O4Si2]+	217.0890 – [C9H17O4Si]+	129.0915 – [C7H13O2]+
Histidine	Without methoxymation	Amino acids, peptides, and analogues	Endogenous	17.07	155.0695	C15H33N3O2Si3	254.1493 – [C11H24N3Si2]+	218.1024 – [C11H14N2Si2]+	154.0680 – [C11H8N]+
Homoserine	Without methoxymation	Amino acids, peptides, and analogues	Endogenous	11.99	119.0582	C13H33NO3Si3	320.1525 – [C12H30NO3Si3]	218.1095 – [C9H21O2Si2]+	130.0666 – [C5H14Si2]+
Hydrocaffeic acid	Without methoxymation	Phenylpropanoic acids	Exogenous	17.33	182.0579	C18H34O4Si3	398.1762 – [C18H34O4Si3]+	267.0148 – [C14H7O4Si]+	179.0884 – [C10H15OSi]+

Continuation Supplementary Table 2

Hydrocinnamic acid	Without methoxymation	Phenylpropanoic acids	Endogenous	11.40	164.0473	C12H18O2Si	222.1029 – [C12H18O2Si]+	207.0325 – [C11H15O2Si]+	104.0632 – [C8H8]+
Hydroquinone	Without methoxymation	Phenols and derivatives	Exogenous	11.15	110.0368	C12H22O2Si2	254.0143 – [C12H22O2Si2]+	239.2368 - [C11H20OSi]+	137.0406 - [C7H9OSi]+
Hydroxybutyric acid	Both protocols	Hydroxy fatty acids	Endogenous	7.42	104.0473	C10H24O3Si2	233.1025 – [C9H21O3Si2]+	205.1067 – [C8H21O2Si2]+	191.0922 – [C7H19O2Si2]+
Hypoxanthine	Without methoxymation	Purines and purine derivatives	Endogenous	15.92	136.0380	C11H20N4OSi2	280.1164 – [C11H20N4OSi2]+	265.0935 – [C10H17N4OSi2]+	193.0465 – [C9H13OSi2]+
Indole-3-acetic acid	Without methoxymation	Indoles and derivatives	Endogenous	17.53	175.0630	C16H25NO2Si2	319.1422 – [C16H25NO2Si2]+	202.1045 – [C12H16NSi]+	130.0666 – [C9H8N]+
Inositol	Both protocols	Carbohydrates and derivatives	Endogenous	18.42	180.0634	C24H60O6Si6	507.2271 – [C20H47O5Si5]+	305.1408 – [C12H29O3Si3]+	217.1080 – [C9H21O2Si2]+
Isoleucine	Both protocols	Amino acids, peptides, and analogues	Endogenous	9.64	131.0946	C12H29NO2Si2	260.1482 – [C11H26NO2Si2]	232.1533 – [C10H26NOSi2]+	158.1355 – [C8H20NSi]+
Itaconic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	10.40	130.0266	C11H22O4Si2	259.1159 – [C11H23O3Si2]+	215.0915 – [C9H19O2Si2]+	131.0346 – [C4H11OSi2]+
Lactic acid	Both protocols	Hydroxy acids and derivatives	Endogenous	6.16	90.0320	C9H22O3Si2	233.1064 – [C9H21O3Si2]+	219.0923 – [C8H19O3Si2]+	203.0604 – [C7H15O3Si2]+
Leucine	Without methoxymation	Amino acids, peptides, and analogues	Endogenous	9.33	131.0946	C12H29NO2Si2	260.1489 – [C11H26NO2Si2]+	232.1535 – [C10H26NOSi2]+	158.1357 – [C8H20NSi]+
Lysine	Both protocols	Amino acids, peptides, and analogues	Endogenous	15.21	146.1055	C15H38N2O2Si3	362.2227 – [C15H38N2O2Si3]+	230.1162 – [C10H22NOSi2]+	156.1187 – [C8H18NSi]+
Malic acid	Both protocols	Hydroxy acids and derivatives	Endogenous	12.33	134.0215	C13H30O5Si3	245.0662 – [C9H17O4Si2]+	233.1020 - [C9H21O3Si2]+	217.0715 - [C8H17O3Si2]+
Malonic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	11.58	104.0110	C9H20O4Si2	233.1001 – [C9H21O3Si2]+	147.0658 – [C5H15OSi2]+	73.0464 – [C3H9Si]+
Maltose	Both protocols	Carbohydrates and derivatives	Endogenous	24.15	342.1162	C36H86O11Si8	361.1670 - [C16H29O7Si]+	204.1009 - [C9H16O5]+	191.0923- [C8H15O5]+
Mannitol	Both protocols	Carbohydrates and derivatives	Endogenous	17.59	182.0790	C24H62O6Si6	520.1592 - [C23H36O6Si4]+	319.1569 – [C13H31O3Si3]+	205.1043 – [C12H17OSi]+
Mannofuranose	Both protocols	Carbohydrates and derivatives	Endogenous	16.76	180.0634	C21H52O6Si5	540.2631- [C21H52O6Si5]+	437.2020 - [C17H41O5Si4]+	265.0939 - [C14H26O6Si]+
Mannopyranose	Both protocols	Carbohydrates and derivatives	Endogenous	16.92	180.0634	C21H52O6Si5	435.1877– [C21H35O4Si3]+	204.1003 - [C8H20O2Si2]+	147.0656 - [C5H15OSi2]+
Methionine	Both protocols	Amino acids, peptides, and analogues	Endogenous	12.62	149.0510	C11H27NO2SSi2	293.1272 – [C11H27NO2SSi2]+	176.0921 – [C6H18NOSi2]+	128.0524 – [C5H10NOSi]+
Monopalmitin	Both protocols	Glycerolipids	Endogenous	23.36	330.2770	C25H54O4Si2	459.3341 – [C24H51O4Si2]+	371.2978 – [C21H43O3Si]+	313.2553 – [C18H37O2Si]+
Monostearin	Both protocols	Glycerolipids	Endogenous	24.36	358.3083	C27H58O4Si2	487.3626 – [C26H55O4Si2]+	399.3283 – [C23H47O3Si]+	341.2852 – [C20H41O2Si]+
M-Pyrol	Both protocols	Pyrrolidines	Exogenous	5.96	99.0684	C5H9NO	98.0593 – [C5H8NO]+	70.0772 – [C5H10]+	42.0442 – [C3H6]+
Myoinositol	Both protocols	Carbohydrates and derivatives	Endogenous	18.87	180.0634	C24H60O6Si6	612.2926 – [C24H60O6Si6]+	305.1403 – [C12H29O3Si3]+	217.1059 – [C9H21O2Si2]+

Continuation Supplementary Table 2

N-Acetyltyrosine ethyl ester	Both protocols	Amino acids, peptides, and analogues	Endogenous	20.36	251.1152	C13H17NO4	251.1149 – [C13H17NO4]+	192.0659 – [C10H10NO3]+	192.0659 – [C10H10NO3]+
Oleic acid, methyl ester	Both protocols	Fatty acids and conjugates	Endogenous	18.99	296.4950	C19H36O2	296.2712 – [C19H36O2]+	264.2442 – [C18H32O]+	224.1564 – [C17H20]+
Ornithine	Both protocols	Amino acids, peptides, and analogues	Endogenous	13.76	132.0899	C14H36N2O2Si3	348.2071 – [C14H36N2O2Si3]+	142.1030 – [C7H16NSi]+	115.0788 – [C5H13NSi]+
Oxalic acid isomer 1	Both protocols	Dicarboxylic acids and derivatives	Endogenous	6.16	89.9953	C8H18O4Si2	219.0482 – [C7H15O4Si2]+	190.0830 – [C7H18O2Si2]+	131.0344 – [C4H11OSi2]+
Oxalic acid isomer 2	Both protocols	Dicarboxylic acids and derivatives	Endogenous	7.19	89.9953	C8H18O4Si2	219.0485 – [C7H15O4Si2]+	190.0832 – [C7H18O2Si2]+	131.0346 – [C4H11OSi2]+
Palmitic acid, methyl ester	Both protocols	Fatty acids and conjugates	Endogenous	17.59	270.2559	C17H34O2	270.2557 – [C17H34O2]+	239.2376 – [C16H31O]+	
Phenylacetic acid	Both protocols	Benzene and substituted derivatives	Endogenous	11.50	136.0524	C16H26O2Si	278.1652 – [C16H26O2Si]+	263.1483 – [C15H23O2Si]+	204.0982 – [C12H16OSi]+
Phenylalanine	Both protocols	Amino acids, peptides, and analogues	Endogenous	14.08	165.0790	C15H27NO2Si2	294.1339 – [C14H24NO2Si2]+	218.1021 – [C8H20NO2Si2]+	192.1194 – [C11H18NSi]+
Phosphoric acid	Both protocols	Inorganic compounds	Endogenous	9.31	97.9769	C9H27O4PSi3	314.0950 – [C9H27O4PSi3]+	299.0722 – [C8H24O4PSi3]+	283.0427 – [C7H20O4PSi3]+
Pipecolinic acid	Without methoxymation	Carboxylic acids and derivatives	Endogenous	12.66	129.0790	C12H27NO2Si2	273.1577 – [C12H27NO2Si2]	229.0948 – [C9H19NO2Si2]+	156.0475 – [C6H10NO2Si]+
Proline	Both protocols	Amino acids, peptides, and analogues	Endogenous	9.72	115.0633	C11H25NO2Si2	216.1224 – [C9H22NOSi2]+	170.0620 – [C7H12NO2Si]+	142.1043 – [C7H16NSi]+
Pyroglutamic acid isomer 1	Both protocols	Amino acids, peptides, and analogues	Endogenous	12.50	129.0426	C8H15NO3Si	201.0813 – [C8H15NO3Si]+	186.0553 – [C7H12NO3Si]+	103.0213 – [C3H7O2Si]+
Pyroglutamic acid isomer 2	Both protocols	Amino acids, peptides, and analogues	Endogenous	12.76	129.0426	C8H15NO3Si	201.0812 – [C8H15NO3Si]+	186.0554 – [C7H12NO3Si]+	103.0211 – [C3H7O2Si]+
Pyrophosphate	Both protocols	Inorganic compounds	Endogenous	14.25	173.9119	C12H36O7P2Si4	466.1002 – [C12H36O7P2Si4]+	451.0778 – [C11H33O7P2Si4]+	363.0428 – [C8H25O6P2Si3]+
Pyuvic acid	With methoxymation	Keto acids and derivatives/Carboxylic acids	Endogenous	6.01	88.0160	C7H15NO3Si	189.0812 – [C7H15NO3Si]+	174.0581 – [C6H12NO3Si]+	158.0621 – [C6H12NO2Si]+
Ribofuranose	Without methoxymation	Carbohydrates and derivatives	Endogenous	14.00	150.0528	C17H42O5Si4	305.1395 – [C16H25O2Si2]+	217.1068 – [C9H21O2Si2]+	147.0658 – [C5H15OSi2]+
Ribopyranose	Without methoxymation	Carbohydrates and derivatives	Endogenous	14.17	150.0528	C17H42O5Si4	305.1397 – [C16H25O2Si2]+	217.1066 – [C9H21O2Si2]+	147.0657 – [C5H15OSi2]+
Sarcosine	Both protocols	Amino acids, peptides, and analogues	Endogenous	11.19	89.0477	C9H23NO2Si2	218.1024 – [C8H20NO2Si2]+	116.0891 – [C5H14NSi]+	100.0571 – [C4H10NSi]+
Serine	Both protocols	Amino acids, peptides, and analogues	Endogenous	9.75	105.0426	C12H31NO3Si3	218.1033 – [C8H20NO2Si2]+	204.1283 – [C8H22NOSi2]+	188.0924 – [C7H18NOSi2]+
Sorbitol	Both protocols	Sugar alcohols	Endogenous	16.39	182.0790	C18H44O5Si4	453.2335 – [C18H44O5Si4]+	349.1665 – [C18H29O3Si2]+	217.1084 – [C9H21O2Si2]+
Squalene	Both protocols	Triterpene	Exogenous	24.79	410.3913	C30H50	367.3350 – [C27H43]+	81.0691 – [C6H9]+	69.0691 – [C5H9]+
Stearic acid, methyl ester	Both protocols	Fatty acids and conjugates	Endogenous	19.45	298.2872	C21H44O2Si	341.2855 - [C20H41O2Si]+	129.0367 - [C5H9O2Si]+	117.0733 - [C5H13OSi]+

Continuation Supplementary Table 2

Succinic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	9.95	118.0266	C10H22O4Si2	262.1103 - [C10H22O4Si2]+	172.0564 - [C7H12O3Si]+	147.0643 - [C9H11Si]+
Sulfate	Both protocols	Inorganic compounds	Endogenous	7.89	95.9517	C6H18O4SSi2	227.0231 – [C5H15O4SSi2]+	211.0010 – [C4H11O4SSi2]+	138.9516 – [CH3O4SSi]+
Talofuranose	Without methoxymation	Carbohydrates and derivatives	Endogenous	15.97	180.0634	C21H52O6Si5	435.1874 – [C17H39O5Si4]+	217.1080 – [C9H21O2Si2]+	191.0923 – [C7H19O2Si2]+
Talopyranose	Both protocols	Carbohydrates and derivatives	Endogenous	18.26	180.0634	C21H52O6Si5	305.1395 – [C16H25O2Si2]+	217.1068 – [C9H21O2Si2]+	191.0913 – [C7H19O2Si2]+
Threonic acid	Both protocols	Sugars acids and derivatives	Endogenous	13.23	136.0372	C16H40O5Si4	410.1814 – [C15H38O5Si4]+	378.1146 – [C13H30O5Si4]+	319.1201 – [C12H27O4Si3]+
Threonine	Both protocols	Amino acids, peptides, and analogues	Endogenous	11.11	119.0582	C12H31NO3Si3	264.1464 – [C10H25NO3Si2]+	219.1090 - [C8H21NO2Si2]+	130.0671 - [C5H12NOSi]+
Tricarballylic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	15.10	176.0321	C15H32O6Si3	377.1262 – [C14H29O6Si3]+	289.0551 – [C10H17O6Si2]+	259.0457 – [C9H15O5Si2]+
Tryptophan	Both protocols	Amino acids, peptides, and analogues	Endogenous	19.87	204.0899	C17H28N2O2Si2	231.1274 - [C10H23NO2Si2]+	219.1115 - [C12H15O2Si2]+	201.0735 - [C12H12OSi]+
Tyrosine	Both protocols	Amino acids, peptides, and analogues	Endogenous	17.49	181.0739	C18H35NO3Si3	382.1682 – [C17H32NO3Si3]	354.1731 – [C16H32NO2Si3]+	280.1538 – [C14H26NOSi2]+
Tyrosine-O	Both protocols	Amino acids, peptides, and analogues	Endogenous	16.86	181.0739	C18H35NO3Si3	382.1687 – [C17H32NO3Si3]+	354.1729 – [C16H32NO2Si3]+	280.1542 – [C14H26NOSi2]+
Uracil	Without methoxymation	Pyrimidines and pyrimidine derivatives	Endogenous	10.27	112.0273	C10H20N2O2Si2	255.0964 – [C10H19N2O2Si2]+	241.0807 – [C9H17N2O2Si2]+	147.0658 – [C5H15O2Si2]+
Urea	Both protocols	Organic carbonic acids and derivatives	Endogenous	9.12	60.0324	C7H20N2OSi2	204.1110 – [C7H20N2OSi2]+	189.0885 – [C6H17N2OSi2]+	171.0781 – [C6H15N2Si2]+
Uric acid	Both protocols	Purines and purine derivatives /Xanthines	Endogenous	18.79	168.0283	C17H36N4O3Si4	456.1847 – [C17H36N4O3Si4]+	441.1623– [C16H33N4O3Si4]	369.1238 – [C13H25N4O3Si3]
Uridine	Both protocols	Pyrimidine nucleosides	Endogenous	22.01	244.0695	C18H36N2O6Si3	445.1653 – [C17H33N2O6Si3]+	341.0185 – [C18H9N2O2Si2]+	243.0850 - [C14H15O2Si]+
Urocanic acid	Both protocols	Carboxylic acids and derivatives	Endogenous	18.01	138.0429	C12H22N2O2Si2	282.1218 – [C12H22N2O2Si2]+	267.0965 – [C11H19N2O2Si2]	180.0713 – [C8H12N2OSi]+
Valine	Both protocols	Amino acids, peptides, and analogues	Endogenous	8.50	117.0790	C11H27NO2Si2	246.1324 – [C10H24NO2Si2]+	218.1110 – [C9H22O2Si2]+	203.0809 – [C7H17NO2Si2]+
Vanillic acid	Without methoxymation	Phenols and derivatives	Exogenous	15.48	168.0423	C14H24O4Si2	312.1210 – [C14H24O4Si2]+	297.1019 – [C13H21O4Si2]+	126.0443 – [C6H10OSi]+
Xylose	Without methoxymation	Carbohydrates and derivatives	Endogenous	16.91	150.0528	C17H42O5Si4	305.1395– [C16H25O2Si2]+	217.1068 – [C9H21O2Si2]+	191.0913 – [C7H19O2Si2]+
Xylulose	Without methoxymation	Carbohydrates and derivatives	Endogenous	14.51	150.0528	C17H42O5Si4	306.1433 – [C16H26O2Si2]+	217.1080 – [C9H21O2Si2]+	129.0367 – [C5H9O2Si]+

CHAPTER VI

MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls

Talanta

174 (2017) 29–37

MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls

Mónica Calderón-Santiago[a,b,c], María Asunción López-Bascón[a,b,c], Ángela Peralbo-Molina[a,b,c], Feliciano Priego-Capote[a,b,c,*]

[a]*Maimónides Institute of Biomedical Research Maimónides (IMIBIC), Reina Sofía Hospital, University of Córdoba Córdoba, Spain*

[b]*Department of Analytical Chemistry, University of Córdoba, Annex C-3, Campus of Rabanales, Córdoba, Spain*

[c]*CIBER Fragilidad y Envejecimiento Saludable, Instituto de Salud Carlos III, Spain*

MetaboQC: A tool for correcting untargeted metabolomics data with mass spectrometry detection using quality controls

Mónica Calderón-Santiago, María Asunción López-Bascón, Ángela Peralbo-Molina, Feliciano Priego-Capote

ABSTRACT

Nowadays most metabolomic studies involve the analysis of large sets of samples to find a representative metabolite pattern associated to the factor under study. During a sequence of analyses the instrument signals can be subjected to the influence of experimental variability sources. Implementation of quality control (QC) samples to check the contribution of experimental variability is the most common approach in metabolomics. This practice is based on the filtration of molecular entities experiencing a variation coefficient higher than that measured in the QC data set. Although other robust correction algorithms have been proposed, none of them has provided an easy-to-use and easy-to-install tool capable of correcting experimental variability sources. In this research an R-package –the MetaboQC– has been developed to correct intra-day and inter-days variability using QCs analyzed within a pre-set sequence of experiments. MetaboQC has been tested in two data sets to assess the correction effects by comparing the metabolites variability before and after application of the proposed tool. As a result, the number of entities in QCs significantly different between days was reduced from 86 to 19% in the negative ionization mode and from 100 to 13% in the positive ionization mode. Furthermore, principal component analysis allowed detecting the filtration of instrumental variability associated to the injection order.

Keywords: Quality control; R package; batch effect; data pretreatment; instrumental variability

1. Introduction

Metabolomics is the 'omics' concerned with the identification and/or quantitation of small molecules or metabolites (generally < 1500 Da) present in biofluids, cells or organisms. Metabolomic studies can be performed by following three different strategies: fingerprinting analysis, targeted or untargeted analysis of metabolites [1]. Fingerprinting analysis is aimed at obtaining direct, fast snapshots of the metabolic state of the sample; while targeted analysis is focused on the determination of a known family of metabolites, and untargeted analysis deals with detection and/or identification of the greatest possible number of metabolites present in the sample, usually with the purpose of a (semi)quantitative comparison of the metabolic state of different groups of individuals. Separation techniques coupled to mass spectrometry detection (LC–MS, GC–MS and CE–MS) affords detection of a large number of metabolites in biological samples [2–6]. For this reason, the combination of chromatographic or electrophoretic techniques to mass spectrometry leads to three of the most employed analytical platforms in targeted and untargeted metabolomics analysis.

The current trend in targeted/untargeted metabolomics workflows based on separation techniques coupled to MS detection is the analysis of large sets of samples to obtain representative information from the biological system under study. Thus, most studies require long periods to analyze samples in which the quantitative response can fluctuate by alteration of the instrument performance owing to accumulation of matrix components in different instrumental zones or simply by periodic practices such as instrument calibration or cleaning protocols. These sources of instrumental variability are generally corrected in targeted analysis by using isotopically labeled internal standards [7,8]. However, this is not viable in untargeted analysis owing to the wide chemical heterogeneity of metabolites that would force to spike the sample with an isotopically labeled standard per detected metabolite.

The most common practice to monitor experimental variability (including instrumental variability) in untargeted analysis is the implementation of quality control samples (QCs), periodically inserted in the sequence of analyses programmed for each batch of samples [9–19]. Initially, most studies including QCs have been focused on

identifying metabolites reporting signals that experience fluctuations during the sequence. For this purpose, a cut-off value in terms of variability is set to filter molecular entities or metabolites in the final data set. This is the case of the metabolomics guidelines proposed by Dunn *et al.* [10] that were also lately adopted by Vinaixa *et al.* [11], or the protocol developed with liver samples proposed by Masson *et al.* [17]. Therefore, correction of quantitative or semi-quantitative signals for detected metabolites is not generally carried out before statistical analysis.

More recently some researchers have proposed correction approaches based on QCs looking for the best option to minimize experimental variability. These correction tools are based on different mathematical strategies as that proposed by Karpievitch *et al.* [9], who used singular value decomposition to find systematic trends attributable to bias; or the use of a Gaussian process to reduce bias due to sample collection, sample treatment and instrumental variability through a normalization process [13]. Regression models has also constituted a common practice, and different regression methodologies such as support vector regression have been tested [18,19]. Recent studies have used local regression to correct a given general trend in batches of samples [10,12], but the procedure requires high computational capabilities and is quite sensitive to outliers. Other tools classify the metabolites according to their behavior along the sequence of analyses and use correction functions for each group to correct instrumental variability [9]. However, it is not the injection order the only factor contributing to instrumental variability, which can also be caused by periodic practices as those previously cited (*viz.*, instrument calibration or clean-up protocols). With these premises, the desirable tool should allow correcting the signal of each metabolite by considering its own variability.

The use of QCs to remove instrumental variability acts as a scaling tool, since the (semi)quantitative response of each compound is related to its response in a pool of samples. In this research the statistical package MetaboQC has been developed to study and filtrate experimental variability in data sets generated by MS analysis of sequences developed for several days. This new tool uses QCs to individually correct any tendency on quantitative signals of metabolites that can be associated to experimental variability.

MetaboQC considers the order of injections in batches, as well as the variability detected among different days. For this purpose, the package analyzes the trends of quantitative signals for each detected metabolite in the QCs to define a correction function that is subsequently applied to the batch of samples. It is worth mentioning that QCs should be preferentially prepared with a pool of samples of the cohort to be studied. In this way, representativeness of matrix effects and metabolome composition is ensured. The open source data analysis software R has been used to create the algorithms included in MetaboQC. The applicability of the proposed package has been demonstrated by testing two data sets obtained by analysis of serum samples from a large cohort.

2. Materials and methods

2.1. Chemicals

LC–MS grade acetonitrile from Fisher Scientific (Madrid, Spain) and formic acid from Scharlab (Barcelona, Spain) were used to prepare chromatographic mobile phases. Deionized water (18 MΩ·cm) supplied by a Milli-Q water purification system from Millipore (Bedford, MA) was used for sample treatment and also to prepare the aqueous mobile phase.

2.2. Instruments and apparatus

A Sorvall Legend Micro 21R centrifuge from Thermo Scientific (Waltham, MS) was used to centrifuge samples after deproteination. An Agilent 1200 Series LC system coupled to an Agilent 6540 UHD Accurate-Mass QTOF hybrid mass spectrometer equipped with a dual electrospray (ESI) source (Santa Clara, CA) was used for sample analysis. Chromatographic eluates were monitored by tandem mass spectrometry in high resolution mode.

2.3. Sample collection and pretreatment

Venous blood was collected from 240 healthy donors in evacuated sterile serum tubes (Vacutainer, Becton Dickinson, Franklin Lakes, NJ, USA) containing no additives.

The tubes were incubated at room temperature for 10 min to facilitate coagulation, then centrifuged at 2000 × *g* for 15 min at 4 °C to isolate the serum fraction (processing within 2 h after collection). A serum pool was prepared to be used as QC by taking 100 μL of each sample and mixing them. Both the serum pool (QC samples) and the samples were fractioned in 100 μL aliquots that were stored at −80 °C until analysis.

All steps from serum sampling to analysis were performed in compliance with the guidelines dictated by the World Medical Association Declaration of Helsinki of 2004 and supervised by specialist staff of the Maimonides Biomedical Research Institute (Cordoba, Spain). The study was approved by the ethics committee of the Reina Sofia University Hospital. The individuals selected for this study were previously informed to obtain consent.

2.4. Sample preparation

A 70% of the serum samples were analyzed using two protocols. The remaining 30% were only analyzed using the first approach. For both protocols, serum pool aliquots of 100 μL were unfrozen slowly by immersion in an ice bath. The first protocol consisted of a simple deproteination. For this purpose, 200 μL of methanol was added to each aliquot for protein precipitation, vortexed for 10 min and centrifuged at 4 °C at 13 800× *g* for 5 min. The second protocol consisted of a solid-phase extraction step to isolate phospholipids from the sample. The well-established procedure has been already described in a previous publication [20].

2.5. LC–QTOF MS/MS serum analysis

Chromatographic separation was performed at 25 °C using a C18 reversed phase analytical column (100 mm × 4.6 mm i.d., 3 μm particle size) Mediterranea from Teknokroma (Barcelona, Spain). The mobile phases used were water (phase A) and ACN (phase B) both containing 0.1% formic acid as ionization agent. The LC pump was programmed to operate at a flow rate of 0.8 mL min^{-1} with the following elution gradient for analysis of deproteinized serum: 3% phase B for 1 min; phase B ramp from 3 to 100% from min 1 to min 17; and hold at 100% for 3 min. For the analysis of phospholipids, the elution gradient was 30% phase B for 2 min, phase B ramp from 30 to 70% from min 2

to 17, from 70% to 100% from min 17 to 25, and hold at 100% for 5 min. The injected volume was in both cases 5 μL and the injector needle was washed 10 times with 80% ACN between injections. Also, the needle seat back was flushed with 80% ACN at 4 mL min^{-1} for 10 s to avoid cross contamination. The autosampler was kept at 4 °C to increase sample stability.

The electrospray ionization source was operated in positive and negative ionization modes using the following settings: capillary and fragmentor voltage ±3.5 kV and 175 V, respectively; N_2 nebulizer gas pressure 40 psi; and N_2 drying gas flow rate and temperature 10 L min^{-1} and 325 °C, respectively. The instrument was daily calibrated and tuned as recommended by the manufacturer. MS and MS/MS data were collected in the centroid mode using a rate of 2.0 spectra per second in the extended dynamic range mode (2 GHz). Accurate mass spectra in auto MS/MS mode were acquired over the *m/z* range 60–1100, and MS/MS *m/z* range 31–1100. The instrument gave typical resolution 15 000 FWHM at *m/z* 118.0862 and 30 000 FWHM at *m/z* 922.0098. Mass accuracy in recorded ions was assured by continuous internal calibration during analyses by using the signals at *m/z* 121.0509 (protonated purine) and 922.0098 [protonated hexakis(1*H*,1*H*,3*H*-tetrafluoropropoxy)-phosphazine or HP-921] for positive ionization mode, while in negative ionization mode ions with *m/z* 119.0362 (proton abstracted purine) and *m/z* 966.0007 (formate adduct of HP-921) were used. The auto MS/MS mode was configured with 2 maximum precursors per cycle and an exclusion window of 0.25 min after 2 consecutive selections of the same precursor ion. The collision energy selected was different in the three replicates of each analysis (10, 20 and 40 V) to increase the level of MS/MS information.

2.6. Data processing and pretreatment

The MassHunter Workstation software package (B.07.00 Qualitative Analysis and B.06.00 Profinder, Agilent Technologies, Santa Clara, CA) was used to process all data obtained by LC–QTOF in the MS/MS mode. The recursive feature extraction algorithm in the software MassHunter Profinder was used to extract and align potential molecular features in all injections carried out with each ionization mode. This algorithm initially deconvolutes chromatograms and aligns features across the selected

sample files in terms of mass accuracy and retention time; then, it uses the mass and retention time of each molecular feature for recursive targeted extraction. This two-step procedure reduces the number of both false negatives and false positives in feature extraction. The target parameters for feature extraction included a threshold of 1500 counts for the monoisotopic peak and a maximum charge state of 2. In addition, the isotopic distribution of valid feature had to be defined by two or more ions, with a peak spacing tolerance of 0.0025 *m/z*, plus 10.0 ppm. Adduct formation (+Na, +K, $+NH_4$ in positive ionization mode and +COOH-H for negative ionization mode) was also used together with protonated and deprotonated ions to identify features associated to the same metabolite. The features were aligned by using a tolerance window of 0.30 min and a mass accuracy of 10 ppm for retention time and *m/z* value, respectively, across all data files.

The minimum absolute height required for feature extraction in the recursive step was set at 3000 counts for the sum of all peaks of the isotopic distribution of each molecular entity, which should be fulfilled in at least 75% of samples. Background contribution was removed by subtraction of entities linked to plasticizers, solvent impurities and other contaminants after analysing a blank. For this purpose, molecular features were also extracted from blanks raw data and those presented in the blank were excluded in the list of features detected in the samples. The resulting data set was exported as .csv file, containing the quantitative response expressed as peak area of the potential entities in each sample, and the main characteristics of these entities (retention time and neutral mass). Data pretreatment was based on baselining and normalization by quantile and logarithmic transformation to reduce relatively large differences among molecular features abundances.

2.7. Statistical analysis

The open source data analysis R program was used through its graphical user interface R-studio to apply and develop MetaboQC package by combination of different algorithms. The package is available at the Comprehensive R Archive Network (CRAN) Repository (https://cran.r-pro ject.org/index.html).

Statistical analysis was carried out by using the web based tool Metaboanalyst (http://www.metaboanalyst.ca/). The different data sets were transformed by logarithm prior to statistical analysis. A multivariate analysis by principal component analysis (PCA) was used to assess the variability of the cohorts. The METLIN, Human Metabolome Database (HMDB) and Massbank databases were used for metabolite identification from both MS and MS/MS information.

3. Results and discussion

3.1. Influence of experimental variability on metabolomics experiments

As stated above, most untargeted metabolomic studies based on MS analysis require injection of large batches of samples or even the integration of data from batches analyzed at different times. The main problem associated to this operational mode is the influence of experimental variability, with special emphasis on that affecting the instrumental response. A representative example is found in the analysis of glycerophospholipids, a family of lipids only differentiated by the molecule attached to the phosphate group. According to the bibliography, ionization efficiency of glycerophospholipid molecular species is predominantly dependent on the polar head and only modestly dependent on the acyl chains [21]. Thus, the analysts could expect that the quantitative response for the different families of glycerophospholipids would be similarly affected by the instrumental variability since the only difference is the conjugated polar group. However, this is not the most frequent situation as can be seen in Figure 1, which shows the variability in the quantitative response of glycerophospholipids pertaining to four different families obtained by LC–MS/MS analysis of a set (n=42) of QCs periodically inserted in a batch of 208 samples. As can be seen, while glycerophosphatidylinositols (PI) and glycerophosphatidic acids (PAs) were characterized by a similar pattern along the sequence of analyses, other families such as lysophosphatidylethanolamines (lysoPE) and glycerophosphatidylcholines (PCs) were affected by different effects. Thus, lysoPE(22:6n3) and lysoPE(O-18:1) reported sensitivity increases at different times within the sequence –particularly in

QCs around the 150th sample of the sequence for lysoPE(22:6n3) or in QCs around 75th sample for lysoPE(O-18:1). On the other hand, lysoPE (20:2), (20:4) and (22:6n6) were characterized by a lower variability in the quantitative response as compared to previous lysoPEs. Concerning detected PCs, they experienced a sensitivity increase in the first part of the sequence (up to injection 60); while some of them suffered a sensitivity decay from injection 50 to injection 100. On the other hand, two of the most concentrated PCs –(40:6) and (38:5)– did not report this particular sensitivity decrease. These concentration-dependent effects are typical of the ESI device considering that most of these analogous set of lipids are eluted in a narrow interval; creating a real competition among coeluted lipids to be ionized. A common effect to the four families of glycerophospholipids is that observed in the first part of the analysis sequence. As can be seen in Figure 1, sensitivity is generally increased up to injection 50, which could be attributed to a stabilization effect of the matrix components in the ESI device or in the chromatographic column.

A typical data preprocessing in metabolomics workflows using MS detection is known as Mass Spectrometry Total Useful Signal (MSTUS), frequently applied as such or after a logarithmic transformation. These two common normalization strategies (MSTUS with or without logarithmic transformation) were tested in the glycerophospholipids data set obtained by analysis of QCs to check the correction effects. Supplementary Figure 1 shows the variability trends resulting from plotting the two normalized data sets. The two normalization protocols reduced significantly, but not totally, the contribution of the experimental variability. Despite the use of QCs to monitor and correct instrumental variability has become more popular, there is not a defined optimum protocol to be implemented in sequences of metabolomics experiments to remove this source of variability. In fact, QCs are employed in most cases to filter metabolites that surpass preset cut-off values of variability.

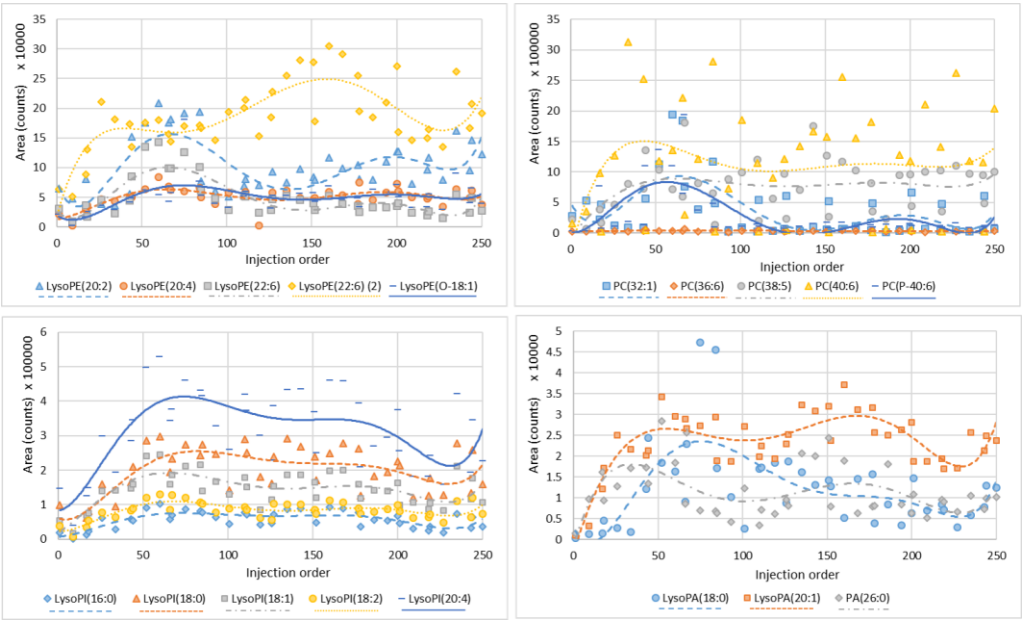


Figure 1. Peak area variability according to the injection order obtained for four classes of detected glycerophospholipids in the QCs inserted in the sequence of samples.

Based on this application, Dunn *et al.* proposed a cut-off value for variability in the quantitative response for each metabolite that should not surpass 20% in LC–MS data sets and 30% in GC–MS data sets [10]. According to this criterion all original data sets from glycerophospholipids would be filtered without normalization. On the other hand, only 20% of glycerophospholipids would be removed in the normalized data set by the MSTUS approach. Despite the normalization effect, the experimental variability is not completely removed, as Supplementary Figure 1 shows. Some other studies reported in the literature have described strategies for correction of experimental variability with the aid of QCs. These procedures are in some cases applied to the complete data set, which means that all metabolites detected in an experiment are equally treated [14]. However, it has been previously shown that not all metabolites are affected in the same way along a sequence of analyses or in batches of samples analyzed at different times. Other strategies only consider a unique variability factor such as the

injection order [18,19]. However, other factors such as the influence of daily cleaning or calibration procedures can introduce additional variability in the data set.

3.2. Theoretical performance

The package proposed here uses the QCs to examine the variability trends in the quantitative response of detected metabolites along the sequence of analysis, considering the number of batches and the injection order. For this purpose, QCs should be periodically inserted in the list of samples to be analyzed with the same method. The package performance can be split into three main steps, namely: (i) evaluation of variability trends in the quantitative response of each metabolite in QCs and samples; (ii) selection of an algorithm fitting the variability trend for each metabolite in the QCs; (iii) application of the algorithm to each metabolite in QCs and samples to evaluate its correction effect. The scheme of the operational pipeline is shown in Figure 2. The data inserted in the R-studio must include four columns with the following information: (i) sample name, (ii) injection order, (iii) classification code, sample or QC, and finally (iv) date of analysis. The data set is constituted by the response from the detected metabolites in each analysis (samples and QCs).

The process starts with the study of the instrumental variability by monitoring the quantitative response of the metabolites detected both in the samples and QCs, the latter being injected at least three times per day. For this purpose, the package includes a function that shows the evolution of the quantitative response for each metabolite along the sequence to obtain its variability trend. This function, named *graphQC*, generates a *.pdf* file with one graph per compound representing the quantitative response (peak area) of the metabolite *versus* the injection order for samples and QCs. These graphs can aid either to detect general variability trends during the analysis time or to identify random variability effects between days. Furthermore, it can also help to assess the suitability of the QCs, particularly, if they have a similar behavior to samples, since the correction is performed for each compound by construction of the trend curve for QCs and correction of the samples according to this curve.

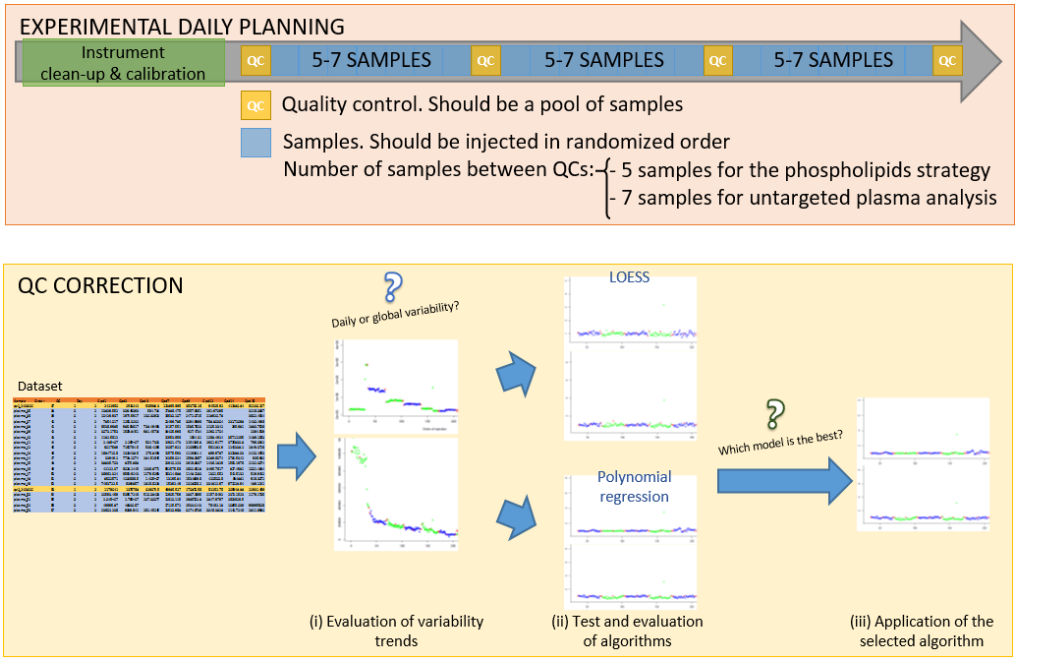


Figure 2. Experimental daily planning and operational pipeline of the MetaboQC package: (i) evaluation of variability trends in the quantitative response of each metabolite in QCs and samples; (ii) selection of an algorithm fitting properly the variability trend for each metabolite in the QCs; (iii) application of the correction algorithm for each metabolite in the QCs and samples to evaluate the correction effect.

Different algorithms with the same objective were programmed in this package to fit the experimental variability observed in the quantitative response of QCs. These algorithms differ in the function employed to build the trend curve –which can be either polynomial or based on local polynomial regression (LOESS)–, and also in the number of generated models, one for the whole sequence or one model per day when between-days variability is significant. Application of LOESS requires as many QCs per day as possible to obtain representative models. For this reason, when the number of QCs per day is below 4–5, this strategy should be applied by considering the complete sequence. To take into consideration applications involving three or a higher number of QCs injected per day, functions of polynomial equations with order 3, 4 and 6 can also been applied. The list of functions is included in the manual as Supplementary material S1.

The function reporting the best fit would be selected to correct the quantitative response of detected metabolites in the samples. Several functions can be tested to compare the correction effect. Additionally, the experimental variability can be estimated for each metabolite to compare the correction effect provided by each function. Once the correction algorithm is selected, it is applied to the complete data set obtained by analysis of the cohort under study. This operation generates a table in the workspace that can be easily exported to a *.csv* file containing the corrected responses, which are now ready for statistical analysis. It is worth mentioning that the *graphQC* function can be employed with any data set, normalized or not, original data sets or those from other software. Furthermore, since the process is easy and fast, the matrices obtained with different algorithms can be combined in case that some groups of metabolites perform better with regression models different from the rest of metabolites.

The practical operability of the tool proposed here was tested using two independent data sets provided by the analysis of serum samples with two different sample treatments: protein precipitation and analysis of the supernatant phase and SPE for isolation of glycerophospholipids and subsequent analysis. Serum was selected to test the efficiency of this tool as it is the most common clinical biofluid used in metabolomics studies and contains a wide chemical diversity of metabolites. Deproteinized samples were injected for 8 days, while samples after SPE were injected for 12 days. In both cases, at least three QCs were injected per day: one at the beginning of the daily sequence, other in the middle and the last one at the end of the day.

The selection of these two studies was supported on the experimental variability detected along the complete sequence of analysis. In fact, the two cases were identified as the two most common situations found by metabolomics analysts using MS based methods. In the case of glycerophospholipids analysis a variability trend was observed along the sequence of analysis with a clear increase of sensitivity. On the other hand, the analysis of deproteinized serum revealed variability trends with significant changes in sensitivity among days but also within a day. These two cases are separately exposed below.

3.3. Correction of variability trends along the sequence of analysis

As stated above, the package includes a function that generates a set of plots that can be used to study either the suitability of QCs to correct instrumental variability and/or if this is dominated by between-days or within-day variability. With the data set provided by analysis of glycerophospholipids, the plots obtained after this first step showed a continuous trend along the complete sequence, as shows Figure 1. The variability for glycerophospholipids detected in the sequence of samples was quite similar to that found in QCs; therefore, the experimental variability affecting the quantitative response of the metabolites can be modeled with that observed in the QCs. As Supplementary Figure 2 shows for three dominating lysoPC forms –particularly, lysoPC(18:1), (18:3) and (22:6)–, the experimental variability was characterized by a common pattern with a sensitivity increase along the sequence. Nevertheless, the relative variation for lysoPC(18:1) was clearly higher than that observed for the other two lysoPC forms. The use of an internal standard, even if this is an isotopically labeled compound pertaining to the same chemical family –lysophosphatidylcholines–, would not exert an optimum correction since the three selected lysoPCs were affected by the instrumental variability in a different way. The application of an unsupervised analysis by PCA on QCs showed a clear discrimination between analyses carried out in the first four days (Figure 3); thus, the instrumental variability trend detected for most metabolites could be associated to the analysis sequence.

In this scenario, the most adequate functions to fit these variability trends are those supported on a single equation either based on a polynomial model or on the LOESS algorithm. In the batch of samples under study, both functions were tested (*QCcorrectionLOESS* and *QCcorrectionSinglePolyX*, being X the grade of the polynomial function).

In the case of the polynomial function, grade 3, 4 and 6 were tested to check the effect on the variability correction due to the high number of QCs considered to build the model. However, the classical polynomial regression sometimes causes wrong trends that are avoided in the case of LOESS. For this reason, LOESS provided better results in this example as compared to polynomial regression.

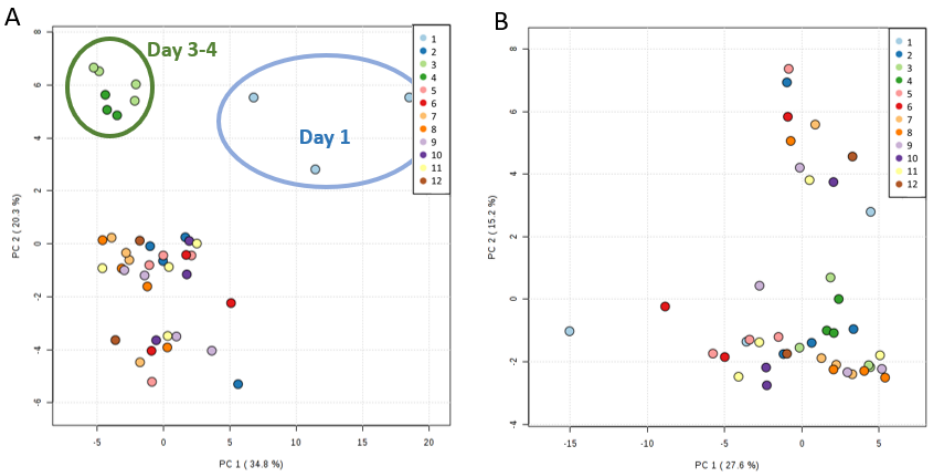


Figure 3. PCA obtained for the phospholipids data set before (A) and after (B) variability correction using MetaboQC package.

Supplementary Figure 2 illustrates the variability plots provided by application of the LOESS and grade 4 and 6 polynomial algorithms that exhibit a noticeable correction effect. To compare in overall terms the correction effect for all detected metabolites, a visual alternative is to represent the corrected variability in QCs expressed as RSD percentage as compared to the variability of the original data set. An example of this is found in Figure 4 that shows the correction effect after application of the LOESS and grade 4 polynomial functions to the glycerophospholipids data set. Positive values imply a reduction of RSD attained by application of the algorithm, while negative values highlight that variability was not reduced by removal of experimental contribution.

Several effects can be deduced from this representation. First of all, a different behavior of the two functions according to the ionization mode was detected. Thus, the polynomial function was able to reduce RSD more efficiently in the positive ionization mode than did the LOESS function. The opposite situation was found in negative ionization, where LOESS led to a more efficient correction in general, although some exceptions were observed.

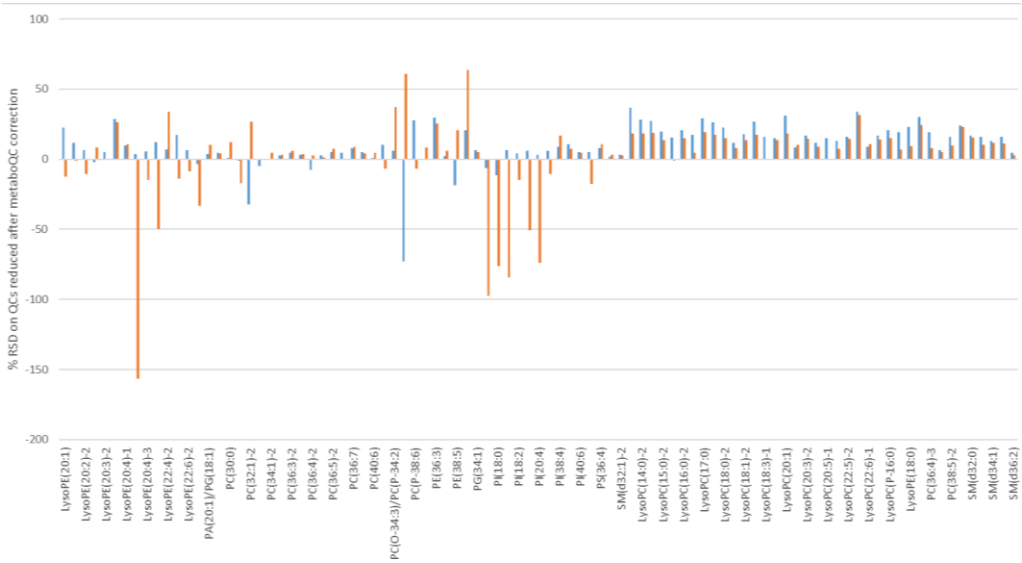


Figure 4. Percentage of variability (expressed as RSD) corrected after application of the LOESS algorithm (grey) and the polynomial grade 4 algorithm (black). Negative values imply an increase in variability after the application of the algorithm.

Thus, the most relevant change was found for glycerophosphatidylinositols, PI, for which the application of the LOESS function did not work properly. On the other hand, the LOESS function suitably corrected the instrumental variability in PCs and PEs. With these results, we could select the correction function to be applied for each ionization mode and, more exhaustively, the correction for each family of metabolites or even for each compound. With the application of the suited algorithm for each independent metabolite according to the corrected RSD, the effect was evaluated by comparing the number of metabolites presenting significant differences between days along the sequence. This number was considerably reduced from 50 (86%) to 11 metabolites (19%) in negative ionization mode and from 38 (100%) to 5 metabolites (13%) in positive ionization mode, which means that the proportion of significant glycerophospholipids was reduced in both cases from more than 80% to less than 20%. Additionally, the corrected data set was analyzed by PCA to detect variability trends and no groupings as a function of the sequence day were detected, as Figure 3 shows.

3.4. Correction of between-days variability trends

The experiment about untargeted analysis of serum samples was used as a model to correct between-days variability trends. This experiment was affected by a strong influence of experimental variability on the quantitative response. Figure 5 shows an example of this influence for four representative metabolites detected in serum, which were selected according to their retention time. The variability trends were completely different to those observed in the previous case. This experiment was characterized by implementation of a daily protocol for cleaning of the ESI source, which was always applied at the beginning of the day. Additionally, a protocol for calibration of the TOF instrument was also daily programmed. These practices are clearly the main reason contributing to the observed variability between two consecutive days. As can be seen, this strong between-days variability affected in the same way to QCs.

In this particular case, it is recommended to use multiple models, one per day, to correct instrumental variability, as there is no a continuous trend along the complete sequence. As expected, the variability plots were different for the four representative metabolites, proving that the instrumental variability does not affect in the same manner to all the target compounds. The two available algorithms were used to prepare the corresponding functions to be applied along the different days of the sequence. These functions were named *QCcorrectionMultiLOESS* and *QCcorrectionMultiPolyX*, being *X* the grade of the polynomial function (typically 3, 4 and 6). Both the *MultiLOESS* and *MultiPolyX* functions are recommended in cases in which there are 4 or more QCs per day. On the other hand, when the number of QCs per day is limited to three, the *QCcorrectionMultiPoly3* should give an optimum fit.

Nevertheless, due to the simplicity of the process, all models can be easily tested to finally select the best for each specific case. In the present study, the *QCcorrectionMultiPoly3* function was selected since only three QCs per day were implemented in the injections sequence. The application of this correction model leads to data following the distributions reported in Figure 6. As can be seen, the quantitative responses for the selected metabolites in QCs were scaled to one and the instrumental variability preliminary observed in the sequence of samples was considerably decreased.

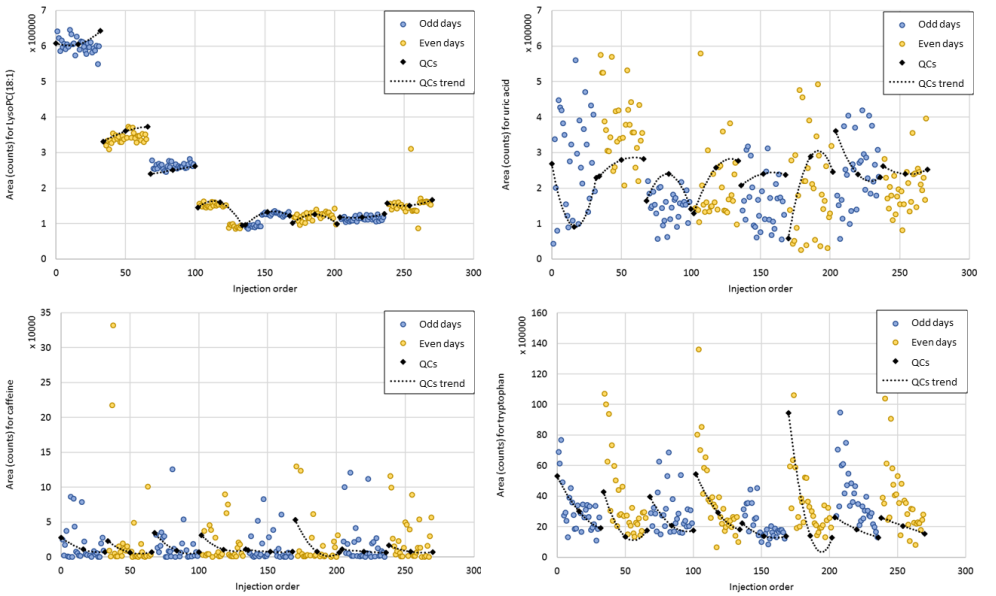


Figure 5. Peak area variability according to the injection order of four characteristic compounds (LysoPC(18:1), caffeine, uric acid and tryptophan) detected in serum samples and QCs during the injections sequence.

Similarly, to the previous study, the PCA for the uncorrected data showed a clear discrimination through PC1 for samples injected within the first four days and those injected from the fifth day until the end of the sequence. Additionally, PC2 allowed separating samples according to the injection order within each day. These separation trends were removed when the PCA was applied to the corrected data set since the PCA after correction does not show this separation or any trend that could be associated to the injection order.

The reduction of instrumental variability can be also observed on the violin plots obtained for the four representative metabolites previously selected. The violin plot is a visual tool providing an idea of the distribution or density of a parameter in a data set. As can be seen in Supplementary Figure 3, the variability was reduced in all cases with special emphasis on the cases of lysoPC(18:1) and uric acid.

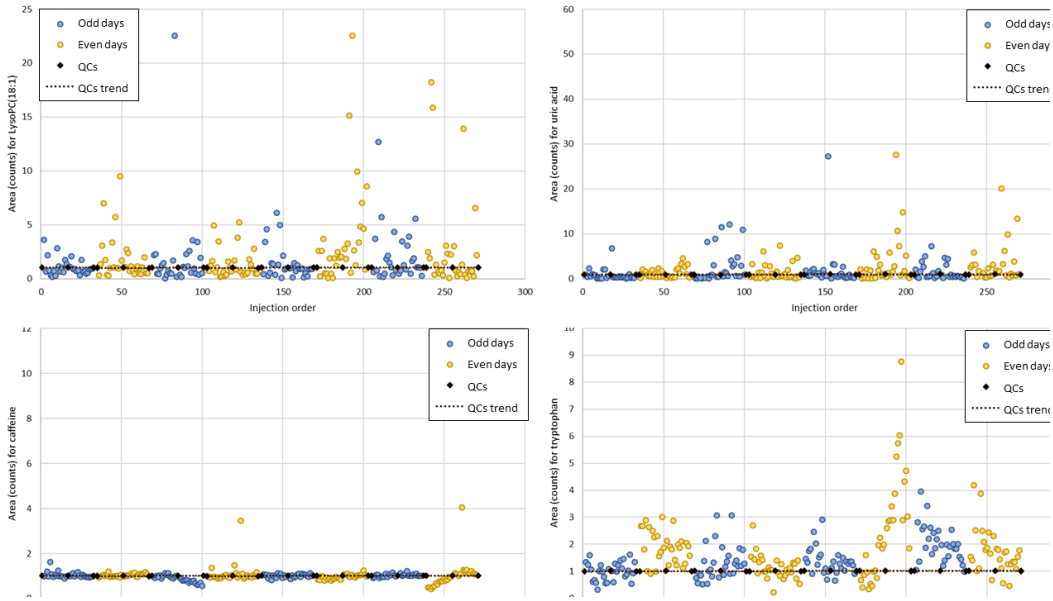


Figure 6. Peak area variability according to injection order for four compounds (lysoPC(18:1), uric acid, caffeine and tryptophan) detected in serum samples and QCs after application of a multipoly3 correction with the MetaboQC package.

4. Conclusions

A tool for correction of experimental variability associated to the instrumental quantitative response has been developed for implementation in metabolomics workflows based on MS detection. The proposed packaged is based on functions that can be used to correct variability on data sets obtained in metabolomics studies based on large set of samples. Different functions have been considered to make the package useful to remove experimental variability sources affecting long sequences of analysis but also those affecting within a given day. This type of tools can be particularly efficient in cases in which the use of internal standards is not operative due to chemical diversity of metabolome composition.

The strategy included in the proposed package involves that each metabolite is corrected according to the function that best fits its variability trend. Therefore, correction is independently applied to each metabolite. The only requirement for its application is the implementation of QCs, preferentially prepared with the same samples of the cohort, in the sequence of analysis following a given plan.

Akcnowledgments

This research has been jointly financed by the Spanish Ministerio de Economía y Competitividad, ISCIII-Subdirección General de Evaluación, Fondo Europeo de Desarrollo Regional (FEDER) and Junta de Andalucía through Projects CTQ2015-68813-R, PIE14/00005/1 (FLEXI-MET) and FQM-1602. M.A. Lopez-Bascon also thanks the MICINN for an FPU scholarship (FPU15/02359).

References

[1] B. Álvarez-Sánchez, F. Priego-Capote, M.D. Luque de Castro, Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation, TrAC - Trends Anal. Chem. 29 (2010) 111–119.

[2] R. Ramautar, G.W. Somsen, G.J. de Jong, The Role of CE-MS in Metabolomics, in: Metabolomics Pract., Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, 2013: pp. 177–208.

[3] O. Fiehn, Metabolomics by Gas Chromatography-Mass Spectrometry: Combined Targeted and Untargeted Profiling, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2016.

[4] P.A. Vorkas, G. Isaac, M.A. Anwar, A.H. Davies, E.J. Want, J.K. Nicholson, et al., Untargeted UPLC-MS profiling pipeline to expand tissue metabolome coverage: application to cardiovascular disease, Anal. Chem. 87 (2015) 4184–4193.

[5] M.M. Delgado-Povedano, M. Calderón-Santiago, F. Priego-Capote, M.D. Luque de Castro, Development of a method for enhancing metabolomics coverage of human sweat by gas chromatographyemass spectrometry in high resolution mode, Anal. Chim. Acta 905 (2016) 115–125.

[6] M. Calderón-Santiago, F. Priego-Capote, M.D.L. de Castro, Enhancing detection coverage in untargeted metabolomics analysis by solid-phase extraction on-line coupled to LC-MS/MS, Electrophoresis 36 (2015) 2179–2187.

[7] Peng, Y.T. Chen, C.L. Chen, L. Li, Development of a universal metabolome-standard method for long-term LC-MS metabolome profiling and its application for bladder cancer urine-metabolite- biomarker discovery, Anal. Chem. 86 (2014) 6540–6547.

[8] C. Bueschl, R. Krska, B. Kluger, R. Schuhmacher, Isotopic labeling-assisted metabolomics using LC–MS, Anal. Bioanal. Chem. 405 (2013) 27–33.

[9] Y. V Karpievitch, S.B. Nikolic, R. Wilson, J.E. Sharman, L.M. Edwards, Metabolomics data normalization with EigenMS, PLoS One 9 (2014) e116221.

[10] W.B. Dunn, D. Broadhurst, P. Begley, E. Zelena, S. Francis-McIntyre, N. Anderson, *et al.*, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry, Nat. Protoc. 6 (2011) 1060–1083.

[11] M. Vinaixa, S. Samino, I. Saez, J. Duran, J.J. Guinovart, O. Yanes, A guideline to univariate statistical analysis for LC/MS-based untargeted metabolomics-derived data, Metabolites 2 (2012) 775–795.

[12] M. Chen, R.S.P. Rao, Y. Zhang, C.X. Zhong, J.J. Thelen, A modified data normalization method for GC-MS-based metabolomics to minimize batch variation., Springerplus 3 (2014) 439.

[13] M.R.N. Ranjbar, M.G. Tadesse, Y. Wang, H.W. Ressom, Normalization of LC-MS data using Gaussian process, 2012 IEEE Int. Work. Genomic Signal Process. Stat. (2012) 187–190.

[14] M.R. Nezami Ranjbar, Y. Zhao, M.G. Tadesse, Y. Wang, H.W. Ressom, Gaussian process regression model for normalization of LC-MS data using scan-level information., Proteome Sci. 11 (2013) S13.

[15] B.A. Ejigu, D. Valkenborg, G. Baggerman, M. Vanaerschot, E. Witters, J.C. Dujardin, *et al.*, Evaluation of normalization methods to pave the way towards large-scale LC-MS-based metabolomics profiling experiments, Omics 17 (2013) 473–85.

[16] T. Sangster, H. Major, R. Plumb, A.J. Wilson, I.D. Wilson, A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis, Analyst 131 (2006) 1075–1078.

[17] P. Masson, A.C. Alves, T.M.D. Ebbels, J.K. Nicholson, E.J. Want, Optimization and evaluation of metabolite extraction protocols for untargeted metabolic profiling of liver samples by UPLC-MS, Anal. Chem. 82 (2010) 7779–7786.

[18] X. Shen, X. Gong, Y. Cai, Y. Guo, J. Tu, H. Li, *et al.*, Normalization and integration of large-scale metabolomics data using support vector regression, Metabolomics 12 (2016) 89.

[19] R. Wehrens, B.A. Jos Hageman, B. Fred van Eeuwijk, B. Rik Kooke, B.J. Pádraic Flood, B. Erik Wijnker, et al., Improved batch correction in untargeted MS-based metabolomics, Metabolomics 12 (2016) 88.

[20] M. Calderón-Santiago, F. Priego-Capote, J.G. Galache-Osuna, M.D. Luque de Castro,

Analysis of serum phospholipid profiles by liquid chromatography-tandem mass spectrometry in high resolution mode for evaluation of atherosclerotic patients, J. Chromatogr. A 1371 (2014) 154–62.

[21] X. Han, R.W. Gross, Shotgun lipidomics: Electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples, Mass Spectrom. Rev. 24 (2005) 367–412.

Supplementary Information

Figure Supplementary 1. Peak area variation of some glycerophospholipids detected in the QCs injected along a sequence. Original data were normalized by MSTUS with and without the application of logarithmic transformation: (A) Data normalized by MSTUS, (B) Data normalized by MSTUS and transformed by application of base 2 logarithm.

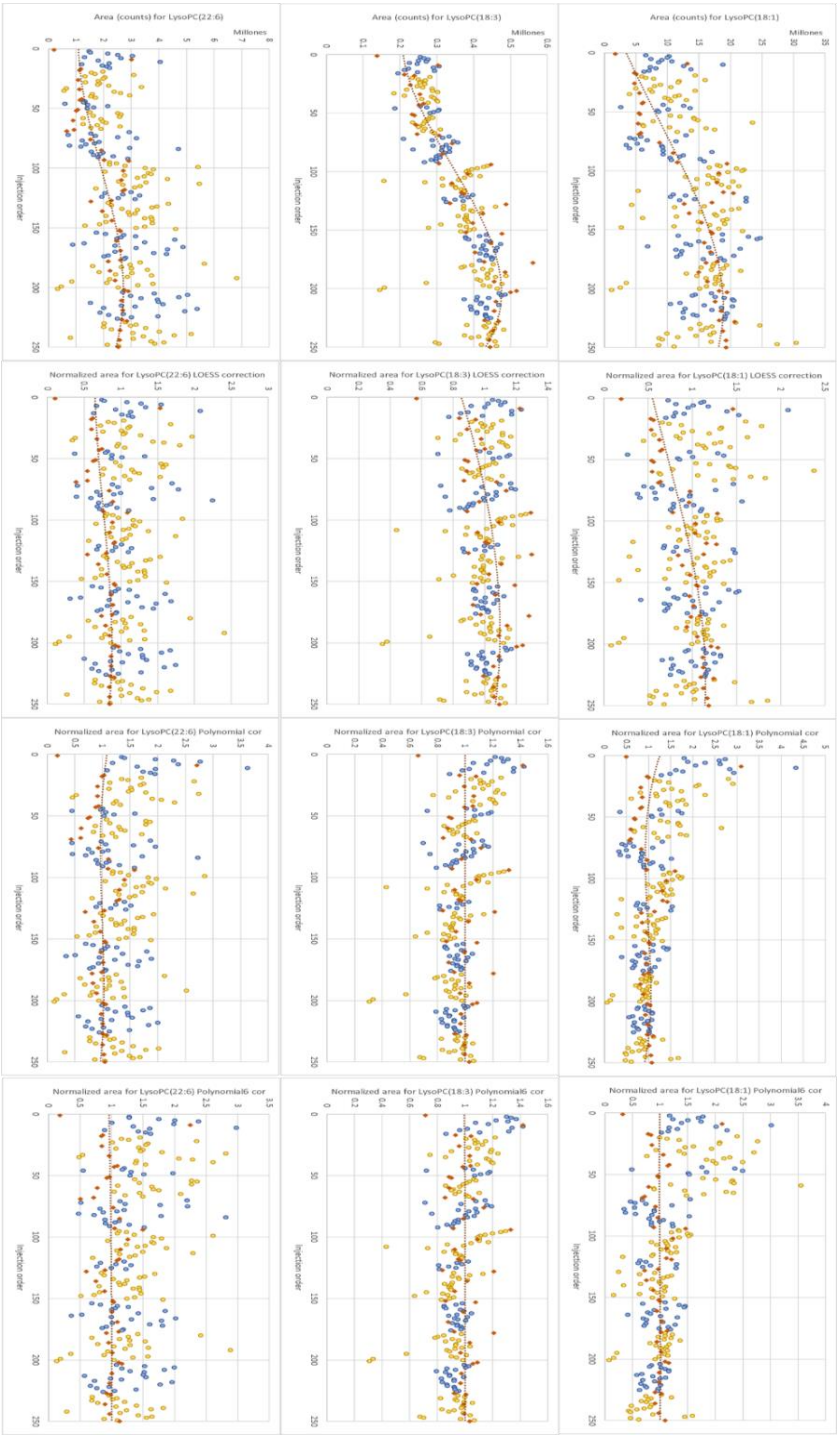


Figure Supplementary S2. Peak area variability according to the injection order for lysoPC(18:1), lysoPC(18:3) and lysoPC(22:6) detected in serum samples and QCs without correction, and after application of the LOESS, polynomial grade 4 and polynomial grade 6 algorithms included in the MetaboQC package.

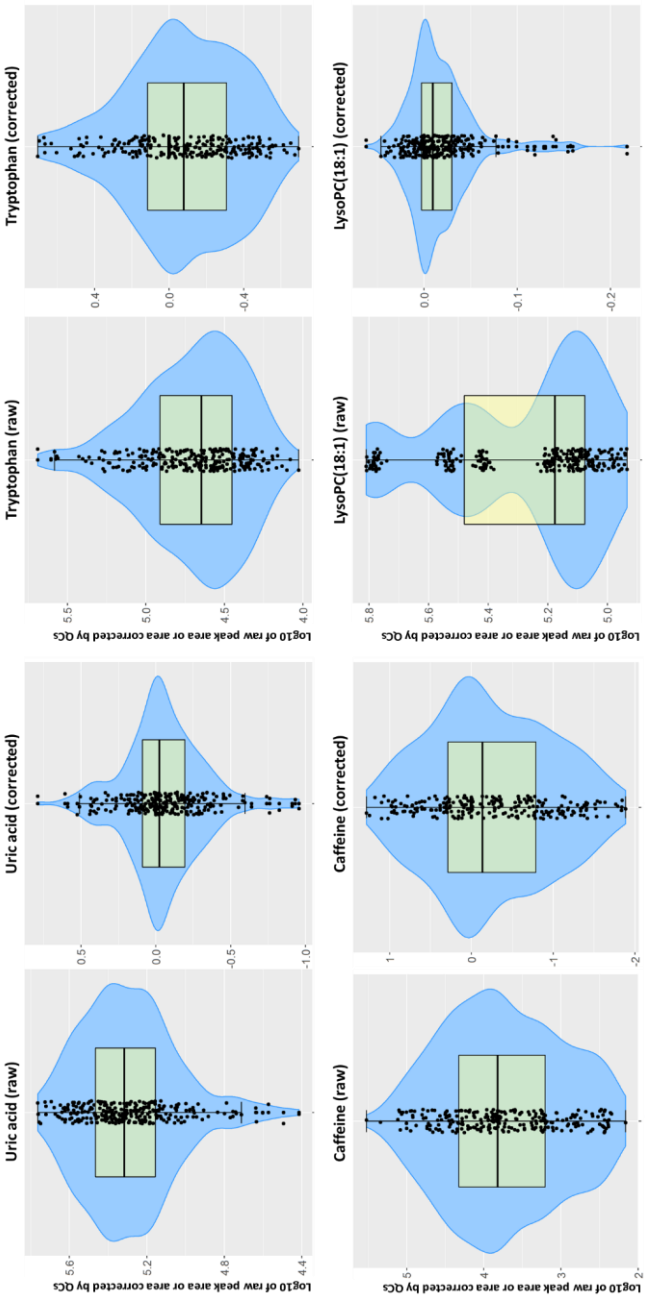


Figure Supplementary S3. Violin plots obtained for the log 10 of the raw peak area value and after QC correction for uric acid, caffeine, tryptophan and lysoPC(18:1).

MetaboQC

Correct instrumental variability from metabolomic studies using quality controls

Description:

Use the peak area of each compound detected in quality control samples to find the instrumental variability trend and use the generated curve for correction in samples.

Requirements:

Required package: plyr

Functions:

	Function	**Description**
	graphQC	Gives a pdf file with one graph per compound showing the variability of the peak area along a sequence of analysis
Polynomial correction	*QCcorrectionSinglePoly*	Uses an order 3 polynomial function to fit the variability trend of the peak area of each metabolite
	QCcorrectionSinglePoly4	Uses an order 4 polynomial function to fit the variability trend of the peak area of each metabolite
	QCcorrectionSinglePoly6	Uses an order 6 polynomial function to fit the variability trend of the peak area of each metabolite
	QCcorrectionMultiPoly	Uses an order 3 polynomial function to fit the variability trend of the peak area of each metabolite per day
	QCcorrectionMultiPoly4	Uses an order 4 polynomial function to fit the variability trend of the peak area of each metabolite per day
	QCcorrectionMultiPoly6	Uses an order 6 polynomial function to fit the variability trend of the peak area of each metabolite per day

LOESS correction	*QCcorrectionLOESS*	Uses local polynomial regression to fit the variability of the peak area of each metabolite along a sequence of analysis
	QCcorrectionMultiLOESS	Uses local multipolynomial regression to fit the variability of the peak area of each metabolite along a sequence of analysis

Performance:

1. *graphQC*

The data set is introduced in R as a .csv file containing the following columns:

- A first column with sample name.
- A second one with the injection order, called 'Order'.
- A third column specifies the quality control analysis. This column is named 'QC' and samples are represented by '0' while QC are represented by '1'.
- The 4th column includes the day of analysis. This column, called 'Day' contains a numeric value for each day. It is preferable to start with day 1 and, then, continue the list.
- The following remaining columns contain the peak area of each potential metabolite detected in the sample list.

Sample Name	Order	QC	Day	Compound 1	Compound 2	Compound 3	...
QC1-Day1	1	1	1	21122232	28151347	475068	
Sample1	2	0	1	33825160	764442	197000	
Sample2	3	0	1	19700000	9336217	425023	
QC2-Day1	4	1	1	20845767	28421872	495345	
QC1-Day2	5	1	2	25300000	32466685	345722	

The matrix should be introduced in the workspace using the *read.table* function, with the next parameters (sep="","", header=TRUE, row.names=NULL, as.is=1).

Example: LCMSData<-read.table(file="LCMSdata.csv", sep="","", header=TRUE, row.names=NULL, as.is=1)

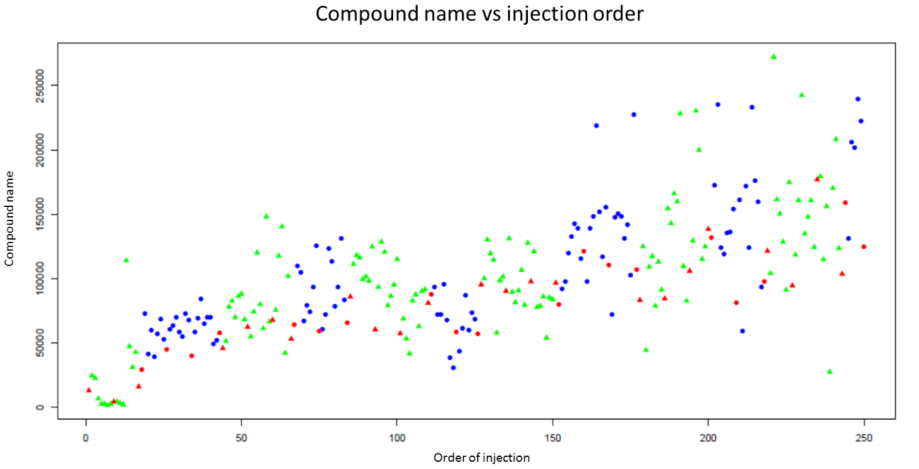
Note: Files that are going to be incorporated to the workspace or environment should be in the working directory, so first of all it is recommended to establish the correspondent working directory.

Usage: *graphQC*(LCMSData, N, "Text")

Arguments:

LCMSData	The matrix that is going to be normalized (Sample Name, Order, QC, Day and multiple columns with the peak area of the compounds). It can be also a normalized data set
N	A number corresponding to the number of metabolites to be studied. A variability plot will be obtained per metabolite
Text	A text to be included as name of the pdf file generated to differentiate it

The pdf file generated is called: 'xy-Plot for *N* compounds form *Text*.pdf' and each page contains a graph similar to that:



As can be seen, QCs are represented in red color, odd days are represented by triangles, while even days are represented by circles. Furthermore, samples from odd days are colored in green, while that from even days are colored in blue.

2. *QCcorrection functions (QCcorrectionSinglePoly, QCcorrectionSinglePoly4, QCcorrectionSinglePoly6, QCcorrectionMultiPoly, QCcorrectionMultiPoly4, QCcorrectionMultiPoly6, QCcorrectionLOESS, QCcorrectionMultiLOESS)*

Usage: LCMSDataNormalized <- *QCcorrectionfunction*(LCMSData)

Arguments:

LCMSData	The matrix that is going to be normalized (Sample Name, Order, QC, Day and multiple columns with the peak area of the compounds)

QCcorrectionfunction	One of the functions used to fit the variability trend of metabolites detected in QC and and apply to samples
LCMSDataNormalized	The final dataset already normalized according to the function selected. This table can be exported into a *.csv* file using the *write.table* function (example: write.table(LCMSDataNormalized, file="LCMSDataNormalized.csv", sep=",", row.names=FALSE))

Example of the generated tables:

Sample Name	Order	QC	Day	Compound 1_Norm	Compound 2_Norm	Compound 3_Norm	...
QC1-Day1	1	1	1	1.006	0.995	0.997	
Sample1	2	0	1	1.612	0.027	0.410	
Sample2	3	0	1	0.934	0.330	0.876	
QC2-Day1	4	1	1	0.993	1.004	1.003	
QC1-Day2	5	1	2	1.010	0.996	1.001	

SECTION III

Strategies for searching potential biomarkers in clinical and agro-food studies

Metabolomics is an interesting tool for assessing the nutritional status of subjects, the food consumption, the biological consequences of following a nutritional intervention, or the study of metabolic mechanisms associated with a disease in response to a diet depending on a particular metabolic phenotype. Section III of this PhD Book is devoted to strategies for searching potential biomarkers in nutrimetabolomics with clinical and agro-food applications.

Chapter VII, VIII and IX are focused on clinical applications. It is widely known that the postprandial response to a meal depends on many factors and involves multiple processes that include energy storage and metabolic switch in several organs such as liver, muscle and adipose tissue, accompanied by several compensating processes such as inflammation and oxidative stress. The aim of chapter VII was to detect postprandial alterations in the level of plasma metabolites after an OFTT. For this purpose, plasma samples were collected just before and four hours after the OFTT. These samples were analyzed by an untargeted approach using LC–QTOF MS/MS and GC–TOF/MS.

For Chapter VIII the strategy was quite similar, but in this case the aim was to create a panel of biomarkers to predict the occurrence of T2DM by examining the postprandial response of 215 non-diabetic patients after OFTT. In the baseline, plasma samples were collected just before and four hours after the OFFT and analyzed by an untargeted approach using LC–QTOF MS/MS and GC–TOF/MS. Patients were clinically monitored for 5 years and, after this period, 107 patients were diagnosed with T2DM. Metabolic changes occurring in the postprandial OFTT were used to predict the development of the disease and were associated to the diagnostic date by Cox Regression Analysis.

The last clinical application was aimed to elucidate the early events preceding the onset of islet autoimmunity and overt T1DM (Chapter IX). LC–QTOF MS/MS and GC–Q/MS approaches were used to determine levels of

molecular lipids and polar metabolites in human PBMCs isolated from prospective samples collected in the DIPP study.

On the other hand, Chapter X is focused on an agro-food application. Discrimination among the types of feeding regimes for Iberian pigs is currently a highly demanded challenge by the Iberian pig sector. For this reason, the combination of two analytical methods, previously used independently without success, was used to developed discrimination models and search panels of biomarkers with capability to classify slaughtered animals by their feeding regime. These analytical methods were based on FAMEs analysis by GC–FID and determination of $\delta^{13}C$ by IRMS.

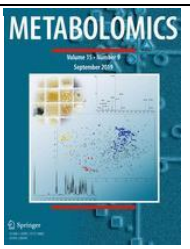
CHAPTER VII

Postprandial metabolic response to the oral fat tolerance test (OFTT) by plasma metabolomics analysis

Springer

Sent to

Metabolomics



Postprandial metabolic response to the oral fat tolerance test (OFTT) by plasma metabolomics analysis

María Asunción López-Bascón[a,b,c,d], Mónica Calderón-Santiago[a,b,c,d*], Antonio Camargo[a,e,f], José López-Miranda[a,e,f], Feliciano Priego-Capote[a,b,c,d*]

[a]*Maimónides Institute for Biomedical Research (IMIBIC), Reina Sofía University Hospital, Córdoba, Spain.*

[b]*Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain.*

[c]*CeiA3 Agroalimentary Excellence Campus, University of Córdoba, Córdoba, Spain.*

[d]*CIBER Fragilidad y Envejecimiento Saludable (CIBERFES), Carlos III Health Institute, Madrid, Spain.*

[e]*Lipid and Atherosclerosis Unit, Reina Sofía University Hospital, Cordoba, Spain.*

[f]*CIBER Fisiopatologia Obesidad y Nutricion (CIBEROBN), Carlos III Health Institute, Madrid, Spain.*

Postprandial metabolic response to the oral fat tolerance test (OFTT) by plasma metabolomics analysis

María Asunción López-Bascón, Mónica Calderón-Santiago, Antonio Camargo, José López-Miranda, Feliciano Priego-Capote

ABSTRACT

Oral fat tolerance test (OFTT) is used to evaluate postprandial lipemia by monitoring the concentration of triglycerides after a weight-adjusted meal containing lipids, carbohydrates and proteins. With these premises, the present research is focused on detecting postprandial alterations in the level of plasma metabolites after the OFTT on 215 patients involved in the CORDIOPREV study. Samples were collected just before and four hours after the OFTT and analyzed by LC–QTOF MS/MS and GC–TOF/MS in two different batches including 57 and 158 individuals. A total number of 365 metabolites were tentatively identified by combination of both analytical platforms. The paired *t*-test led to the identification of 33 metabolites significantly altered ($p<0.05$) in both batches due to the OFTT. Special attention was paid to fatty acids, their derivatives, bile acids and acylcarnitines, since these families comprised the 75% of significant metabolites. The most important pathways affected by OFTT were those involved in inflammatory and oxidative processes, *de novo* lipogenesis, metabolism of acylcarnitines and primary and secondary bile acids, and the synthesis of cortisol. In most cases, significant metabolites increased after OFTT, with special relevance of docosapentaenoic acid (C22:5n3) and palmitoleic acid (C16:1n7) that presented a fold-change (FC) above 60, followed by octadecenoylcarnitine, which reported an FC above 38. Within metabolites reducing their levels after OFTT, acetylcarnitine, carnitine and dodecanoylcarnitine presented the lower values with FC equal or below –0.8. This information opens the door to increase the applicability of the OFTT in the evaluation of lipemia as well as other metabolic diseases.

Keywords: Oral fat tolerance test; metabolomics profile; plasma; postprandial period; untargeted analysis

CHAPTER VIII

Multimetabolite panels predict the occurrence of type 2 diabetes mellitus by examining the postprandial response

Sent to

Molecular and Cellular Endocrinology

Multimetabolite panels predict the occurrence of type 2 diabetes mellitus by examining the postprandial response

María Asunción López-Bascón[a,b,c,d], Mónica Calderón-Santiago[a,b,c,d*], Antonio Camargo[a,e,f], José López-Miranda[a,e,f], Feliciano Priego-Capote[a,b,c,d*]

[a]*Maimónides Institute for Biomedical Research (IMIBIC), Reina Sofía University Hospital, Córdoba, Spain.*

[b]*Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain.*

[c]*CeiA3 Agroalimentary Excellence Campus, University of Córdoba, Córdoba, Spain.*

[d]*CIBER Fragilidad y Envejecimiento Saludable (CIBERFES), Carlos III Health Institute, Madrid, Spain.*

[e]*Lipid and Atherosclerosis Unit, Reina Sofía University Hospital, Cordoba, Spain.*

[f]*CIBER Fisiopatologia Obesidad y Nutricion (CIBEROBN), Carlos III Health Institute, Madrid, Spain.*

Multimetabolite panels predict the occurrence of type 2 diabetes mellitus by examining the postprandial response

María Asunción López-Bascón, Mónica Calderón-Santiago, Antonio Camargo, José López-Miranda, Feliciano Priego-Capote

ABSTRACT

Understanding the pathogenesis of type 2 diabetes mellitus (T2DM) is a requisite for the development of preventive strategies. Particularly, the identification of early metabolic alterations is a promising challenge in the study of etiological pathways and may further help to identify high-risk individuals. The present research is focused on predicting the occurrence of T2DM in individuals by examining their postprandial response to the oral fat tolerance test (OFTT). For this purpose, plasma samples at baseline from 215 non-diabetic patients (107 out of them were diagnosed with T2DM in the following five years) were analyzed using a metabolomics untargeted approach. Sixty metabolites showed significant concentration changes in the postprandial associated with the development of T2DM. Two multimetabolite panels were created to predict the occurrence of T2DM by prioritizing specificity (panel 1) or sensitivity (panel 2). The combination of both panels reported sensitivity of 86.6% and specificity of 71.6%, while external validation led to 90.0% of sensitivity and 73.3% of specificity. The risk of T2DM development was evaluated by Cox Regression Analysis that provided Hazard Ratio values of 6.5 (3.7–11.4) and 5.4 (3.0–9.6) for panels 1 and 2, respectively. Metabolites included in the panels were associated to relevant pathways such as oxidation, insulin secretion, and mitochondrial and peroxisomes activity.

Keywords: Metabolomics profile, type 2 diabetes mellitus, biomarkers, plasma, oral fat tolerance test, postprandial period

CHAPTER IX

Metabolic alterations in human peripheral blood mononuclear cells associate with progression to islet autoimmunity and type 1 diabetes

Springer

Accepted in

Diabetologia

Metabolic alterations in human peripheral blood mononuclear cells associate with progression to islet autoimmunity and type 1 diabetes

Partho Sen[a]*, Alex M. Dickens[a], María Asunción López-Bascón[b,c], Tuomas Lindeman[a], Esko Kemppainen[a], Santosh Lamichhane[a], Tuukka Rönkkö[a], Jorma Ilonen[d,e], Jorma Toppari[f,g], Riitta Veijola[g,h,i], Heikki Hyöty[k,m], Tuulia Hyötyläinen[c], Mikael Knip[n,ñ]*, Matej Orešič[a,o]*

[a]*Turku Bioscience, University of Turku and Åbo Akademi University, Turku, Finland*

[b]*Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain*

[c]*Department of Chemistry, Örebro University, Örebro, Sweden*

[d]*Immunogenetics Laboratory, Institute of Biomedicine, University of Turku, Turku, Finland*

[e]*Clinical Microbiology, Turku University Hospital, Turku, Finland*

[f]*Department of Paediatrics and Adolescent Medicine, Turku University Hospital, Turku, Finland*

[g]*Institute of Biomedicine, Centre for Integrative Physiology and Pharmacology, University of Turku, Turku, Finland*

[h]*Department of Pediatrics, PEDEGO Research Unit, Medical Research Centre, University of Oulu, Oulu, Finland*

[i]*Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland*

[j]*Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden*

[k]*Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland*
[m]*Fimlab Laboratories, Pirkanmaa Hospital District, Tampere, Finland*
[n]*Children's Hospital, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland*
[ñ]*Tampere Centre for Child Health Research, Tampere University Hospital, Tampere, Finland*
[o]*School of Medical Sciences, Örebro University, Örebro, Sweden*

Metabolic alterations in human peripheral blood mononuclear cells associate with progression to islet autoimmunity and type 1 diabetes

Partho Sen, Alex M. Dickens, María Asunción López-Bascón, Tuomas Lindeman, Esko Kemppainen, Santosh Lamichhane, Tuukka Rönkkö, Jorma Ilonen, Jorma Toppari, Riitta Veijola, Heikki Hyöty, Tuulia Hyötyläinen, Mikael Knip, Matej Orešič

ABSTRACT

Previous metabolomics studies suggest that type 1 diabetes (T1DM) is preceded by specific metabolic disturbances. Here we asked whether distinct metabolic patterns occur in peripheral blood mononuclear cells (PBMCs) of children later developing pancreatic β-cell autoimmunity or overt T1DM. In a longitudinal cohort setting, PBMC metabolomic analysis was applied in children who either (1) progressed to T1DM (PT1D, n=34), (2) seroconverted to ≥1 islet autoantibody without progressing to T1DM (P1Ab, n=27), or (3) remained autoantibody negative during follow-up (CTRL, n=10). During the first year of life, levels of most lipids and polar metabolites were lower in PT1D and P1Ab, *vs.* CTRLs. Pathway overrepresentation analysis suggested alanine, aspartate, glutamate, glycerophospholipid and sphingolipid metabolism were overrepresented in PT1D. Genome-scale metabolic models of PBMCs in T1DM progression were developed using available transcriptomics data and constrained with metabolomics data from our study. Metabolic modeling confirmed altered ceramide pathways as specifically associated with T1DM progression.

Keywords: birth cohort, ceramides, genome-scale metabolic modeling, lipidomics, metabolomics, PBMCs, peripheral blood mononuclear cells, sphingolipid metabolism, type 1 diabetes mellitus

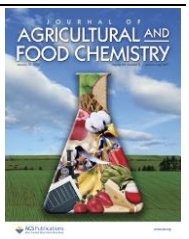
CHAPTER X

Determination of fatty acids and stable carbon isotopic ratio in subcutaneous fat to identify the feeding regime of Iberian pigs

ACS Publications

Journal of Agricultural and Food Chemistry

63 (2015) 692–699



Determination of fatty acids and stable carbon isotopic ratio in subcutaneous fat to identify the feeding regime of Iberian pigs

María Asunción López-Bascón[a,b,c], Feliciano Priego-Capote[a,b,c*], Mónica Calderón-Santiago[a,b,c], Verónica Sánchez de Medina[a,b,c], José Manuel Moreno-Rojas[d], Juan María García-Casco[e], María Dolores Luque de Castro[a,b,c*]

[a]*Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain*

[b]*CeiA3 Agroalimentary Excellence Campus, University of Córdoba, Córdoba, Spain*

[c]*Maimónides Institute for Biomedical Research (IMIBIC), University of Córdoba, Reina Sofía University Hospital, Córdoba, Spain*

[d]*Postharvest Technology and Agrifood Industry, IFAPA Center, Alameda del Obispo, Córdoba, Spain*

[e]*INIA Center I+D of Cerdo Ibérico, Ctra Ex101 Km, Zafra, Badajoz, Spain*

Determination of fatty acids and stable carbon isotopic ratio in subcutaneous fat to identify the feeding regime of Iberian pigs

María Asunción López-Bascón, Feliciano Priego-Capote, Mónica Calderón-Santiago, Verónica Sánchez de Medina, José Manuel Moreno-Rojas, Juan María García-Casco, María Dolores Luque de Castro

ABSTRACT

Discrimination among the types of feeding regimes for Iberian pigs is currently a highly demanded challenge by the Iberian pig sector. In the present research, discrimination among feeding regimes has been achieved by combination of two analytical methods (based on FAMEs analysis by GC–FID and determination of $\delta^{13}C$ by IRMS) previously used independently without success. In the present study, 80 samples of adipose tissue from Iberian pigs subjected to four different feedings were analyzed. The study of the variables more influenced by the feeding regime has allowed us to configure panels of markers with predictive power for the studied feedings by multivariate ROC analysis. The results provided values of specificity and sensitivity higher than 85% in most cases. The statistical combination of results from different analytical methods could be the key to develop models for the correct discrimination of Iberian pigs according to the feeding regime.

Keywords: Iberian pig, subcutaneous fat, IRMS, GC–FID, feeding regimes, ROC analysis, fatty acid methyl esters, stable carbon isotopes.

1. Introduction

The Iberian pig is well adapted to the environmental and ecological conditions of dehesa, where it eats basically acorns, the fruit of the genus *Quercus*. Studies have shown that this feeding habit yields Iberian pigs with large amounts of monounsaturated fatty acids (MFAs), triglycerides and phospholipids;[1,2] therefore, the fat composition in these animals is characterized by a content of polyunsaturated fatty acids (PUFAs) lower than in those fed with prepared feeds.[3]

In the last 30 years the increased demand for Iberian pork products has led to a rise in production, which has not been in-parallel with the available natural resources. The dehesa forest cannot support the density of animals demanded by consumers; therefore, Iberian pigs are most times fed with grains and prepared feeds[4] and the obtained products are endowed with nutritional and organoleptical properties different from those of pigs subjected to *montanera* feeding (traditional name for the feeding regime exclusively based on the natural resources of the dehesa). Differences in the feeding regime have a direct impact on the economic value of cured products obtained from slaughtered pigs. The high economic value of cured products from animals fed under *montanera* regime has made mandatory regulation and identification of Iberian products. Therefore, a legislation (a Royal Decree, RD) to regulate their quality according to the animals feeding was promoted and approved in 2001.[5] Designation of the different types of feeding has undergone changes along time, which are defined according to the Spanish RD[6] as follows:

1. *Bellota* (*montanera*): feeding based on acorns, grass and other natural resources present in the dehesa without any contribution of prepared feeds (mainly composed by cereals and legumes).

2. *Recebo*: feeding up to a minimum weight in *montanera* conditions, then supplemented with prepared feeds.

3. *Cebo de Campo*: feeding based on prepared feeds, but with a final stay in the dehesa for at least 60 days combined with prepared feeds.

4. *Cebo*: feeding based exclusively on prepared feeds.

Currently, some modifications in the new RD of 2014,[7] have included the *Recebo* class into the *Cebo de Campo* class. Thus, the new *Cebo de Campo* class is a type of feeding based on prepared feeds but also, in some cases, on acorns and grasses of the dehesa.

Among the methods used for classification of products according to the feeding regime of Iberian pigs, the most common approach has been analysis of the relative proportions of FAs in adipose tissues by gas chromatography (GC) separation and flame ionization detection (FID).[3,4] Analysis of FAs is of great interest in the meat industry because the technological, nutritional and sensory aspects resulting from their composition. Nevertheless, the determination of FAs in the subcutaneous tissue to be related with feeding regime of pigs has given rise to an increase of prepared feeds defined by a composition of FAs similar to that in acorns.[8] Therefore, the composition of the most important fatty acids in animals fed with prepared feeds is very similar to that in those fed by *montanera* regime and the power of the analytical method to distinguish among the categories decreases as a result.[4]

The feasibility of using near-infrared spectroscopy (NIRS) for discrimination among subcutaneous fat from Iberian pigs reared on different fattening diets has also been evaluated in diverse studies.[9,10] This technology is based on the analysis of the spectra in the near infrared region provided by direct determination of cured products or fresh meat. Nevertheless, the models developed so far do not allow successful discrimination of all types of diets. It is clear that new and more specific analytical methods are needed at present to discriminate among feeding regimes. The application of isotope ratio mass spectrometry (IRMS) has become an interesting tool to analyze the quality and authenticity of food.[8] Some of the methods based on this technique have been adopted as official in the EU —for example, for characterization of wine[11] and fruit juices.[12] Isotopic analysis of carbon and sulfur in adipose tissue has also led to an alternative method to discriminate between *Bellota* and *Cebo* regimes.[13,14] Pioneers in the study of the influence of diet on stable carbon isotope ratio in animals were De Niro *et al.*[15] Their results showed the possibility to perform dietary analysis based on the

determination of the $^{13}C/^{12}C$ ratio. In the field of our research, González-Martín *et al.*[13] were pioneers in the determination of the isotopic $^{13}C/^{12}C$ ratio in subcutaneous fat of Iberian pigs to differentiate feeding regimes. Subsequently, other authors have studied the applicability of this technique in other fresh pig tissues —mainly muscle,[13] liver,[4,13–16] adipose tissue [4,11,15] and intramuscular fat[3,14,18,19] for prediction of feeding regimes in Iberian pigs[4]. In general, the different techniques presented difficulty in discerning between other feedings and *Recebo*. The authors emphasized that application of only one method is not enough for discrimination of four feeding categories; therefore, they combined several methods. Nevertheless, they did not combine their data to develop a unique discrimination model.

The aim of the research here presented was to combine the classical method for determination of FAs based GC–FID and that for monitoring carbon isotopic abundances by IRMS to develop discrimination models and search for panels of markers with capability to classify slaughtered animals by their feeding regime.

2. Materials and methods

2.1. Samples

Eighty Iberian pigs were selected in different farms located at the North of the province of Sevilla (Spain) within an area delimited by 25 km from South to North and 50 km from East to West. Collection of samples was carried out in 2010 by personnel belonging to the National Institute of Agriculture and Food Research and Technology (INIA, Extremadura, Spain). The animals were classified into four groups according to the feeding regime as the Official Legislation[6] set. Each group was composed of 20 samples. Subcutaneous fat samples from slaughtered pigs were collected following the established Spanish official method.[20] The fat tissues were cut and put into individual microwave-resistant glass containers, then heated in a domestic microwave oven (700 W power and 2450 MHz microwave frequency) for 5 min at the maximum power, after which the melted samples were homogenized prior to sampling 5 mL of the liquid fat.[21] The treated samples were then stored at −20 °C until analysis.

2.2. Reagents and standards

n-Hexane from Scharlau (Barcelona, Spain) was used for fat solubilization. Commercial standards of FA methyl esters (FAMEs) were acquired from Fluka Analytical (Buches, Switzerland), and corresponded to the following acids: lauric (C12:0), myristic (C14:0), palmitic (C16:0), palmitoleic (C16:1n9), hepatadecanoic (C17:0), stearic (C18:0), oleic (C18:1n9), eicosanoic (C20:0), eicosenoic (C20:1), eicosadienoic (C20:2n6), eicosatrienoic (C20:3n6), eicosatetraenoic (C20:4n3), arachidonic (C20:4n6), eicosapentaenoic (20:5n3), behenic (C22:0), tetracosenoic (C24:0), docosatetraenoic (C22:4n6), docosapentanoic (C22:5n3), docosahexaenoic (C22:6n3), hexacosanoic (C26:0) and octacosanoic (C28:0). A mixture of *cis*/*trans* isomers of linoleic (C18:2) and linolenic (C18:3) FAMEs was from Supelco Analytical Sigma–Aldrich (St. Louis, MO, USA). Tridecanoic acid methyl ester (C13:0) was used as external standard (ES) in the determination step. It is not present in the samples and its physico-chemical properties are similar to those of the target FAs. Individual stock standard solutions and multistandard solutions at different concentrations were prepared by dilution in *n*-hexane. The concentrations were selected according to those of the FAs in subcutaneous fat in pork, and the solutions were stored at –20 ºC. The international isotopic standards (IIS) used for determination of $\delta^{13}C$ values by IRMS were mineral oil (IAEA NBS22) and sacarose (IAEA CH_6), provided by the International Atomic Energy Agency (IAEA, Vienna, Austria).

2.3. Instruments

An Agilent 7820A GC System equipped with an FID was used to separate and determine the esterified FAs in each extract. The GC was furnished with an autosampler and a split/splitless injector. An SP™-2380 fused silica capillary column (60 m × 0.25 mm I.D., 0.2 μm film thickness) provided by Supelco (Bellefonte, PA, USA) was used as analytical column. EZ Chrom Elite Compact software (version 3.3.2 from Agilent Technologies, Santa Clara, CA, USA) was used for acquisition and processing of data in the GC–FID system. A Delta V Advantage Isotope Ratio Mass Spectrometer from Thermo Fisher Scientific (Bremen, Germany) was also used. This system was equipped with a ConFlo IV Universal Interface for continuous flow analysis and a Flash 2000 HT

elemental analyzer. A molecular sieve packed column (5 Å, 1 m × 1/8" × 2 mm) from Thermo Scientific (Bremen, Germany) was used for sample clean-up. Isodat Gas Isotope Ratio MS Software (version 3.0 from Thermo Scientific, Bremen, Germany) was used to acquire and process the signal obtained by IRMS analysis.

Statgraphics Centurion XV Version 15.1.02 for Windows and Mass Profiler Professional (MPP) software package (version 2.0 from Agilent Technologies, Santa Clara, CA, USA) were used for statistical analysis of the generated data. Roccet software (Version on-line, URL http://www.roccet.ca/) was used for data pretreatment and analysis of Receiver Operating Characteristic (ROC) curves.

2.4. Preparation of FAMEs

Conversion of FAs into their more volatile FAMEs is mandatory prior to individual GC separation. 0.1 g of subcutaneous fat was shaken with 2 mL of hexane and vortexed for 1 min to complete solubilization. 0.2 mL of 2 M methanolic potassium hydroxide solution was added and the byphasic system was vortexed for 1 min. The two phases were easily separated and 1.5 mL of the upper phase —containing the esterified FAs— was pipetted in a glass vial. This solution was 1:100 (v/v) diluted in hexane and the resulting analytical sample was spiked with the ES (40 μg mL^{-1}) prior to injection into the GC.

2.5. Determination of FAMEs by GC–FID

2 μL of each analytical sample was injected into the GC–FID equipment for analysis of FAMEs. The injector and detector were maintained at 250 and 295 ºC, respectively. All samples were prepared and analyzed in triplicate. The injection was 1:10 split and the program of the temperature gradient was as follows: the initial oven temperature was kept at 50 ºC for 2 min, then raised at 15 ºC min^{-1} to a final temperature of 250 ºC, which was held for 15 min. The equilibration time was 3 min. The mobile phase was helium at a flow rate of 1 mL min^{-1}. The corresponding standard was used to confirm identification of each FAME in the fat samples. Quantitative analysis was carried out by preparation of a calibration model for each FA by using the multistandard

solution at different concentrations spiked with the ES at 40 μg mL^{-1}. The concentrations of the target FAs were expressed as percentage in relative terms.

2.6. Determination of $^{13}C/^{12}C$ isotopic ratio by IRMS.

Sacarose ($\delta^{13}C$–10.45‰) and NBS22-Oil ($\delta^{13}C$ –30.03‰) were used as IIS within the ranges 350–450 and 550–750 μg, respectively, to achieve a signal intensity of 7600 mV. Samples (within the range 320–420 μg) were weighed into a tin capsule (3.3 × 5 mm, IVA Analysentechnik e.K., Düsseldorf, Germany). The analyses were carried out in batches of 10 samples, sandwiched by standards as internal control. The carrier and reference gas (He in both cases) were circulated at 80 and 200 mL min^{-1}, respectively. The temperatures of the elemental analyzer were set at 65 and 1020 ºC for the oven temperature and combustion reactor, respectively. The results of the carbon isotope ratio ($\delta^{13}C$) analyses were reported as the per mil (‰) enrichment relative to the international standard, V-PDB (Vienna Pee Dee Belemnite) for carbon isotope ratio, according to the equation:

$$X(‰) = [(R_{sample}/R_{reference}) - 1] \times 1000,$$

where X is the ratio of the heavy to the light stable isotope (*e.g.* $^{13}C/^{12}C$) in the sample (R_{Sample}) and in the standard ($R_{Reference}$)[22] The fat samples were analyzed in duplicate with a variability less than 0.20‰ for $\delta^{13}C$ measurements.

2.7. Statistical analysis

The raw data files were used to create a data set formed by samples (columns) and the quantitative variables (concentrations of FAs and $\delta^{13}C$ values). The data set, formed by 11 variables × 80 samples, was exported as comma separated value files (.csv) into the Roccet and MPP softwares for processing.

Normalization by logarithmic transformation was used as pre-processing step since the data obtained from the two different analyses were combined. An ANOVA test was executed with the normalized data to evaluate the significance of each variable (*p*-value lower than 0.05). Statistical analysis was carried out by Principal Component Analysis (PCA) and Partial Least Squares Discriminant Analysis (PLS-DA). In the latter

case, an N-Fold cross-validation model was used. With this model, the classes of the input data are randomly divided into N equal parts; N-1 parts are used for model training, and the remaining part is used for testing. The process is iteratively repeated N times using each time a different subset for testing. Thus, each row is used at least once in training and once in testing, and a confusion matrix is generated. This whole process can then be repeated as many times as specified by the number of repetitions. Ten repetitions and a fold number of three were selected in all PLS-DA models.

The ROC curves were individually obtained for different panels of variables in order to find discrimination trends between pairs of feedings.

3. Results and discussion

3.1. Development of calibration models for analysis of FAMEs

The calibration models were prepared from 25 FAMEs prior to analysis of the samples by using multistandard solutions at different concentrations as a function of the abundance of each compound in subcutaneous pork fat.[1,3,23] The most concentrated FAs —C16:0, C18:0, and particularly, C18:1n9— ranged from 1.8 to 187.5 μg mL^{-1}; a second set (C14:0, C:16:1n9 and the C18:2 isomers) ranged from 0.90 to 93.75 μg mL^{-1}, and a third calibration range (from 0.48 to 48.75 μg mL^{-1}) was applied to C12:0, C17:0, C18:3 isomers, C20:1n9, C20:2, C20:3 and C20:4. Finally, the rest of FAs were modelled from 0.125 to 12.5 μg mL^{-1}. Six dilutions (1:2, 1:5, 1:10, 1:25, 1:50, 1:100 and 1:250) of the multistandard stock solution, each injected in triplicate, were used to run the calibration curves. All the solutions at the different concentrations were spiked with the ES at 40 μg mL^{-1}.

The calibration curves were obtained by plotting the ratio between the peak area for each FA and the peak area of the ES. The calibration equations and regression coefficients are listed in Supplementary Table 1. As can be seen, the regression coefficients were above 0.99 for all the compounds. It is also worth mentioning that the eight C18:3 isomers could not be independently quantified, except for the *t,t,t* and *c,c,c*

isomers. A similar situation was found for C20:3 and C20:4 owing to overlapping of their chromatographic peaks.

3.2. Analysis of FAs in the target samples by GC–FID and IRMS characterization

Taking into account the variable concentration of the FAs in subcutaneous pig fat, different dilution factors were tested for analysis of the derivatized samples to obtain quantitative information of the maximum number of FAs. Thus, an 1/100 (v/v) dilution factor was adopted for analysis of the 10 most concentrated FAs (C14:0, C16:0, C16:1n9, C17:0, C18:0, C18:1n9, c, c C18:2, c,c,c C18:3, C20:0 and C20:2), while the rest of FAs were detected at lower dilution factors at which quantitative analysis of the most concentrated FAs was compromised; therefore, they were discarded from the results obtained at dilutions lower than 1/100. Supplementary Figure 1 shows a GC–FID chromatogram obtained by analysis of subcutaneous fat from a *Bellota* sample.

After analysis, the data were processed to obtain the FAs concentration profile from each sample. Table 1 summarizes the results provided by the 80 samples with information of the maximum and minimum concentrations found for each FA and the mean concentration and standard deviation for each target feeding. The $\delta^{13}C$ value is also included to show the variability of this parameter. As Table 1 shows, four FAs (C18:1n9, C16:1n9, C18:0 and c,c C18:2) clearly dominated the concentration profile of subcutaneous fat, which is in agreement with characterization studies previously reported.[1,3] The rest of FAs were present at relative concentrations below 5%.

The comparison of the FAs profile among the four types of samples grouped by feeding revealed that the content of major FAs was similar for *Cebo* and *Cebo de Campo* feedings (*p*-value above 0.0713), while samples from *Recebo* and *Bellota* also provided a similar profile (*p*-value above 0.1239); a quite foreseeable behavior taking into account that these two feedings are the most similar among the four classes under study. It is also worth emphasizing that C18:1n9 was the most concentrated FA in the four groups of samples (about 50% of relative concentration), but mean values were higher for *Recebo* and *Bellota* (56.03 and 55.44%, respectively) that could be ascribed to similarity between these two feeding regimes. Animals subjected to *Recebo* regime are fattened in

montanera conditions up to they reach a cut-off weight, while *Bellota* animals are fed under *montanera* conditions up to they are slaughtered.

Table 1. Maximum, minimum and mean concentration values expressed as percentage of FAs and $\delta^{13}C$ as per thousand as a function of the feeding regime.

Type of sample		C14:0	C16:0	C16:1n9	C17:0	C18:0	C18:1 n9	c,c,C18:2	c,c,cC18:3	C20:0	C20:2	$\delta^{13}C$
Bellota	*Max*	1.38	21.42	2.48	0.40	9.76	60.73	10.05	1.03	1.42	0.59	-28.01
	Min	1.03	17.07	1.44	0.27	7.22	51.29	8.10	0.40	0.71	0.40	-24.77
	Mean	1.18	19.62	1.84	0.33	8.35	55.44	9.00	0.65	0.93	0.49	-27.05
	N	20	20	20	20	20	20	20	20	20	20	20
	*RSD (%)**	9.78	5.31	13.60	10.81	9.72	4.86	8.65	27.84	20.24	13.45	2.59
Recebo	*Max*	1.54	22.22	2.51	0.45	10.79	59.92	10.07	0.83	1.49	0.63	-27.83
	Min	1.00	17.75	1.69	0.29	7.40	49.07	7.12	0.34	0.68	0.04	-25.48
	Mean	1.25	20.12	1.94	0.34	8.82	56.03	8.59	0.54	1.01	0.41	-26.26
	N	20	20	20	20	20	20	20	20	20	20	20
	*RSD (%)**	10.51	6.10	9.98	12.40	11.96	5.82	10.00	24.57	23.37	36.70	3.02
Cebo de Campo	*Max*	1.77	27.24	3.49	0.54	13.60	53.80	11.11	0.80	1.53	0.75	-25.38
	Min	1.13	17.68	1.67	0.27	7.15	37.85	7.15	0.41	0.95	0.31	-24.54
	Mean	1.43	23.57	2.50	0.36	11.16	44.73	8.98	0.52	1.27	0.59	-24.93
	N	20	20	20	20	20	20	20	20	20	20	20
	*RSD (%)**	9.67	9.59	19.18	17.06	15.15	8.85	13.02	17.08	11.65	18.26	1.08
Cebo	*Max*	1.74	26.80	3.06	0.89	14.45	52.90	9.71	0.54	1.01	0.55	-27.80
	Min	1.12	21.32	1.66	0.18	9.08	41.50	6.17	0.29	0.53	0.22	-25.55
	Mean	1.39	23.99	2.17	0.39	11.28	46.92	8.35	0.42	0.72	0.39	-26.40
	N	20	20	20	20	20	20	20	20	20	20	20
	*RSD (%)**	13.91	9.40	18.48	56.45	15.03	7.98	14.69	16.86	20.02	24.92	2.33

*RSD: Relative standard deviation.

The only official method for discrimination of feeding regimes of Iberian pigs was based on GC–FID analysis of FAs composition in subcutaneous adipose tissue.[6,18] This method was discarded as official method in 2007 as it was not able to detect frauds imvolving Iberigan pigs fed with prepared feeds.[18] B.O.E. 2004[20] includes a summary of the criteria used for classification of the three defined feedings: *Cebo*, *Recebo* and *Bellota*. The relative standard deviation (RSD %) was quite similar among the four groups under study (Table 1): below 10% for all FAs, except for the two most relevant: C16:0 and C18:1.

The $\delta^{13}C$ values (see Supplementary Figure 1 that shows an IRMS spectrum provided by analysis of a subcutaneous adipose *Bellota* sample) ranged from –28.01 to

−24.54‰, with RSD below 3‰. Attending to this variability, the inclusion of $\delta^{13}C$ in classical discrimination models based on FAMEs analysis could aid to feedings classification. In this research, only carbon isotopic ratios were studied. The isotopic analysis of hydrogen and oxygen has been applied mainly for geographical discrimination of different types of food such as durum wheat semolina and potatoes.[4,16,25-30] Geographical discrimination is out of the scope of the present research as all samples were collected in a delimited area, as specified under materials and methods.

3.3. Influence of feeding on the FAs profile and $\delta^{13}C$ value

The data set was normalized by logarithmic transformation because of the two analytical methods used. The effect of this pre-processing step was proved by application of the Kurtosis and Skewness tests. Then, the normalized data set was analyzed by ANOVA to evaluate the influence of the feeding type on both the concentration of FAs and the value of $^{13}C/^{12}C$ isotopic ratio (dependent variables).

Table 2 summarizes the results provided by ANOVA of the 11 dependent variables as a function of the predictive factor, the feeding type. The variance is decomposed into two components: between-groups variability (inter-groups) and within-group variability (intra-group). As can be seen, most of the variables reported a very high significance with *p*-values below 0.001, except for C17:0 and *c,c* C18:2 that did not show any level of significance. Therefore, the feeding factor exerted a noticeable influence on the FAs profile and $^{13}C/^{12}C$ ratio. In fact, feeding was identified as the main variability source to explain the levels of the most concentrated FAs (C16:0, C18:0, C18:1n9 and C20:0) as well as the $^{13}C/^{12}C$ ratio, which was deduced by comparing the sum of squares between and within the target groups.

Figure 1 shows the Box-and-Whisker plots obtained by normalized concentrations of the significant variables, which allowed detecting some differences among the four groups. Thus, the $\delta^{13}C$ values revealed differences between *Cebo de Campo* samples and the other three feedings. On the other hand, C16:0, C16:1n9 and C18:0 were characterized by a similar trend since they were clearly more concentrated in *Cebo de Campo* and *Cebo* samples than in those from *Bellota* and *Recebo*. The opposite situation

was observed with C18:1n9, which was more concentrated in animals subjected to *montanera* regime during a given period. A different trend was observed for C20:0, which provided discrimination in concentration between *Cebo de Campo* and *Cebo* samples, while *Bellota* and *Recebo* samples gave levels between the two previous groups.

Table 2. Sum of squares between and within groups, mean square and significance (*p*-values) obtained after application of the ANOVA test.

Compound	Variability	Sum of squares	D.f.	Mean square	*p-V*alue
C14:0	Between-groups	0.95	3	0.32	0.000
	Within-group	1.93	76	0.03	
	Total	2.88	79		
C16:0	Between-groups	1.32	3	0.44	0.000
	Within-groups	1.03	76	0.01	
	Total	2.35	79		
C16:1n9	Between-groups	2.15	3	0.72	0.000
	Within-groups	3.66	76	0.05	
	Total	5.82	79		
C17:0	Between-groups	0.13	3	0.04	0.826
	Within-groups	10.71	76	0.14	
	Total	10.83	79		
C18:0	Between-groups	2.92	3	0.97	0.000
	Within-groups	2.83	76	0.04	
	Total	5.75	79		
C18:1n9	Between-groups	1.68	3	0.56	0.000
	Within-groups	0.78	76	0.01	
	Total	2.46	79		
c, c, C18:2	Between-groups	0.18	3	0.06	0.121
	Within-groups	2.31	76	0.03	
	Total	2.50	79		
c, c, c C18:3	Between-groups	3.52	3	1.17	0.000
	Withn-groups	7.19	76	0.09	
	Total	10.71	79		
C20:0	Between-groups	6.96	3	2.32	0.0000
	Within-groups	5.35	76	0.07	
	Total	12.32	79		
C20:2	Between-groups	5.06	3	1.69	0.000
	Within-groups	14.19	76	0.19	
	Total	19.25	79		
δ13C	Between-groups	0.15	3	0.05	0.000
	Within-groups	0.09	76	0.00	
	Total	0.24	79		

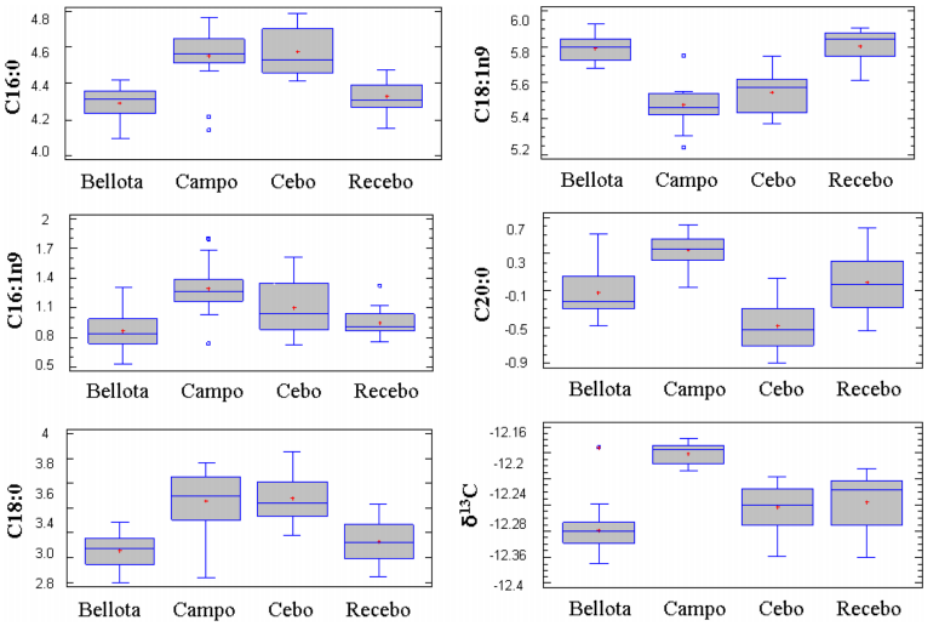


Figure 1. Box-and-whisker plots provided by comparison of normalized concentrations of significant FAs (%) and $\delta^{13}C$ (‰) according to the feeding regime.

The influence of feeding on the FAs profile and $\delta^{13}C$ values led to apply multivariate analysis to find discrimination trends among groups. Thus, PCA was first applied to check the incidence of the breeding type on the FAs profile and the $^{13}C/^{12}C$ ratio. The three-dimensional PCA scores plot with samples identified by symbols as a function of the type of feeding is shown in Supplementary Figure 2. The 3D-plot allowed explaining a 70.6% of the total variability. One cluster formed by *Cebo de Campo* samples can be clearly differentiated from the other three groups. Thus, the *Cebo de Campo* class was the feeding type providing the most discriminating pattern in the FAs profile and the $^{13}C/^{12}C$ ratio. On the other hand, a partial separation can be observed for the other three feeding classes (*Cebo*, *Recebo* and *Bellota*), which allows setting differences. A particular behavior was visualized for *Cebo* samples as two sub-clusters were found, one of them completely separated, and a second sub-cluster that overlapped with *Recebo* and *Bellota* samples. No details were provided by farmers about this result, but it could be associated with differences in the provided feeding. The results obtained with the

PCA seems to be quite logical since the maximum overlapping was detected between *Recebo* and *Bellota* samples that constitute the most similar classes since animals subjected to *Recebo* feeding must reach a minimum weight in the dehesa (*montanera* regime).

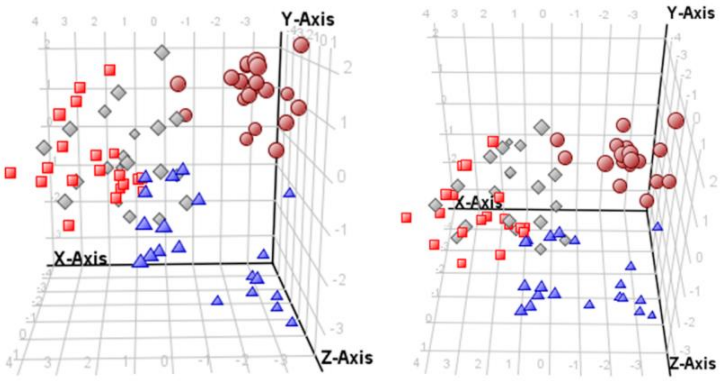


Figure 2.PLS-DA scores plot (different views) of samples from slaughtered pigs subjected to different feeding regimes.X-axis, t0; Y-axis, t1; Z-axis, t2. Explained variability: X-axis, 40.95%; Y-axis, 19.56%; Z-axis, 9.65%. Squares, *Bellota*; circles, *Cebo de Campo*; triangles, *Cebo*; rhombuses, *Recebo*.

The next step was application of supervised statistical analysis by PLS-DA to evaluate the capability of the data set for development of predictive/discrimination models. Figure 2 shows the PLS scores plot that allows discrimination between *Cebo de Campo* and *Cebo* samples, while *Bellota* and *Recebo* samples overlapped. This behavior was checked in the confusion matrix provided by this PLS-DA model both in the training and validation steps, which were carried out by cross-validation. Thus, *Bellota, Cebo de Campo* and *Cebo* were classified with accuracy above 90%, while the model failed in the classification of *Recebo* samples with a 45% of accuracy. In fact, the model confused 35% of the *Recebo* samples, which were erroneously classified as *Bellota* samples.

3.4. Multivariate analysis between pairs of feedings

Since the variability in the concentrations of FAs and $\delta^{13}C$ values does not allow complete discrimination of the four groups of feedings, multivariate analysis was

applied to pairs of feedings to precise the patterns that differentiate them. For this purpose, one PLS-DA model was created for each pair of feedings, and then, ANOVA analysis revealed the main statistical differences in subcutaneous fat of animals subjected to each feeding (Supplementary Figure 3).

Supplementary Table 2 shows the discrimination capability for training and cross-validation obtained from the six cases. As expected, high-discriminating models were found in all pairs of feedings, except for that involving *Bellota* and *Recebo* samples, which provided discrimination capabilities above 75% in the training step and 65% in the validation step. Supplementary Figure 3 shows the 3D-plots obtained by PLS-DA with clear separation between each pair of feedings. Therefore, if the variability study is focused on two types of feedings the discrimination models gain in prediction capability, even for *Bellota* and *Recebo* samples, which seems to be the two most similar feedings.

After supervised analysis, an ANOVA test was independently applied to each pair of feedings to evaluate its influence on the concentration of FAs and $\delta^{13}C$. Table 3 lists the variables that are significantly different by comparison of two feedings and the *p*-values obtained for them. As can be seen, the $\delta^{13}C$ value was significantly altered by the feeding in all cases, with *p*-value below 0.001, except for the pair *Recebo vs. Cebo*. The most concentrated FAs in subcutaneous fat, C16:0, C18:0 and C18:1n9, were significant in most of the evaluated pairs, except for *Cebo de Campo vs. Cebo* and *Recebo vs. Bellota*. It is worth mentioning that the significance level was quite high in all cases with *p*-values below 0.001. Only two FAs, C17:0 and *c,c* C18:2, were not significantly affected by the type of feeding.

3.5. Prediction analysis by multivariate ROC analysis

The influence of feeding on the variability of FAs concentration and $\delta^{13}C$ values in subcutaneous fat from Iberian pigs supports the assessment of predictive models based on panels formed by combination of significant variables according to the ANOVA test. ROC curves are generally considered the method of choice for evaluating the performance of potential biomarkers associated to biological processes. In this research, panels were prepared by multivariate ROC analysis, which involves a preliminary study

of the individual predictive capability of all the variables and, then, the combination of variables in panels to maximize the predictive capability.

Table 3. *p*-Values obtained by application of ANOVA test between pairs of regimes.

***p*-Value**	***Cebo de Campo vs. Recebo***	***Cebo de Campo vs. Bellota***	***Cebo de Campo vs.* Cebo**
0.05-0.01			C16:1n9
0.01-0.001	C20:2	C18:3, C20:2	
<0.001	δ^{13}C, C16:0 C18:0, C18:1n9 C14:0 C20:0, C16:1n9	δ^{13}C, C16:0 C18:0, C18:1n9 C14:0 C20:0, C16:1n9	δ^{13}C, C18:3 C20:0, C20:2

***p*-Value**	***Bellota vs.* Cebo**	***Recebo vs. Bellota***	***Recebo vs.* Cebo**
0.05-0.01	C18:2	C18:3, C20:2	C14:0, C16:1n9
0.01-0.001	δ^{13}C, C16:1n9	δ^{13}C	C18:3
<0.001	C18:0, C18:1n9 C16:0, C18:3 C20:0, C14:0 C20:2		C18:0, C18:1n9 C16:0, C20:0

Taking into account that *Bellota* is the most appreciated class, prediction models were created for classification of *Bellota vs. Cebo de Campo*, *Bellota vs. Cebo* and *Bellota vs. Recebo*. Other panels with commercial interest were also configured for discrimination between *Cebo vs. Cebo de Campo* and *Cebo vs. Recebo*. Predictive models were developed by dividing the set of samples into training and validation sets, which included 75% and 25% of the samples of each condition, respectively. The samples used for validation were not included in the training step, which means that this was carried out with an external set. The algorithm used for classification and selection of variables in the panels was PLS-DA with three latent variables. All the panels were formed by three variables, except for *Bellota vs. Recebo*, in which the best predictive behavior was provided by δ^{13}C, and no improvement provided the inclusion of additional variables. In the studies involving other combinations of feeding regimes, the sensitivity and/or specificity was not improved by inclusion of more variables either. It is also worth mentioning that all panels constituted by three variables were formed by δ^{13}C and the most concentrated FA: oleic acid. Additionally, stearic acid contributed to

predict *Bellota vs. Cebo de Campo* and *Bellota vs. Cebo*, while palmitic acid and eicosanoic acid improved the models for discrimination of *Cebo vs. Recebo* and *Cebo vs. Cebo de Campo*, respectively. The models involving *Bellota vs. Cebo de Campo* and *Cebo* reported high sensitivity and specificity, always above 90%, which lead to quite good predictive models. The most complicated case was the model involving *Bellota vs. Recebo*, based on a unique variable, $\delta^{13}C$, which reported a specificity of 80%, while sensitivity decreased to 63%. This result agrees with that provided by the PLS-DA study for the same two feeding regimes. The efficiency of this model can be compared with that provided by previous studies based on IRMS analysis, such as that carried out by García Casco *et al.*,[4] who provided a model based on the analysis of stable isotopes of carbon ($\delta^{13}C$) and sulphur ($\delta^{34}S$) with lower discrimination capability than the model obtained in this research. The sensitivity and specificity of the models for discrimination of *Cebo vs. Recebo* and *Cebo vs. Cebo de Campo* were 93% and 85%, respectively. The most compromised case was the model involving *Recebo vs. Bellota* that reported a specificity of 80% and sensitivity of 63%. Thus, these models enabled to classify samples according to their feeding regimes. These good prediction parameters can be explained by the ROC curves illustrated in Figure 3.

The results of sensitivity and specificity for each panel can be combined with the estimations of prediction capability for each class and the false positive rates listed in Supplementary Table 3. Thus, these values associated with the validation set were quite consistent with the results provided by the training set. As the validation set was formed by samples not included in the training set, the quality and statistical support is considerably enhanced.

At present, there are not discrimination methods able to identify the feeding regime used with capability to detect frauds. Different techniques have been independently assayed. However, no methods based on the application of different analytical techniques had been combined so far. In view of these results, the statistical combination of the results obtained by different analytical methods could be the key for the correct discrimination of feeding regimes in the sector of Iberian pig. Taking into account the demand for methods to discriminate feeding regimes, this approach could

be implemented in routine laboratories since the analysis of fatty acids is already carried out at this level, while IRMS analysis is frequently used in reference food laboratories.

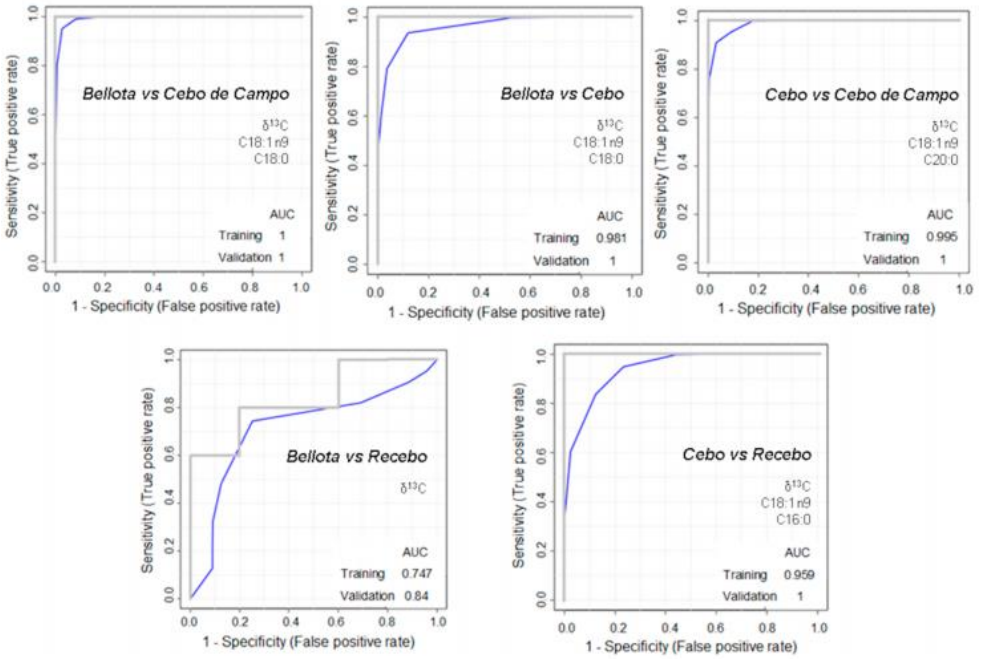


Figure 3. Individual ROC curves obtained from panels of biomarkers for discrimination between pairs of diets. The thick line corresponds to validation (external) and the fine line corresponds to training. *Bellota vs. Cebo de Campo* variables: δ13C, C18:1n9, C18:0; *Bellota vs. Cebo* variables: δ13C, C18:1n9, C18:0; *Cebo vs. Cebo de Campo* variables: δ13C, C18:1n9, C20:0; *Bellota vs. Recebo* variables: δ13C; *Cebo vs. Recebo* variables: δ13C, C18:1n9, C16:0.

Acknowledgments

The Spanish Ministerio de Economía y Competitividad (MINECO) and European FEDER program are thanked for financial support through project CTQ2012-37428. F.P.C. is also grateful to the Ministerio de Ciencia e Innovación (MICINN) for a Ramón y Cajal contract (RYC-2009-03921).

References

[1] Rey, A. I.; Isabel, B.; Cava, R.; Lopez-Bote, C. J. Dietary acorns provide a source of gamma-tocopherol to pigs raised extensively. *Can. J. Anim. Sci.* 1998, *78*, 441–443.

[2] Tejerina, D.; García-Torres, S.; Cabeza de Vaca, M.; Vázquez, F.M.; Cava, R. Acorns [*Quercus rotundifolia Lam.*] and grass as natural sources of antioxidants and fatty acids in the "montanera" feeding of Iberian pig: Intra- and inter-annual variations. *Food Chem.* 2011, *124*, 997–1004.

[3] Cava, R.; Ruiz, J.; López-Bote, C.; Martín, L.; García, C.; Ventanas, J.; Antequera, T. Influence of finishing diet on fatty acid profiles of intramuscular lipids, triglycerides and phospholipids in muscles of the Iberian pig. *Meat Sci.* 1997, *45*, 263–270.

[4] García Casco, J.M.; Muñoz, M.; González, E. Predictive ability of the feeding system in Iberian pig by means of several analytical methods. *Grasas Aceites* 2013, *64*, 191–200.

[5] Royal Decree, 1083/2001, 5 October. Quality standard for Iberian products made in Spain. B.O.E. 2001, *247*, 37830–37833.

[6] Royal Decree, 1469/2007, 2 November. Quality standard for Iberian products made in Spain. B.O.E. 2007, *264*, 45087–45104.

[7] Royal Decree, 4/2014, 10 January. New quality standard for Iberian products made in Spain. B.O.E. 2014, *10*, 1569–1585.

[8] Vinci, G.; Preti, R.; Tieri, A; Vieri, S. Authenticity and quality of animal origin food investigated by stable-isotope ratio analysis. *J. Sci. Food Agric.* 2012, *93*, 439–448.

[9] Arce, L.; Domínguez-Vidal, A.; Rodríguez-Estévez, V.; López- Vidal, S.; Ayora-Canada, M.J.; Valcárcel, M. Feasibility study on the use of infrared spectroscopy for the direct authentication of Iberian pig fattening diet. *Anal. Chim. Acta* 2009, *636*, 183–189.

[10] Fernández-Cabanás, V.M.; Garrido-Varo, A.; García Olmo, J.; De Pedro, E.; Dardenne, P.Optimisation of the spectral pre-treatments used for Iberian pig fat NIR calibrations. *Chemometrics and Intelligent Laboratory Systems* 2007, *87*, 104–112.

[11] Regulation of the European Commission, Nº 2676/90, 17 September 1990. Community methods of analysis applicable in the wine sector. 1990, 1–192.

[12] Jamin E.; Martin F.; Martin G.G. Determination of the 13C/12C ratio of ethanol derived from fruit juices and maple syrup by isotope ratio mass spectrometry: collaborative study. *J. AOAC Int.* 2004, *87*, 621–31.

[13] González-Martín, I.; González-Pérez, C.; Hernández Menéndez, J.; Marqués-Macías, E.; Sanz Poveda, F. Use of isotope analysis to characterize meat from Iberian breed swine. *Meat Sci.* 1999, *52*, 437–441.

[14] González-Martín, I.; González-Pérez, C.; Hernández-Menéndez, J.; Sánchez-González, C. Differentiation of dietary regimen of Iberian swine by means of isotopic analysis of carbon and sulphur in hepatic tissue. *Meat Sci.* 2001, *58*, 25–30.

[15] Deniro, M. J.; Samuel Epstein, S. Influence of diet on the distribution of carbon isotopes in animals. *Geochim. Cosmochim. Acta* 1978, *42*, 495–506.

[16] Sanabria, C.; Martín-Mateos, M. J.; González-Cantillo, N.; Moreno-Indias, I.; García-Casco, J. M. Fatty acid analysis of subcutaneous fat from animals with a reliable and safe feeding. *Grasas Aceites* 2013, *64*, 148–156.

[17] Nürnberg, K.; Wegner, J.; Ender, K. Factors influencing fat composition in muscle and adipose tissue of farm animals. *Livestock Production Science* 1998, *56*, 145–156.

[18] Pascual, J. V.; Rafecas, M.; Canela, M. A.; Boatella, J. R.; Barroeta, A. C.; Codony, R. Effect of increasing amounts of a linoleic-rich dietary fat on the fat composition of four pig breeds. Part II: Fatty acid composition in muscle and fat tissues. *Food Chem*. 2007, *100*, 1639–1648.

[19] Narváez-Rivas, M.; Vicario, I. M.; Alcalde, M. J.; Camacho, M. L. Volatile hydrocarbon profile of Iberian dry-cured hams. A possible tool for authentication of hams according to the fattening diet. *Talanta* 2010, *81*, 1224–1228.

[20] Orden PRE/3844/2004, 18 November. Official methods of sampling channels Iberian pigs and the method of analysis for determining the fatty acid composition of total lipids from subcutaneous fat of Iberian pigs. B.O.E. 2004, *283*, 38770–38779.

[21] De Pedro, E.; Casillas, M.; Miranda, C. M. Microwave oven application in the extraction of fat from the subcutaneous tissue of Iberian pig ham. *Meat Sci.* 1997, *45*, 45–51.

[22] Moreno-Rojas, J.M.; Vasta, V.; Lanza, A.; Luciano, G.; Ladroue1, V.; Claude Guillou, C.; Priolo, A. Stable isotopes to discriminate lambs fed herbage or concentrate both obtained from C3 plants. *Rapid Commun. Mass Spectrom*. 2008, *22*, 3701–3705.

[23] García-Olmo. J.; De Pedro, E.; Garrido, A.; Paredes, A.; Sanabria, C.; Santolallad, M.; Salas, J.; García-Hierro, J.R.; Gonzalez, I.; García-Cachan, M.D.; Guirao, J. Determination of the precision of the fatty acid analysis of Iberian pig fat by gas chromatography. Results of a mini collaborative study. *Meat Sci.* 2002, *60*, 103–109.

[24] Sacco, D.; Brescia, M.A.; Buccolieri, A.; Caputi Jambrenghi, A. Geographical origin and breed discrimination of Apulian lamb meat samples by means of analytical and spectroscopic determinations. *Meat Sci.* 2005, *71*, 542–548.

[25] Camin F.; Bontempo, L.; Heinrich, K.; Horacek, M.; Kelly, S. D.; Schlicht, C.; Thomas, F.; Monahan, F. J.; Hoogewerff, J.; Rossmann, A. Multi-element (H,C,N,S] stable isotope characteristics of lamb meat from different European regions. *Anal. Bioanal. Chem*. 2007, *389*, 309-320.

[26] Schmidt, O.; Quilter, J. M.; Bahar, B.; Moloney, A. P.; Scrimgeour, C. M.; Begley, I. S.; Monahan, F. J. Inferring the origin and dietary history of beef from C, N and S stable isotope ratio analysis. *Food Chem*. 2005, *91*, 545–549.

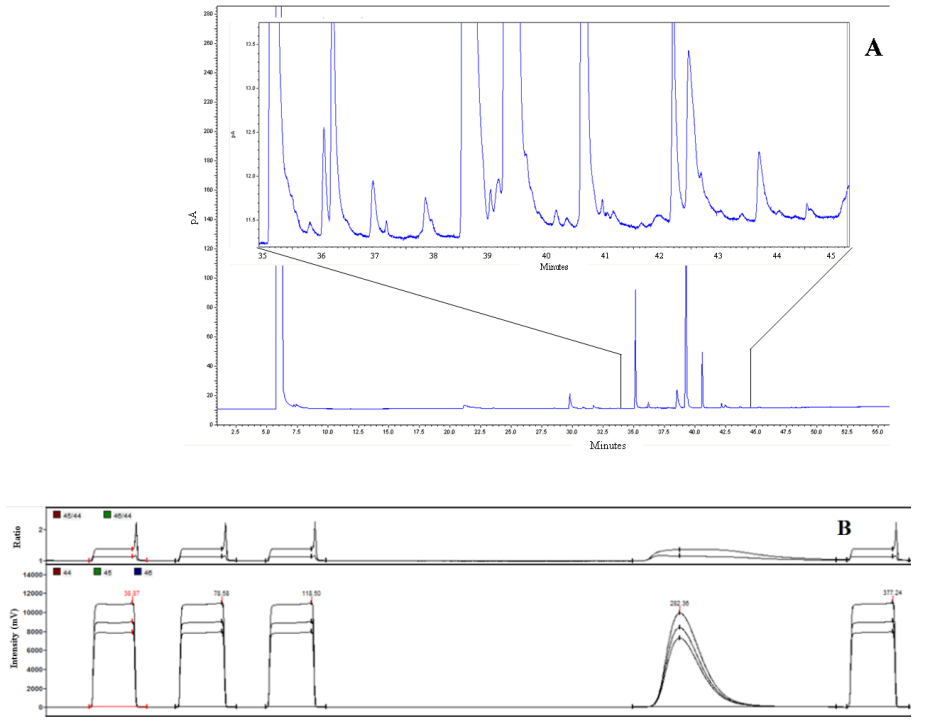
[27] Osorio, M. T.; Moloney, A. P.; Schmidt, O.; Monahan, F. J. Multielement isotope analysis of bovine muscle for determination of international geographical origin of meat. *J. Agric. Food Chem*. 2011, *59*, 3285–3294.

[28] Horacek, M.; Min, J. S. Discrimination of Korean beef from beef of other origin by stable isotope measurements. *Food Chem.* 2010, *121*, 517–520.

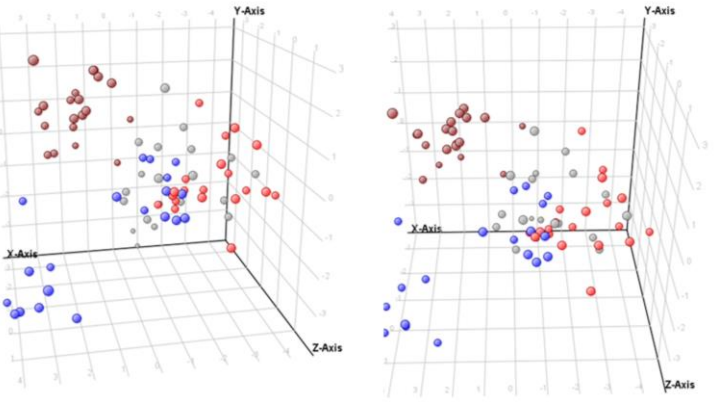
[29] Horacek, M., Min, J. S., Heo, S. C.; Soja, G. Discrimination between ginseng from Korea and China by light stable isotope analysis. *Anal. Chim. Acta* 2010, *682*, 77–81.

[30] Longobardi, F.; Casiello, G.; Sacco, D.; Tedone, L.; Sacco, A. Characterisation of the geographical origin of Italian potatoes, based on stable isotope and volatile compound analyses. *Food Chem.* 2010, *124*, 1708–1713.

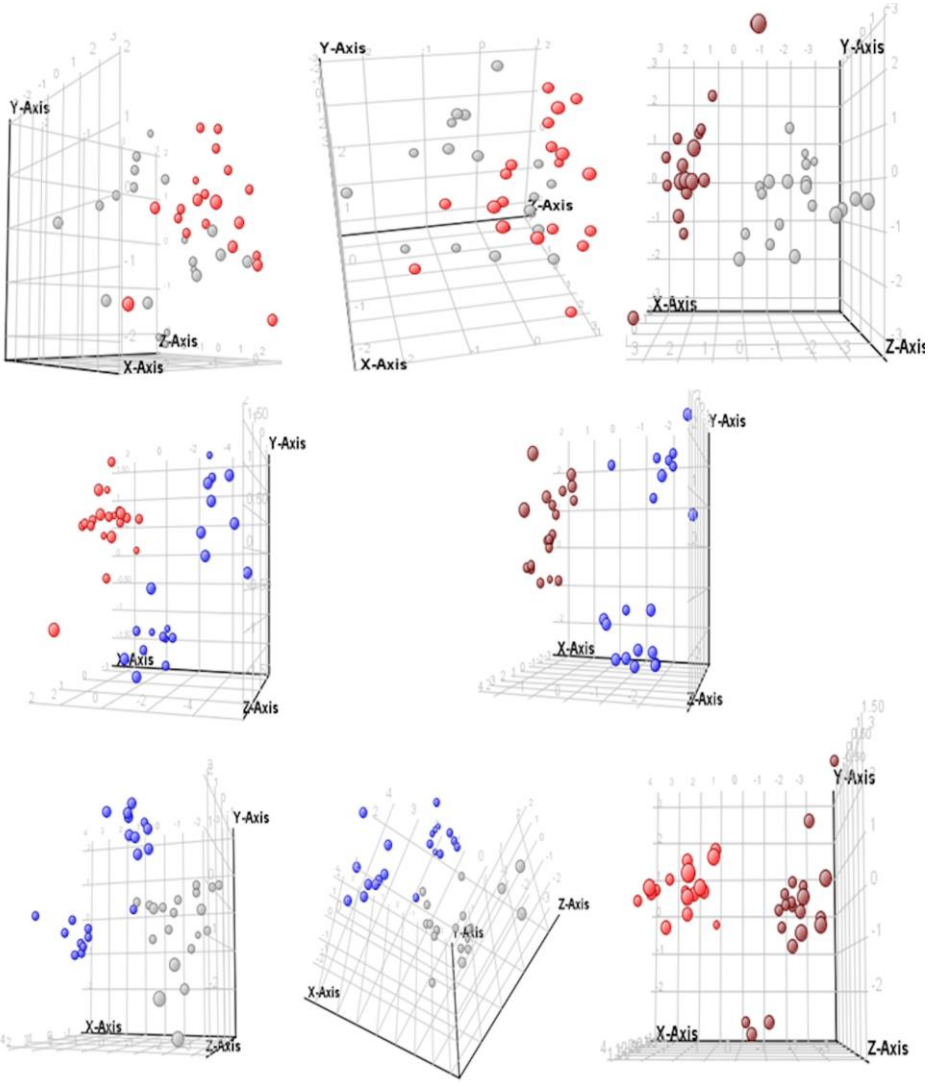
Supplementary information



Supplementary Figure 1. GC–FID chromatogram (A) and IRMS spectrum (B) both from a sample of *Bellota* subcutaneous fat.



Supplementary Figure 2. PCA score plots (different views) associated to FAs and δ13C as a function of the feeding regime.



Supplementary Figure 3. 3D-plots obtained by PLS-DA analysis of FAs and $\delta^{13}C$ to discriminate pair of diets.

Supplementary Table 1. Calibration models for quantitative analyses of FAMEs by GC–FID.

Compound Name	RT (min)*	Calibration Range (μg mL-1)	Calibration equation**	Regression coefficient (R^2)
Methyllaureate,C12:0	27.6	48.75–0.4875	Y=0.0864·X–0.0171	0.9994
Methyltridecanoate,C13:0	29.6	External standard		
Methylmyristate,C14:0	31.5	0.9375–93.75	Y=0.4403·X-0.1007	0.9964
Methylpalmitate,C16:0	35.2	1.875–187.5	Y=0.0209·X+0.0191	0.9999
Methylpalmitoleate,C16:1n9	36.2	0.9375–93.75	Y=0.0041·X+0.0081	0.9992
Methylheptadecanoate,C17:0	36.8	0.4875–48.75	Y=0.0131·X+0.0061	0.9997
Methylstearate,C18:0	39.1	1.875–187.5	Y=0.0258·X-0.0806	0.9998
Methyloleate,C18:1n9	39.3	1.875–187.5	Y=0.0544·X+0.1724	0.9998
t9,t12Methyllinoleate,C18:2	40.1	0.9375–93.75	Y=0.0974·X+0.0063	0.9999
c9,t12Methyllinoleate,C18:2	40.4		Y=0.0039·X+0.0028	0.9999
t,9c12Methyllinoleate,C18:2	40.5		Y=0.004·X-0.0009	0.9999
c9,c12Methyllinoleate,C18:2	40.6		Y=0.0261·X+0.0094	0.9990
t9,t12,t15Methyllinolenate,C18:3	41.3	0.4875–48.75	Y=0.0079·X+0.004	0.9990
t9,t12,c15Methyllinolenate,C18:3	41.6		Y=0.0293·X+0.001	0.9984
t9,c12,c15Methyllinolenate,C18:3				
c9,t12,t15Methyllinolenate,C18:3	41.7		Y=0.0034·X+0.0053	0.9973
c9,c12,t15Methyllinolenate,C18:3				
c9,t12,c15Methyllinolenate,C18:3	42.0			
t9,c12,c15Methyllinolenate,C18:3				
c9,c12,c15Methyllinolenate,C18:3	42.2		Y=0.0085·X+0.003	0.9941
Methyleicosanoate, C20:0	42.3	0.125–12.5	Y=0.0736·X+0.0126	0.9997
Methyleicosenoate, C20:1n9	41.6		Y=0.0293·X+0.001	0.9984
Methyleicosadienoate, C20:2	43.6	0.4875–48.75	Y=0.0264·X+0.0076	0.9999
Methylbehenate, C22:0	44.5	0.125–12.5	Y=0.0359·X+0.0037	0.9997
Methyleicosatrienoate, C20:3n6	45.1	0.4875–48.75	Y=0.025·X+0.0061	0.9999
Methyleicosatetraenoate, C20:4n3				
Methylaraquidonate, C20:4n6				
Methyldocosadienoate, C22:2n6	46.3	0.125–12.5	Y=0.0245·X+0.0035	0.9996
Methyleicosapentaenoate, C20:5n3	46.6	0.125–12.5	Y=0.0221·X+0.0058	0.9994
Methyltetracosenoate, C24:0	47.1	0.125–12.5	Y=0.033·X+0.004	0.9995
Methyldocosatetraenoate, C22:5n6	47.9	0.125–12.5	Y=0.0316·X+0.0058	0.9996
Methyldocosapentanoate, C22:5n3	49.4	0.125–12.5	Y=0.041·X+0.0141	0.9984
Methylhexacosanoate, C26:0	49.6	0.125–12.5	Y=0.0505·X+0.0027	0.9990
Methyldocosahexaenoate, C22:6n3	49.9	0.125–12.5	Y=0.016·X+0.0079	0.9947
Methyloctacosanoate, C28:0	51.9	0.125–12.5	Y=0.0308·X+0.0409	0.9938

*RT: retention time; **Y: peak area of the target FAME/peak area of the ES; X: concentration.

Supplementary Table 2. Accuracy values (expressed as %) for the training and validation sets obtained by PLS-DA as a function of feeding regime.

Training	**Predicted Samples**				
	[B]	**[R]**	**[CA]**	**[C]**	**Accuracy**
***Bellota*[B]**	19	1	0	0	95
***Recebo*[R]**	7	9	2	2	45
***CebodeCampo*[CA]**	0	0	20	0	100
***Cebo*[C]**	0	0	1	19	95
Average					83.75

Validation	**Predicted Samples**				
	[B]	**[R]**	**[CA]**	**[C]**	**Accuracy**
***Bellota*[B]**	18	2	0	0	90
***Recebo*[R]**	7	9	2	2	45
***CebodeCampo*[CA]**	0	1	19	0	95
***Cebo*[C]**	0	0	1	19	95
Average					81.25

Supplementary Table 3. Prediction capability values and false positive rates (expressed as %) obtained for the training and validation models created for different pairs of diets**.**

Bellota vs. Cebo de Campo	***Bellota* Prediction Capability**	***Cebo de Campo* Prediction Capability**	**False Positive for *Bellota***	**False Positive for *Cebo de Campo***
Training model	100	93.3	6.3	0
Validation model	100	100	0	0

Bellota vs. Recebo	***Bellota* Prediction Capability**	***Recebo* Prediction Capability**	**False Positive for *Bellota***	**False Positive for *Recebo***
Training model	80	73.3	25.0	21.4
Validation model	100	60	28.57	0

Bellota vs. Cebo	***Bellota* Prediction Capability**	***Cebo* Prediction Capability**	**False Positive for *Bellota***	**False Positive for *Cebo***
Training model	100	86.7	1.27	0
Validation model	100	60	28.58	0

Cebo vs. Recebo	***Cebo* Prediction Capability**	***Recebo* Prediction Capability**	**False Positive for *Cebo***	**False Positive for *Recebo***
Training model	86.7	86.7	13.39	13.3
Validation model	80	100	0	16.6

***Cebo de Campo vs.* Cebo**	***Cebo de Campo* Prediction Capability**	***Cebo* Prediction Capability**	**False Positive for *Cebo de Campo***	**False Positive for *Cebo***
Training model	100	93.3	6.3	0
Validation model	100	100	0	0

DISCUSSION OF THE RESULTS

The present Thesis Book is based on the format of articles compilation (published or next to publication) regulated by University of Córdoba. Therefore, articles were included as such. A joint discussion of the results obtained according to the objectives initially planned is necessary to provide a global vision of the main results derived from the Doctoral Thesis.

The highest present challenge in metabolomics is to maximize the detection capability of the analytical methods to achieve unequivocal identification of thousands of metabolites, thus making possible to generate representative results of complex metabolomes. Associated to this challenge, the general objective of this Doctoral Thesis was to propose innovations in the different stages of the analytical process involving sampling, sample preparation, detection and data analysis to improve the detection capacity of analytical methods. The research developed in this Thesis Book is divided into three sections according to the aim and the specific metabolomic strategy used for each purpose. Thus, Section I is concerned to demonstrate that the versatility of the triple quadrupole analyzer allows improving the identification/quantification of certain families of metabolites. Section II deals with untargeted metabolomics analysis to study those aspects related with the detection coverage with the aim of improving the identification. By contrast, Section III focuses on strategies for searching potential biomarkers in nutrimetabolomics studies, with clinical and agro-food applications.

In this part of the Doctoral Thesis Book the most relevant results obtained throughout the experimental development proposed in the different chapters are presented. One common link among sections is the main detection technique: mass spectrometry with several variations. The analytical sensitivity, selectivity, accuracy, precision and resolution of mass spectrometry make this technique the most preferred detection tool for targeted and untargeted metabolomic analysis.

Section I. Methodological development in targeted metabolomics analysis: versatility of the triple quadrupole mass spectrometer

Low resolution mass spectrometry, essentially with a QqQ configuration, is an ideal tool to undertake quantitative and confirmatory analysis in metabolomics and, therefore, to work in targeted analysis. However, the QqQ mass spectrometer offers additional features that make it a detector that can play an important role in qualitative analysis [1,2]. For this purpose, in Section I, the development of DDM was used with the aim of increasing the identification power in the analysis of certain families of lipids: FAHFAs and polar lipids (ceramides, glycero-phospholipids and sphingolipids).

A novel class of endogenous mammalian lipids endowed with antidiabetic and anti-inflammatory properties were recently discovered: FAHFAs formed by condensation between a hydroxy fatty acid and a fatty acid [3]. Research on FAHFAs has revealed multiple effects that improve glucose-insulin homeostasis. FAHFAs are present in human serum and tissues at low nanomolar concentrations, similar to other signaling lipids such as endocannabinoids. Therefore, high sensitive and selective profiling analysis of these compounds in clinical samples is demanded.

For this purpose, in Chapter I, an automated qualitative and quantitative method based on SPE–LC–MS/MS has been developed for determination of FAHFAs in serum with the required sensitivity and selectivity. Matrix effects were evaluated by preparation of calibration models in serum and methanol. Recovery factors ranged between 73.8 and 100% in serum. The within-day variability ranged from 7.1 to 13.8 %, and the between-days variability varied from 9.3 to 21.6 %, which are quite acceptable values taking into account the low concentration levels at which the target analytes are found.

The fragmentation scheme for PAHSA, as a model compound to explain the product ions selected for each FAHFA, is shown in Figure 1. As can be seen, three

main product ions were formed according to this fragmentation scheme. The most intense product ion corresponded to the fatty acid —in this example to palmitic acid carboxylate ion at *m/z* 255. The hydroxy fatty acid moiety fit the second product ion, which was formed by cleavage of the ester bound. This ion was obtained at *m/z* 299 for PAHSA. A third representative fragment was obtained by dehydration of the last ion — in this case at *m/z* 281.

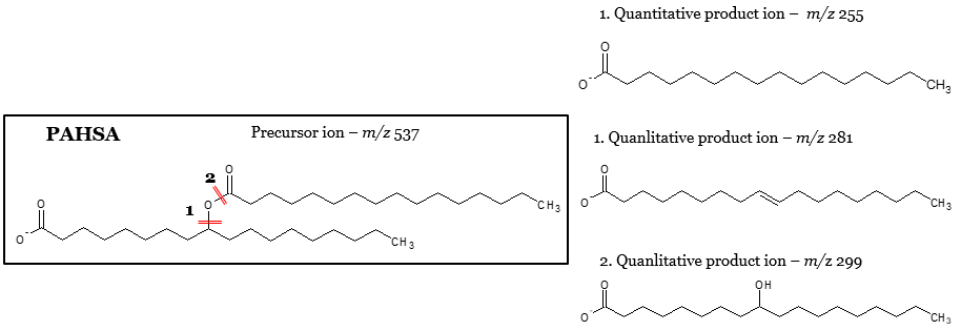


Figure 1. PAHSA precursor ion and product ions in negative ionization mode.

According to the sensitivity of the different transitions from precursor to product ions, those leading to the fatty acid carboxylate fragments were selected for quantitative purposes, while the other two transitions were selected with qualitative interest in the MRM method. Additionally, FAHFAs were confirmed using the NLS, considering two neutral losses for each compound (example in Figure 2). These were formed by generation of the two ions selected as qualifiers. Therefore, the MRM method was configured for quantification of FAHFAs in real samples.

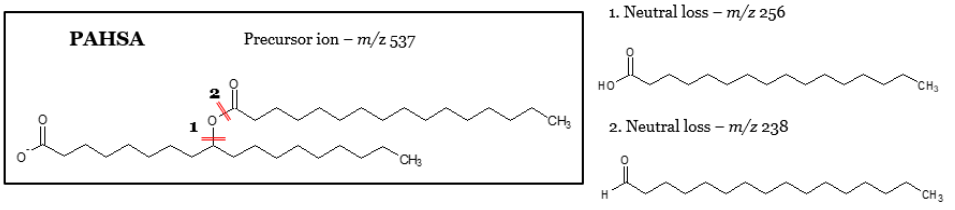


Figure 2. PAHSA precursor ion and neutral losses in negative ionization mode.

The method was applied to a cohort of human individuals to evaluate the influence of the glycaemic state and BMI on FAHFA levels. The application of the method to the studied cohort allowed detecting 11 FAHFAs in at least 75% of the samples. The most concentrated FAHFAs in serum were PAHSA and PAHOA, with relative concentrations around 35 and 25%, respectively; followed by SAHOA, OAHOA and OAHSA, with relative concentrations from 5 to 15%. Considering the glycaemic state, two compounds reported significant differences in their concentrations: PAHPA, which resulted significant between diabetic and prediabetic individuals (*p*-value 0.0149); while POHPO was significant for the comparison between prediabetic and normoglycaemic individuals (*p*-value 0.0251). On the other hand, the BMI only contributed to explain the variability in the concentration of POHPA when normal and overweighed individuals were compared (*p*-value 0.0300). This automated method could be implemented in a similar matrix, such as plasma, with minimum user assistance.

Polar lipids, especially glycerophospholipids, constitute the main components of cell membranes and are precursors of signaling molecules in many cellular and physiological processes. For this reason, the development of methods with high capability for multiple detection of polar lipids in biological samples is required.

In Chapter II, the objective was to develop a method for massive qualitative/quantitative determination of polar lipids in plasma by combination of acquisition methods with a triple quadrupole mass analyzer. The methods have been targeted at the following families of polar lipids: CERs, SMs, SBs, LPLs, PCs, PEs, PAs, PGs, PIs, PSs and plasmalogens (O-alkyls and O-alkenyls). The strategy was optimized in two steps: a) a first step for detection of lipids by monitoring selective fragmentation patterns representative of each lipid family; and b) a second step for confirmation of lipid species by detection and identification of product ions associated with the conjugated fatty acids.

A practical example of strategy used for lysoPC(18:0) is shown in Figure 3. In the first step the representative fragment of PCs, which corresponds to the

phosphorylcholine moiety *m/z* 184.1, was monitored in positive ionization mode, while the R-COO⁻ ion of stearic acid (*m/z* 283.3) was monitored in negative ionization mode in the second step. The retention time in both steps was the same.

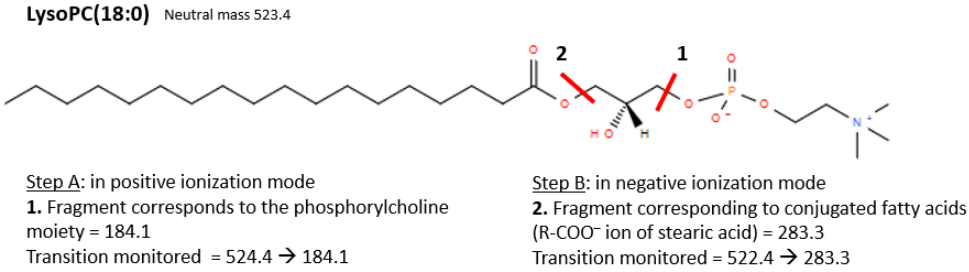


Figure 3. Practical example of the strategy used for development of the methods.

The acquisition list was divided in two MRM methods to ensure the detection of all transitions with enough instrumental sensitivity according to chromatographic retention time and relative abundance in plasma.

The combination of the two MRM methods allowed the detection of 398 polar lipids in plasma in 64 min (32 min per run). This strategy has been applied to a cohort formed by 384 individuals in order to obtain a qualitative and quantitative distribution of polar lipids in human plasma. The most concentrated lipid families in relative terms were LPLs, plasmalogens and PCs, with mean relative concentration of 58.0, 17.1 and 8.3%, respectively. Then, SMs and PEs reported a relative concentration of 2.0%, followed by PSs with 1.1%.

The proposed method could be used in lipidomic analysis of biological samples to provide qualitative and semiquantitative information of lipid polar families. The QqQ analyzer combines the qualitative potential of DDMs, namely NLS, PIS and PrIS, with the quantitative possibilitities of the MRM mode once the suited transitions have been defined.

Section II. Methodological development in untargeted metabolomics analysis: improvement of the analytical process

As mentioned, one of the main challenges of metabolomics is to improve the detection capacity and, subsequently, the identification of metabolites. This is specially relevant in untargeted analysis, where it is also critical to reduce instrumental variability in large batches of samples. These weaknesses of untargeted analysis by mass spectrometry are generally associated to a critical step of the analytical process: detection. However, other steps such as sampling, sample preparation and data analysis should be also considered. For this reason, the main aims of Section II were to develop new strategies applied throughout these other steps of the analytical process —sampling (Chapter III), preparation of the sample (Chapter IV and V) and data analysis (Chapter VI)— to improve the detection and identification capacity in untargeted metabolomic analysis or to reduce the effect of variability sources on large sets of samples. The results provided by each step (sampling, sample preparation and data analysis) are summarized below.

Sampling

Human blood is the most common biofluid used by clinicians, but blood collection is a frequently forgotten aspect that can be responsible for uncontrolled variability sources in metabolomic analysis [4,5]. There are several types of commercial tubes (with different stopper, stopper lubricant, separator gel, clot activator, etc.) for sampling serum or plasma, which are widely used in metabolomic analysis [6]. However, most of the studies dealing with the influence of blood collection tubes on metabolomics profiling have been restricted to conventional tubes for plasma and serum [4,5]. Polymeric gel tubes, which are frequently proposed to accelerate the separation of serum and plasma, had not been studied.

The research described in Chapter III is focused on the study of the differences at metabolite level between serum and plasma obtained with

conventional tubes and polymeric gel tubes from 13 healthy volunteers. This study was addressed by application of an untargeted approach based on GC–TOF/MS analysis.

A total of 65 metabolites were identified in the analyzed samples. The metabolites were identified at least in all the samples pertaining to one of the four target types (serum, plasma, serum-gel and plasma-gel). The main classes of identified compounds were amino acids (14 metabolites), fatty acids (14 metabolites), carbohydrates (11 metabolites) and keto acids (5 metabolites).

The main metabolic differences between serum and plasma collected in conventional tubes affected to critical pathways such as the citric acid cycle, metabolism of amino acids, fructose and mannose metabolism and that of glycerolipids, and pentose and glucuronate interconversion.

On the other hand, significant changes attributable to the polymeric gel were only detected in serum, while no differences were observed in plasma, which in overall terms provided a metabolite profile similar to that of plasma collected in conventional tubes. Changes occurring in serum were mainly found in the metabolism of amino acids, particularly alanine, proline and threonine; in the metabolism of glycerolipids, detected through changes in the levels of glycerol and two important monoglycerides such as monopalmitin and monostearin, and in two metabolites (aconitic acid and lactic acid) involved in primary pathways. Although no overall changes were observed in plasma, significant differences in the concentration of five metabolites, four amino acids and phosphate, were found. An additional issue was to evaluate the metabolite differences between serum and plasma collected from the same group of individuals in conventional tubes. These differences affected to critical pathways such as the citric acid cycle, the metabolism of amino acids, the fructose and mannose metabolism and that of glycerolipids, and pentose/glucuronate interconversions.

It is essential to take into account these alterations when an experimental protocol for metabolomic analysis, using serum or plasma, is planned to select the most adequate sample.

Sample preparation

The influence of sample preparation on detection coverage was studied in Chapters IV and V with two applications: lipidomics analysis of human adipose tissue and untargeted metabolomics analysis of fecal samples from pigs to be lately correlated with microbiota alterations (Annex I).

The main limitations of lipidomics analysis are the chemical complexity of the different families of lipids [7,8], the range of concentrations at which they exist [8], and the diversity of samples usually analyzed. These limitations particularly affect the characterization of polar lipids owing to the interference of neutral lipids, essentially acylglycerides, which are at high concentration and suppress ionization of low concentrated lipids in mass spectrometry detection [9].

Chapter IV was planned to evaluate the influence of sample preparation on lipidomics analysis of polar lipids in visceral adipose tissue by LC–QTOF MS/MS. Two common extractants used for lipids isolation, MeOH:$CHCl_3$ and MTBE, were qualitatively and quantitatively compared for the extraction of the main families of lipids. Also, the implementation of an SPE step for selective isolation of glycerophospholipids prior to LC–MS/MS analysis was assayed to evaluate its influence on lipids detection coverage as compared to direct analysis. This step was critical to enhance the detection coverage of glycerophospholipids by removal of ionization suppression effects caused by acylglycerides.

A total of 94 compounds pertaining to 13 lipid classes were tentatively identified by LC–QTOF MS/MS. The PEs, PIs, and plasmalogens were the families with a higher number of identified compounds: 13%, 12% and 12% of the total identified lipids, respectively.

The influence of the extractants on the analysis of polar lipids was evaluated. The number of metabolites identified was quite similar, 89 and 94 for MeOH:$CHCl_3$ and MTBE, respectively. Despite the protocols seem to be qualitatively similar, in terms of identified lipids, a semiquantitative approach was necessary to clarify the efficiency of each protocol. With this purpose, each familiy

of lipids was idependently studied. In general, the use of MTBE as extractant led to higher extraction efficiency for unsaturated fatty acids, glycerophospholipids and ceramides, while MeOH:$CHCl_3$ favored the isolation of saturated fatty acids and plasmalogens.

Concerning the implementation of SPE for selective isolation of glycerophospholipids, this step clearly improved the detection of minor glycerophospholipids as compared to direct analysis of the extracts, increasing by 50% the number of detected glycerophospholipids.

According to these results, the recommended sample preparation for analysis of polar lipids in adipose tissue would be liquid–liquid extraction combined with an SPE step to enhance detection of glycerophospholipids. Concerning the extractant, MTBE favored the detection of less abundant lipids such as ceramides and unsaturated fatty acids and, therefore, it would be suggested for untargeted analysis of polar lipids.

Regarding Chapter V, blood (serum and plasma) is the most commonly used biofluid for examinig alteration of metabolites because they are reasonably easy to obtain and are collected in a relatively non-invasive manner [5]. However, feces are an interesting biological sample to be implemented in metabolomics experiments by virtue of the information that can be deduced from the interaction between host and microbiome [10,11]. Despite of this fact, fecal samples have received scant attention, especially in untargeted metabolomics studies.

With these premises, an analytical strategy was planned in Chapter V to maximize the identification coverage of metabolites found in pig fecal samples. Intestinal pathologies caused by zoonotic pathogens such as *Salmonella spp* are considered one of the most important risk factors affecting pig farms. This intestinal disease can be transmitted to humans. Furthermore, the pig is extensively used as preferred animal model for analysis of a wide range of physiological functions and diseases. For these reasons, untargeted metabolomic analysis of pig fecal samples can be key to elucidate biological mechanisms and facilitate early diagnosis of pig diseases.

The influence of sample preparation on the identification coverage for analysis of pig fecal samples by LC–MS/MS and GC–MS has been studied in this research. Concerning sample preparation six extractants with different polarity grade were tested to evaluate the extraction performance and, in the case of GC–MS, two derivatization protocols were compared.

A total of 303 compounds by combination of all the extractants and analytical platforms were tentatively identified. The main identified families were amino acids, fatty acids and derivatives, carbohydrates and carboxylic acids. According to the results obtained, it should be recommended the utilization of MeOH/water as extractant prior to GC–MS analysis. This extractant allowed identifying 126 metabolites by this approach. On the other hand, for LC–MS/MS analysis a dual extraction approach with MeOH or MeOH/water, and ethyl acetate is proposed to enhance the detection of polar and non-polar metabolites, respectively. 20 and 42 metabolites were identified exclusively by MeOH/water and ethyl acetate, respectively. Therefore, a fractionation of the metabolome by LLE improved the identification coverage. Concerning the derivatization step, the implementation of methoximation prior to silylation provided the identification of three α-keto acids that are not detected with the other tested strategies. Since these are important metabolites related to microbiome status, the implementation of a double derivatization strategy could be of interest for some studies.

Regarding the complementarity of the two analytical platforms evaluated, LC–MS/MS and GC–MS, their combined use allowed the identification of 303 metabolites with only 12 common metabolites. Thus, it is obvious that both platforms should be combined to obtain a comprehensive view of the pig feces metabolome.

Data analysis

Chapter VI was focused on the final stage of the analytical process: data analysis. The current trend in metabolomics workflows is the analysis of large sets of samples to obtain representative information from the biological system under study. Thus, most studies require long periods to analyze samples in which the

quantitative response can fluctuate by alteration of the instrument performance owing to accumulation of matrix components in different instrumental zones or simply by periodical practices such as instrument calibration or cleaning protocols. These sources of instrumental variability are generally corrected in targeted analysis by using isotopically labeled internal standards [12]. However, this is not viable in untargeted analysis owing to the wide chemical heterogeneity of metabolites that would force to spike the sample with an isotopically labeled standard per detected metabolite.

Implementation of QC samples to check the contribution of experimental variability is the most common approach in metabolomics [13–15]. This practice is based on filtration of molecular entities experiencing a variation coefficient higher than that measured in the QC data set. Although other robust correction algorithms have been proposed, none of them has provided an easy-to-use and easy-to-install tool capable of correcting experimental variability sources [14,16].

In this chapter an R-package —MetaboQC— was developed to correct intra-day and inter-days variability using QCs analyzed within a pre-set sequence of experiments. The proposed package is based on functions that can be used to correct variability on data sets obtained in metabolomics studies based on large set of samples. Different functions have been considered to make the package useful to remove experimental variability sources affecting long sequences of analysis, but also those affecting within a given day. The strategy included in the proposed package involves that each metabolite is corrected according to the function that best fits its variability trend. Therefore, correction is independently applied to each metabolite. The only requirement for its application is the implementation of QCs, preferentially prepared with the same samples of the cohort, in the sequence of analysis following a given plan.

MetaboQC has been tested in two data sets to assess the correction effects by comparing the metabolites variability before and after application of the proposed tool. As a result, in QCs the number of entities significantly different between days was reduced from 86 to 19% in the negative ionization mode and

from 100 to 13% in the positive ionization mode. Furthermore, principal component analysis allowed detecting the filtration of instrumental variability associated to the injection order.

This type of tools can be recommended in cases in which the use of internal standards is not operative owing to chemical diversity of metabolome composition.

Section III. Strategies for searching potential biomarkers in clinical and agro-food studies

Metabolomics affords detailed characterization of metabolic phenotypes and enables to address precision medicine at different levels, including the characterization of metabolic derangements that underlie a disease, and discovery of biomarkers that may be used to either diagnose disease or dietary patterns [17,18]. This section is focused on strategies for searching potential biomarkers for food intake, dietary patterns and diseases.

Metabolomics allows to evaluate the metabolic response to intake with the aim of predicting metabolic alterations. Modifications on food related metabolome include variations in the concentrations of metabolites directly derived from the ingestion of diet components. These aspects have been studied in clinical (Chapter VII) and agro-food (Chapter X) applications. On the other hand, strategies for searching potential biomarkers of diseases have been considered in Chapter VIII and IX, being T2DM and T1DM the studied diseases.

Clinical applications

It is widely known that the postprandial response to a meal depends on many factors and involves multiple processes that include energy storage and metabolic switch in several organs such as liver, muscle and adipose tissue [19]. In addition, this metabolic response is accompanied by several compensating processes such as inflammation and oxidative stress. Therefore, analysis of the metabolic response after meal ingestion indicate a shift in metabolic pathways

from catabolic (fatty acid oxidation) to anabolic (suppression of ketogenesis and lipolysis) conditions [20]. This shift can be associated with incident pathologies, such as metabolic syndrome and T2DM. In this context, metabolomics studies can lead to the identification of metabolites involved in the mechanisms of a disease by monitoring metabolite changes in predisposed individuals compared with healthy cases.

Chapter VII is focused on detecting postprandial alterations in the level of plasma metabolites after the OFTT, which consisted of a weight-adjusted meal (0.7 g fat and 5 mg cholesterol per kg body weight) with 12% SFAs, 10% PUFAs, 43% MUFAs, 10% protein, and 25% carbohydrates. Meal preparation for the OFTT was performed by nutritionists using olive oil, skimmed milk, white bread, cooked egg yolks, and tomatoes [21]. Plasma samples were collected from 215 patients, just before and four hours after the OFTT. These plasma samples were analyzed by LC–QTOF MS/MS and GC–TOF/MS in two batches including 57 and 158 individuals, respectively.

A total number of 365 metabolites were tentatively identified by combination of both analytical platforms. The paired *t*-test led to the identification of 18 metabolites significantly altered ($p<0.05$) due to the OFTT in both batches. It is worth mentioning that main metabolic alterations affected fatty acids and derivatives, bile acids, bilirubins and neurotransmitters. The significant compounds showed FC positive values, except for taurodeoxycholic acid, four fatty acids, C20:4n6, C20:2n6, C20:5n3 and C20:3n3, and five acylcarnitines. The most important pathways and processes affected by OFTT were inflammatory and oxidative processes, *de novo* lipogenesis, primary and secondary bile acids synthesis and cortisol synthesis.

Palmitoleic acid (C16:1n7) increased its concentration considerably in the OFTT postprandial since its mean FC value is 63.7. Palmitoleic acid is a unique fatty acid that serves as a marker for *de novo* lipogenesis, a process that converts glucose to fatty acids. The level of palmitoleic acid in the diet is limited and, consequently, its concentration is low in tissue. Nevertheless, the concentration of

palmitoleic acid can quickly and substantially increase upon activation of *de novo* lipogenesis. Cao *et al*. identified palmitoleic acid as an adipose tissue-derived lipid hormone that strongly stimulates muscle insulin action and suppresses hepatosteatosis [22]. These authors proved that adipose tissue uses palmitoleic acid as a signal to communicate with distant organs and regulate metabolic homeostasis [22]. This situation would explain the increased levels of palmitoleic acid in the postprandial OFTT.

Regarding PUFAs, comprising the omega-6 (*n*-6) and omega-3 (*n*-3) fatty acids, they have received considerable attention due to their important effects on human health [23,24]. In this study, n-3 and n-6 fatty acids did not report a common pattern after OFTT. Docosapentaenoic acid (C22:5n3) and docosatetraenoic acid (C22:4n6), significant PUFAs, increased their concentration in plasma levels after OFTT (69.9 and 2.8 FC values). Emerging evidence suggests that C22:5n3, the intermediate fatty acid between EPA and docosahexanoic acid (C22:6n3), may also play a role in imparting the health benefits previously attributed solely to EPA and C22:6n3 [23]. On the other hand, EPA and arachidonic acid, precursors of C22:5n3 and C22:4n6, decreased their concentration significantly according to the FC values (−0.16 and −0.07 respectively). EPA and arachidonic acid are precursors of series 3 and 2 prostaglandins, thromboxanes and leukotrienes, which are directly involved in the inflammatory process. The decrease of EPA and arachidonic acid is supported by activation in the production of eicosanoids and increase in the concentration of C22:5n3 and arachidonic acid, the latter dependent of elongase activity.

Also, three bile acids experienced significant metabolic alterations in the OFTT postprandial. These were glycohoclic acid, a primary bile acid, and glycoursodeoxycholic acid and taurodeoxycholic acid, two secondary bile acids. The concentration of glycocholic acid and glycoursodeoxycholic acid were increased after the OFTT (0.78 and 5.5 FC value), while taurodeoxycholic acid was decreased (–0.06 FC value).

Fourteen acylcarnitines experienced significant metabolic alterations after OFTT. All acylcarnitines do not follow the same behavior after the OFTT. According to the FC values, half of them increased their concentration and the other half decreased their levels. Free carnitine decreased in the postprandial OFTT, which may be due to its conversion into acylcarnitines [40]. The acylcarnitine that reported a greater increase after OFTT in postprandial time was octadecenoylcarnitine (38.4 FC value). This result is in concordance with the composition of OFTT, with 43% of MUFAs, being oleic acid (C18:1) the most concentrated MUFA in olive oil.

Although other authors have studied the change produced in acylcarnitines after a specific meal, there is no consensus about their behavior, since the concentration of acylcarnitines is significantly affected by individual anthropometric factors, nutritional composition and duration of the test, etc. Shrestha *et al.* reported similar results in human serum after a specific meal intake, but in this case five acylcarnitines reported a decrease after meal [25], including octadecenoylcarnitine and hexadecanoylcarnitine, which increased after OFTT (38.4 and 0.05 FC values, respectively) in our study.

Other important processes affected to cortisol, bilirubin and biliverdin. Increased cortisol production (5.4 FC value) was observed in the OFTT postprandial. Cortisol is considered a stress hormone involved in the response to physical and/or emotional stress and participates in several actions involved in the homeostatic maintenance: blood pressure, immune system, metabolism of proteins and carbohydrates, and anti-inflammatory action [26–30]. On the other hand, bilirubin and biliverdin have been recognized as potent antioxidants [31–33] and both increased after OFTT according to FC values (5.5 and 17.9, respectively).

This identification of postprandial metabolic alterations occurring during the OFTT can lead to interpret deviations associated to metabolic diseases.

T2DM is the most prevalent metabolic disease in the world and is characterized by defects in insulin secretion and peripheral insulin resistance in the skeletal muscle, adipose tissue and liver [34]. However, the pathogenesis of T2DM is still not fully dilucidated and seems to involve multiple factors. As mentioned above, metabolomics studies allow discovering metabolites involved in a disease mechanism [35,36]. Thus, multimetabolite panels with predictive capacity of T2DM can be developed.

Chapter VIII is focused on predicting the occurrence of T2DM in individuals by examining their postprandial response to the OFTT between non-T2DM and incident-T2DM. For this purpose, plasma samples from 215 individuals (107 classified as incident-T2DM and 108 as non-T2DM regarding T2DM development in the following 5 years) were analyzed using a metabolomic untargeted approach.

This research allows identifying 60 metabolites that showed significant differences associated with the development of T2DM. Significant metabolites were used to develop multimetabolite panels with predictive capacity. Two multimetabolite panels were created to predict the occurrence of T2DM by prioritizing specificity (panel 1) and sensitivity (panel 2). The combination of both panels reported a sensitivity of 86.6% and specificity of 71.6%, while external validation led to 90.0% of sensitivity and 73.3% of specificity. The risk of T2DM development was evaluated by Cox Regression Analysis that provided HR values of 6.5 (3.7–11.4) and 5.4 (3.0–9.6) for panels 1 and 2, respectively.

Finally, the pathways and processes affected by OFTT between non-T2DM and incident-T2DM were studied, being the most important the oxidation process, insulin secretion and, mitochondrial and peroxisomes activity.

T1DM is a chronic, immune-mediated disease characterized by selective loss of insulin-producing β-cells in the pancreatic islets of genetically susceptible individuals [37,38]. Over the past few decades, the incidence of T1DM in most Western countries has been increasing, particularly among children below the age of five [39]. About 70% of children with T1DM carry increased risk-associated

genotypes in HLA loci. On the other hand, only 3–7% of the population with the same risk alleles develop T1DM [40].

Previous metabolomics studies suggest that T1DM is preceded by specific metabolic disturbances. Here, we asked whether distinct metabolic patterns occur in PBMCs of children later developing pancreatic β-cell autoimmunity or overt T1DM. In a longitudinal cohort setting, PBMC metabolomic analysis was applied in children who either (1) progressed to T1DM (PT1D, n=34), (2) seroconverted to ≥1 islet autoantibody without progressing to T1DM (P1Ab, n=27), or (3) remained autoantibody negative during follow-up (CTRL, n=10).

The observed metabolic changes in PBMCs were dominated by glycerophospholipids, triacylglycerols, sphingomyelins, ceramides, amino acids and fatty acids. During the first year of life, levels of most lipids and polar metabolites were lower in PT1D and P1Ab, *versus* controls. These results suggest that progression to T1DM is accompanied by metabolic abnormalities in PBMCs. These changes may be related to impaired *de novo* lipogenesis and metabolism of amino acids, glycerophospholipids and sphingolipids. Since specific differences were also observed between progressors and non-progressors to T1DM after their seroconversion to islet autoimmunity, our findings also highlight specific pathways in immune cells, such as sphingolipid metabolism, which appear to play a role in protection from and progression to T1DM.

Agro-food application

Concerning the agro-food area, discrimination among the types of feeding regimes for Iberian pigs is currently a highly demanded challenge by the Iberian pig sector [41]. At present, there are not discrimination methods able to identify the feeding regime used with capability to detect frauds. Different methods have been independently assayed [42–44]. However, no methods based on the application of different analytical techniques had been combined so far.

In Chapter X, discrimination among feeding regimes was achieved by the combination of two analytical methods (based on FAMEs analysis by GC–FID and

determination of $\delta^{13}C$ by IRMS) previously used independently without success. For this purpose, 80 samples of adipose tissue from Iberian pigs subjected to four different feedings were analyzed.

An ANOVA test was independently applied to each pair of feedings to evaluate its influence on the concentration of FAs and $\delta^{13}C$. The most concentrated FAs in subcutaneous fat, C16:0, C18:0, and C18:1n9, were significant in most of the evaluated pairs, except for Cebo de Campo *vs.* Cebo and Recebo *vs.* Bellota. It is worth mentioning that the significance level was quite high in all cases with *p*-values below 0.001. Only two FAs, C17:0 and C18:2, were not significantly affected by the type of feeding. The study of the most influenced variables by the feeding regime allowed configuring panels of markers with predictive power for the studied feedings by multivariate ROC analysis. Taking into account that Bellota is the most appreciated class, prediction models were created for the classification of Bellota *vs.* Cebo de Campo, Bellota *vs.* Cebo, and Bellota *vs.* Recebo. The models involving Bellota *vs.* Cebo de Campo and Cebo reported high sensitivity and specificity, always above 90%, which led to quite good predictive models. The most complicated case was the model involving Bellota *vs.* Recebo, based on a unique variable, $\delta^{13}C$, which reported a specificity of 80%, while sensitivity decreased to 63%.

In view of these results, the statistical combination of the data obtained by different analytical methods could be the key for the correct discrimination of feeding regimes in the sector of Iberian pig. Taking into account the demand for methods to discriminate feeding regimes, this approach could be implemented in routine laboratories since the analysis of fatty acids is already carried out at this level, while IRMS analysis is frequently used in reference food laboratories.

The next step in the research in Section III is to validate the developed methods with more subjects to prove the robustness of the proposed models.

References

[1] A. Technology, Mass Spectrometry Fundamentals–Theory (accessed on 12th May 2019), (2016).

[2] H.G. Gika, G.A. Theodoridis, R.S. Plumb, I.D. Wilson, Current practice of liquid chromatography – mass spectrometry in metabolomics and metabonomics, J. Pharm. Biomed. Anal. 87 (2014) 12–25.

[3] M.M. Yore, I. Syed, P.M. Moraes-Vieira, T. Zhang, M.A. Herman, E.A. Homan, R.T. Patel, J. Lee, S. Chen, O.D. Peroni, A.S. Dhaneshwar, A. Hammarstedt, U. Smith, T.E. McGraw, A. Saghatelian, B.B. Kahn, Discovery of a class of endogenous mammalian lipids with anti-diabetic and anti-inflammatory effects, Cell 159 (2014) 318–332.

[4] C. Prehn, Y. He, P. Belcredi, G. Mo, Z. Yu, G. Kastenmu, Differences between human plasma and serum metabolite profiles, PLoS One 6 (2011) e21230.

[5] M.J.N. de Paiva, H.C. Menezesa, Z. de L. Cardeala, Sampling and analysis of metabolomes in biological fluids, Analyst 139 (2014) 3683–3694.

[6] S. Cuhadar, A. Atay, M. Koseoglu, A. Dirican, A. Hur, Stability studies of common biochemical analytes in serum separator tubes with or without gel barrier subjected to various storage conditions, Biochem. Med. 22 (2012) 202–2014.

[7] E. Fahy, S. Subramaniam, H.A. Brown, C.K. Glass, A.H. Merrill, R.C. Murphy, C.R.H. Raetz, D.W. Russell, Y. Seyama, W. Shaw, T. Shimizu, F. Spener, G. van Meer, M.S. VanNieuwenhze, S.H. White, J.L. Witztum, E. Dennis, A comprehensive classification system for lipids, J. Lipid Res. 46 (2005) 839–861.

[8] K. Sandra, P. Sandra, Lipidomics from an analytical perspective, Curr. Opin. Chem. Biol. 17 (2013) 847–853.

[9] R.C. Baker, Y. Nikitina, A.R. Subauste, Analysis of Adipose Tissue Lipid using Mass Spectrometry, in: Methods Enzymology, Elsevier, 2014: pp. 89–105.

[10] S. Matysik, C.I. Le Roy, G. Liebisch, S.P. Claus, Metabolomics of fecal samples: a practical consideration, Trends Food Sci. Technol. 57 (2016) 244–255.

[11] Y. Sun, Y. Su, W. Zhu, Microbiome-metabolome responses in the cecum and colon of pig to a high resistant starch diet, Front. Microbiol. 7 (2016) 1–10.

[12] C. Bueschl, R. Krska, B. Kluger, R. Schuhmacher, Isotopic labeling-assisted metabolomics using LC – MS, Anal. Bioanal. Chem. 405 (2013) 27–33.

[13] X. Shen, X. Gong, Y. Cai, Y. Guo, J. Tu, Normalization and integration of large-scale metabolomics data using support vector regression, Metabolomics 12 (2016) 89.

[14] W.B. Dunn, 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, T. Human, S. Metabolome, H. Consortium, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid

chromatography coupled to mass spectrometry, Nat. Protoc. 6 (2011) 1060–1083.

[15] M. Chen, R.S.P. Rao, Y. Zhang, C.X. Zhong, J.J. Thelen, A modified data normalization method for GC-MS-based metabolomics to minimize batch variation, Springerplus 3 (2014) 1–7.

[16] Y. V Karpievitch, S.B. Nikolic, R. Wilson, J.E. Sharman, Metabolomics data normalization with EigenMS, PLoS One 9 (2014) e116221.

[17] R. Llorach, M. Garcia-Aloy, S. Tulipani, R. Vazquez-Fresno, C. Andres-Lacueva, Nutrimetabolomic strategies to develop new biomarkers of intake and health effects, J. Agric. Food Chem. 60 (2012) 8797–8808.

[18] C.B. Clish, Metabolomics: an emerging but powerful tool for precision medicine, Mol. Case Stud. 1 (2015) a000588.

[19] D.J. Vis, J.A. Westerhuis, D.M. Jacobs, J.P.M. van Duynhoven, S. Wopereis, B. van Ommen, M.M.W.B. Hendriks, A.K. Smilde, Analyzing metabolomics-based challenge tests, Metabolomics 11 (2014) 50–63.

[20] A. Shrestha, E. Müllner, K. Poutanen, H. Mykkänen, A.A. Moazzami, Metabolic changes in serum metabolome in response to a meal, Eur. J. Nutr. 56 (2017) 671–681.

[21] A. Leon-Acuña, J.F. Alcala-Diaz, J. Delgado-Lista, J.D. Torres-Peña, J. Lopez-Moreno, A. Camargo, A. Garcia-Rios, C. Marin, F. Gomez-Delgado, J. Caballero, B. Van-Ommen, M.M. Malagon, P. Perez-Martinez, J. Lopez-Miranda, Hepatic insulin resistance both in prediabetic and diabetic patients determines postprandial lipoprotein metabolism: From the CORDIOPREV study, Cardiovasc. Diabetol. 15 (2016) 1–10.

[22] H. Cao, K. Gerhold, J.R. Mayers, M.M. Wiest, S.M. Watkins, Identification of a lipokine, a lipid hormone linking adipose tissue to systemic metabolism, Cell 134 (2008) 933–944.

[23] B.S. Muhlhausler, The Effect of Dietary Modification on Polyunsaturated Fatty Acid Biosynthesis and Metabolism, in: Polyunsaturated Fatty Acid Metabolism, Elsevier, 2018: pp. 181–192.

[24] W. Gładkowski, G. Kiełbowicz, A. Chojnacka, M. Gil, T. Trziszka, C. Wawrzen, Z. Dobrzan, Fatty acid composition of egg yolk phospholipid fractions following feed supplementation of Lohmann Brown hens with humic-fat preparations, Food Chem. 126 (2011) 1013–1018.

[25] A.M. Zivkovic, M.M. Wiest, U. Nguyen, M.L. Nording, M. Watkins, J.B. German, Assessing individual metabolic responsiveness to a lipid challenge using a targeted metabolomic approach, Metabolomics 5 (2010) 209–218.

[26] I. Jeong, The role of cortisol in the pathogenesis of the metabolic syndrome, Diabetes Metab. J. 36 (2012) 207–210.

[27] M.J. Lee, S.K. Fried, S.S. Mundt, Y. Wang, S. Sullivan, A. Stefanni, B.L. Daugherty,

A. Hermanowski-Vosatka, Depot-specific regulation of the conversion of cortisone to cortisol in human adipose tissue, Obesity 16 (2008) 1178–1185.

[28] R.N. Bergman, D. Stefano*vs*.ki, S.P. Kim, Systems analysis and the prediction and prevention of Type 2 diabetes mellitus, Curr. Opin. Biotechnol. 28 (2014) 165–170.

[29] T. Kuo, A. Mcqueen, T.-C. Chen, J.-C. Wang, Regulation of glucose homeostasis by glucocorticoids, Adv. Exp. Med. Biol. (2015) 99–126.

[30] American Diabetes Association, Diagnosis and classification of Diabetes Mellitus: new criteria, Diabetes Care 33 (2010) S62-S69.

[31] L. Vítek, The role of bilirubin in diabetes, metabolic syndrome, and cardiovascular diseases, Front. Pharmacol. 3 (2012) 1–7.

[32] L. Vítek, J.D. Ostrow, Bilirubin chemistry and metabolism; harmful and protective aspects, Curr. Pharm. Des. 15 (2009) 2869–2883.

[33] L. Vítek, H.A. Schwertner, Protective effects of serum bilirubin on peripheral vascular disease, Ann. Hepatol. 7 (2008) 94–95.

[34] M. Guasch-ferr, M.A. Mart, Metabolomics in prediabetes and diabetes: a systematic review and meta-analysis, Diabetes Care 39 (2016) 833–846.

[35] H. Witt, M. Mueller, I. Padberg, E. Peter, S. Gonza, A new metabolomic signature in type-2 diabetes mellitus and its pathophysiology, PLoS One 9 (2014) e85082.

[36] K. Suhre, Metabolic profiling in diabetes, J. Endocrinol. 221 (2014) 75–85.

[37] M. Knip, K. Luopajärvi, T. Härkönen, Early life origin of type 1 diabetes, Semin. Immunopathol. 39 (2017) 653–667.

[38] M.A. Atkinson, G.S. Eisenbarth, Type 1 diabetes: new perspectives on disease pathogenesis and treatment, Lancet 358 (2001) 221–229.

[39] E.J. Mayer-Davis, J.M. Lawrence, D. Dabelea, J. Divers, S. Isom, L. Dolan, G. Imperatore, B. Linder, S. Marcovina, D.J. Pettitt, C. Pihoker, S. Saydah, L. Wagenknecht, Incidence trends of type 1 and type 2 diabetes among youths, 2002-2012, N. Engl. J. Med. 376 (2017) 1419–1429.

[40] P. Achenbach, E. Bonifacio, K. Koczwara, A.-G. Ziegler, Natural history of type 1 diabetes, Diabetes 54 (2005) 25–31.

[41] D. Montet, R.C. Ray, Food Traceability and Authenticity: Analytical Techniques, CRC Press, 2017.

[42] J.M.G. Casco, M. Muñoz, E. González, Predictive ability of the feeding system in Iberian pig by means of several analytical methods, Grasas y Aceites 64 (2013) 191–200.

[43] L. Arce, A. Domínguez-Vidal, V. Rodríguez-Estévez, S. López-Vidal, M.J. Ayora-Cañada, M. Valcárcel, Feasibility study on the use of infrared spectroscopy for the direct authentication of Iberian pig fattening diet, Anal. Chim. Acta 636 (2009)

183–189.

[44] C. Gonza, J. Herna, Â. Me, I. Gonza, F.S. Poveda, E. Marque, Use of isotope analysis to characterize meat from Iberian-breed swine, Meat Sci. 52 (1999) 437–441.

CONCLUSIONS

CONCLUSIONES

The research conducted in this Doctoral Thesis was aimed at developing new analytical strategies based on the use of low- and high-resolution mass spectrometry to improve the detection and identification coverage in metabolomic analysis.

The most salient conclusions drawn from this work can be summarized as follows according to the objectives:

1. Benefits derived from the versatility of the triple quadrupole analyzer to improve the identification/quantification of certain families of metabolites.

    i) An automated method based on SPE on-line coupled to LC–MS/MS has been developed for determination of FAHFAs in serum. Eleven FAHFAs have been identified and quantified in relative terms in serum by application of a confirmatory test. PAHSA and PAHOA were the most concentrated FAHFAs in serum. PAHPA and POHPO reported significant differences between glycaemic states, while only POHPA was found significantly different considering BMI.

    ii) A dual analysis strategy for massive quantitative determination of polar lipids by LC–MS/MS with a QqQ analyzer was optimized. A combination of two MRM methods enabled to monitor 398 polar lipids in 64 minutes (32 minutes per run) after application of a simple protocol for sample preparation. The proposed method can be used in lipidomics analysis of biological samples to provide qualitative and semiquantitative information of lipid polar families.

2. Improvement of the analytical process (sampling, sample preparation and data analysis) through methodological development in untargeted metabolomics analysis.

Sampling

(iii) It is essential to take into account that metabolic alterations occur in applying experimental protocols for metabolomics analysis in blood sampled. Serum and plasma collected in polymeric gel tubes were compared with serum and plasma obtained in conventional tubes using a GC–TOF/MS untargeted approach. Significant changes attributable to the polymeric gel were only detected in serum, while no differences were observed in plasma, which in overall terms provided a metabolite profile similar to that of plasma collected in conventional tubes. An additional issue was to evaluate the metabolite differences between serum and plasma collected from the same group of individuals in conventional tubes. These differences affected to critical pathways such as the citric acid cycle, the metabolism of amino acids, the fructose and mannose metabolism and that of glycerolipids, and pentose/glucuronate interconversions.

Sample preparation

(iv) The influence of sample preparation for lipidomics analysis of polar lipids in adipose tissue was studied. According to these results, the recommended sample preparation for analysis of polar lipids in adipose tissue would be LLE combined with an SPE step to enhance detection of glycerophospholipids. Concerning the extractant, MTBE favored the detection of less abundant lipids such as ceramides and unsaturated fatty acids and, therefore, it would be recommended for untargeted analysis of polar lipids.

(v) The influence of sample preparation on the identification coverage for pig fecal samples analysis by LC–MS/MS and GC–MS has been evaluated. A total number of 303 compounds by combination of all the extractants and analytical platforms were tentatively identified. According to the results obtained, it should be recommended the utilization of MeOH/water as extractant for GC–MS analysis, but for LC–MS/MS analysis the combined analysis of extracts obtained with MeOH or MeOH/water and ethyl acetate can lead to a significant increase of identified compounds. Concerning the derivatization step, the implementtation of methoximation previous silylation provided the identification of three α-keto acids that are not detected by the other tested strategies. Regarding the complementarity of the two analytical platforms, it is obvious that both approaches should be combined to obtain a comprehensive view of the pig feces metabolome.

Data analysis

(vi) A tool for correction of experimental variability associated to the instrumental quantitative response has been developed for implementation in metabolomics workflows based on MS detection. The proposed package is based on functions that can be used to correct variability on data sets obtained in metabolomics studies with large set of samples. The strategy included in this package involves that each metabolite is corrected according to the function that best fits its variability trend. Therefore, correction is independently applied to each metabolite. The only requirement for its application is the implementation of QCs, preferentially prepared with the same samples of the cohort, following a given planning in the sequence of analysis.

3. Development of strategies for searching potential biomarkers in clinical and agro-food applications.

Clinical applications

(vii) Postprandial alterations in the level of plasma metabolites after the OFTT were studied by combination of LC–QTOF MS/MS and GC–TOF/MS. The most important metabolic alterations affected inflammatory and oxidative processes, synthesis of primary and secondary bile acids and cortisol production. This study revealed that OFTT can be used to interpret deviations associated to metabolic diseases, increasing its usefulness.

(viii) The capability of metabolic changes occurring in the OFTT postprandial to predict the development of T2DM has been assessed. Taking into account the complexity of T2DM pathogenesis, two multimetabolite panels were configured to identify future T2DM patients. The combination of the two panels led to a model with sensitivity of 86.6% and specificity of 71.6%. The HR obtained for both panels, 5.4 (3.0–9.6) and 6.5 (3.7–11.4), revealed the predictive power, which reflected metabolic alterations associated to oxidation, insulin secretion and mitochondrial and peroxisomes activity.

(ix) Metabolic patterns occurring in PBMCs of children developing pancreatic β-cell autoimmunity or overt T1DM have been studied. Pathway analysis suggested that alanine, aspartate, glutamate, glycerol phospholipid and sphingolipid metabolism were overrepresented in PT1D. Genome-scale metabolic models of PBMCs in T1DM progression were developed using available transcriptomics data and constrained with metabolomics results. Metabolic modeling confirmed altered ceramide pathways as specifically associated with T1DM progression.

Agro-food application

(x) The statistical combination of the results obtained by different analytical methods (concretely fatty acids determination by GC–FID and $\delta^{13}C$ by IRMS analysis) could be the key for the correct discrimination of feeding regimes in the sector of Iberian pig. Taking into account the demand for methods to discriminate feeding regimes, this approach could be implemented in routine laboratories since the analysis of fatty acids is already carried out at this level, while IRMS analysis is frequently used in reference food laboratories.

La investigación realizada en esta Tesis Doctoral tenía como objetivo desarrollar nuevas estrategias analíticas basadas en el uso de espectrometría de masas de baja y alta resolución para mejorar la capacidad de detección y la cobertura de identificación en análisis metabolómico.

Las conclusiones más destacadas de este trabajo de acuerdo con los objetivos inicialmente propuestos, se resumen a continuación:

1. Beneficios derivados de la versatilidad del analizador de triple cuadrupolo para mejorar la identificación/cuantificación de ciertas familias de metabolitos.

    i) Se ha desarrollado un método automatizado basado en SPE en línea acoplada a LC–MS/MS para la determinación de FAHFAs en suero. El método permitió identificar y cuantificar en términos relativos 11 FAHFAs en suero mediante el desarrollo de una estrategia de análisis confirmatorio. Los FAHFAs más concentrados en suero fueron PAHSA y PAHOA. Se han detectado diferencias significativas en los niveles de PAHPA y POHPO en función del estado glucémico, y de POHPA en función del BMI.

    ii) Se ha optimizado una estrategia de análisis para la identificación y cuantificación masiva de lípidos polares por LC–MS/MS con un analizador QqQ. La combinación de dos métodos MRM permitió monitorizar 398 lípidos polares en 64 minutos (32 minutos por cada método) después de la aplicación de un protocolo de preparación de muestra simple. La estrategia propuesta puede usarse en el análisis lipidómico de muestras biológicas para proporcionar información cualitativa y semicuantitativa de familias de lípidos polares.

2. Mejora del proceso analítico (muestreo, preparación de muestra y análisis de datos) a través del desarrollo metodológico en el análisis metabolómico no dirigido.

Muestreo

(iii) Es esencial tener en cuenta las alteraciones que se producen en los protocolos experimentales de análisis metabolómico cuando se toman muestras de sangre. Se han comparado suero y plasma recogidos en tubos de gel polimérico con suero y plasma obtenidos en tubos convencionales utilizando un enfoque no dirigido mediante GC–TOF/MS. Se detectaron cambios significativos atribuibles al gel polimérico en suero, mientras que no se observaron diferencias en el plasma, que proporcionó un perfil de metabolitos similar al del plasma recogido en tubos convencionales. Adicionalmente, se han evaluado las diferencias de metabolitos entre suero y plasma recolectado en tubos convencionales del mismo grupo de individuos. Estas diferencias afectaron a rutas importantes como el ciclo del ácido cítrico, el metabolismo de los aminoácidos, de la fructosa y la manosa y de los glicerolípidos, y las interconversiones pentosa/glucuronato.

Preparación de la muestra

(iv) Se ha estudiado la influencia de la preparación de la muestra en el análisis lipidómico de lípidos polares en tejido adiposo. Según los resultados obtenidos, se recomienda la LLE seguida de una etapa SPE para mejorar la detección de glicerofosfolípidos. Respecto al extractante, el MTBE favoreció la detección de lípidos menos abundantes como ceramidas y ácidos grasos insaturados y, por lo tanto, se recomienda para el análisis no dirigido de lípidos polares.

(v) Se ha evaluado la influencia de la preparación de la muestra en la capacidad de detección de metabolitos en el análisis de muestras de heces de cerdo mediante LC–MS/MS y GC–MS. Se han

identificado tentativamente un total de 303 compuestos mediante combinación de todos los extractantes y plataformas analíticas estudiadas. Según los resultados obtenidos, se recomienda la utilización de MeOH/agua como extractante para el análisis GC–MS. Sin embargo, para LC–MS/MS el análisis combinado de extractos obtenidos con MeOH o MeOH/agua y acetato de etilo puede conducir a un aumento significativo de los compuestos identificados. Respecto a la etapa de derivatización por sililación, la implementación de una metoximación previa proporcionó la identificación de tres α-cetoácidos que no se detectan con las otras estrategias probadas. Con respecto a la complementariedad de las dos plataformas analíticas, es obvio que ambas deben combinarse para obtener una visión integral del metaboloma de las heces de cerdo.

Análisis de datos

(vi) Se ha desarrollado una herramienta para la corrección de la variabilidad experimental asociada a la respuesta cuantitativa instrumental y su implementación en flujos de trabajo de metabolómica basados en la detección por MS. El paquete propuesto se basa en funciones que pueden usarse para corregir la variabilidad en los datos obtenidos en estudios de metabolómica con un gran conjunto de muestras. La estrategia incluida en este paquete implica que cada metabolito se corrige de acuerdo con la función que mejor se adapte a su tendencia de variabilidad. Por lo tanto, la corrección se aplica independientemente a cada metabolito.

3. Desarrollo de estrategias para la búsqueda de biomarcadores potenciales en aplicaciones clínicas y agroalimentarias.

Aplicaciones clínicas

(vii) Se han estudiado las alteraciones metabólicas producidas en el postprandio del OFTT, mediante la combinación de LC–QTOF MS/MS y GC–TOF/MS. Las alteraciones metabólicas más importantes afectaron a los procesos inflamatorios y oxidativos, la síntesis de ácidos biliares primarios y secundarios, la síntesis de carnitinas y la producción de cortisol. Este estudio reveló que el OFTT se puede utilizar para interpretar las desviaciones asociadas a enfermedades metabólicas, aumentando su utilidad.

(viii) Se ha evaluado la capacidad predictiva de los cambios metabólicos que ocurren en el postprandio del OFTT para el desarrollo de T2DM. Teniendo en cuenta la complejidad de la patogénesis de T2DM, se configuraron dos paneles multimetabolitos para identificar futuros pacientes con T2DM. La combinación de los dos paneles condujo a un modelo con una sensibilidad del 86,6% y una especificidad del 71,6%. El HR obtenido para ambos paneles, 5.4 (3.0–9.6) y 6.5 (3.7–11.4), reveló su poder predictivo, los cuales reflejan alteraciones metabólicas asociadas a la oxidación, la secreción de insulina, y la actividad mitocondrial y de los peroxisomas.

(ix) Se han estudiado los patrones metabólicos que se producen en las PBMCs de niños que han desarrollado autoinmunidad en las células β pancreáticas o T1DM. El análisis de rutas metabólicas sugirió que el metabolismo de la alanina, el aspartato, el glutamato, los glicerofosfolípidos y los esfingolípidos estaban sobreexpresados en PT1D. Se han desarrollado modelos metabólicos a escala del genoma de las CMSP en la progresión de la T1DM utilizando los datos transcriptómicos disponibles y los resultados

del análisis metabolómico. Los modelos desarrollados mostraron alteraciones en la ruta metabólica de las ceramidas, asociadas específicamente con la progresión de la T1DM.

Aplicación agroalimentaria

(x) La combinación estadística de los resultados obtenidos por diferentes métodos analíticos (concretamente la determinación de ácidos grasos por GC–FID y de $\delta^{13}C$ por análisis IRMS) podría ser la clave para la correcta discriminación de los regímenes alimenticios en el sector del cerdo ibérico. Teniendo en cuenta la demanda de métodos para discriminar los regímenes de alimentación, este enfoque podría implementarse en laboratorios de rutina ya que el análisis de ácidos grasos ya se lleva a cabo a este nivel, mientras que el análisis IRMS se usa con frecuencia en laboratorios alimentarios de referencia.

ANNEXES

ANNEX I

Other publications co-authored by PhD student

Other publications co-authored by the PhD student

Collaborations with other members of the group and with other group, which have provided 2 articles published in high impact international journals.

1. Determination of primary fatty acid amides in different biological fluids by LC–MS/MS in MRM mode with synthetic deuterated standards: Influence of biofluid matrix on sample preparation.
   L.S. Castillo-Peinado, ***M.A. López-Bascón****, A. Mena-Bravo, M.D. Luque de Castro, F. Priego-Capote.*
   Talanta 193 (2018) 29-37

---

2. Early Salmonella Typhimurium infection in pigs disrupts Microbiome composition and functionality principally at the ileum mucosa.
   H. Argüello, J. Estellé, S. Zaldívar-López, Á. Jiménez-Marín, A. Carvajal, ***M.A. López-Bascón****, F. Crispie, O. O'Sullivan, P.D. Cotter, F. Priego-Capote, L. Morera, J.J. Garrido.*
   Scientific Reports 7788 (2018) 1-12

Talanta 193 (2019) 29–36

Contents lists available at ScienceDirect

Talanta

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

Determination of primary fatty acid amides in different biological fluids by LC–MS/MS in MRM mode with synthetic deuterated standards: Influence of biofluid matrix on sample preparation

L.S. Castillo-Peinado[a,b,c,d], M.A. López-Bascón[a,b,c,d], A. Mena-Bravo[a,b,c,d], M.D. Luque de Castro[a,b,c,d], F. Priego-Capote[a,b,c,d,*]

[a] Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, Córdoba, Spain
[b] Maimónides Institute of Biomedical Research (IMIBIC), Reina Sofía University Hospital, University of Córdoba, Córdoba, Spain
[c] University of Córdoba Agroalimentary Excellence Campus, ceiA3, Córdoba, Spain
[d] CIBER Fragilidad y Envejecimiento Saludable (CIBERfes), Instituto de Salud Carlos III, Spain

ARTICLE INFO

Keywords:
Primary fatty acid amides
Liquid chromatography–tandem mass spectrometry
Stable isotopically labeled internal standard
Human biofluids
Sample preparation

ABSTRACT

The recent growing interest in primary fatty acid amides (PFAMs) is due to the broad range of physiological effects they exhibit as bioindicator of pathological states. These bioactive lipids are usually in biological samples at the nanomolar level, making their detection and identification a challenging task. A method for quantitative analysis of seven main PFAMs (lauramide, myristamide, linoleamide, palmitamide, oleamide, stearamide and behenamide) in four human biofluids —namely, urine, plasma, saliva and sweat— is here reported. Two sample preparation procedures were compared to test their efficiency in each biofluid: solid-phase extraction (SPE) and protein precipitation. The latter was the best for plasma and urine, while the analysis of saliva and sweat required an SPE step for subsequent suited determination of PFAMs. Detection of the seven metabolites was performed by liquid chromatography coupled to tandem mass spectrometry (LC–MS/MS) in multiple reaction monitoring (MRM) mode. Quantitative analysis was supported on the use of stable isotopically labeled internal standards (SIL-ISs) in the calibration method, which required the synthesis of each IS from the precursor deuterated fatty acids. Detection limits for the target analytes were within 0.3–3 ng mL^{-1}. The method was applied to a small cohort of male and female volunteers ($n = 6$) to estimate the relative concentration profiles in the different biofluids. The analytical features of the method supported its applicability in clinical studies aimed at elucidating the role of PFAMs metabolism.

1. Introduction

Fatty acid amides are a part of a large family of structurally diverse molecules found in humans and other organisms. Particularly, primary fatty acid amides (PFAMs) are a subclass of fatty acyls that contain a carboxamide headgroup and an acyl tail with variable carbon length and unsaturations. Five PFAMs (palmitamide, palmitoleamide, oleamide, elaidamide and linoleamide) were first identified in human luteal phase plasma [1]. The interest in these molecules was modest until the isolation of oleamide and erucamide from cat, rat, and human cerebrospinal fluid were achieved [2]. In general, PFAMs are considered a class of neuromodulatory lipids that regulate the nervous system together with different classes of chemical transmitters. These compounds are related to neuronal communication, neuroplasticity, neuroinflammation and individual behavior [3–5]. Also, these bioactive lipids have been linked to biological processes such as sleep regulation and modulation of monoaminergic systems [6]. Analogous PFAMs seem to exhibit common biological properties, but their functionality seems to be moderately governed by their chemical structure [7–9]. Research on these compounds has been focused mostly on oleamide, known for its role in regulating the sleep/wake cycle, blocking gap junction communication in glial cells, regulating memory processes, decreasing body temperature and locomotor activity, and stimulating Ca^{2+} release, among other functions [3,7,10–12].

Abbreviations: ACN, acetonitrile; c-μSPE, centrifugal micro-solid phase extraction; EIC, extracted ion chromatogram; FAAH, fatty acid amide hydrolase; FA, formic acid; LC–MS/MS, liquid chromatography coupled to tandem mass spectrometry; MeOH, methanol; MRM, multiple reaction monitoring; PFAMs, primary fatty acid amides; PIM, product ion monitoring; RSD, relative standard deviation; SIL-ISs, stable isotopically labeled internal standards; TIC, total ion chromatogram

* Corresponding author at: Department of Analytical Chemistry, Annex Marie Curie Building, Campus of Rabanales, University of Córdoba, Córdoba, Spain.
E-mail address: feliciano.priego@uco.es (F. Priego-Capote).

https://doi.org/10.1016/j.talanta.2018.09.088
Received 10 July 2018; Received in revised form 21 September 2018; Accepted 24 September 2018
Available online 25 September 2018
0039-9140/

www.nature.com/scientificreports

OPEN

Early *Salmonella* Typhimurium infection in pigs disrupts Microbiome composition and functionality principally at the ileum mucosa

Received: 5 October 2017
Accepted: 2 May 2018
Published online: 17 May 2018

Héctor Argüello[1], Jordi Estellé[2], Sara Zaldívar-López[1], Ángeles Jiménez-Marín[1], Ana Carvajal[3], Mª Asunción López-Bascón[4], Fiona Crispie[5,6], Orla O'Sullivan[5,6], Paul D. Cotter[5,6], Feliciano Priego-Capote[4], Luis Morera[1] & Juan J. Garrido[1]

Salmonella **is a major foodborne pathogen which successfully infects animal species for human consumption such as swine. The pathogen has a battery of virulence factors which it uses to colonise and persist within the host. The host microbiota may play a role in resistance to, and may also be indirectly responsible from some of the consequences of, *Salmonella* infection. To investigate this, we used 16S rRNA metagenomic sequencing to determine the changes in the gut microbiota of pigs in response to infection by *Salmonella* Typhimurium at three locations: ileum mucosa, ileum content and faeces. Early infection (2 days post-infection) impacted on the microbiome diversity at the mucosa, reflected in a decrease in representatives of the generally regarded as desirable genera (i.e., *Bifidobacterium* and *Lactobacillus*). Severe damage in the epithelium of the ileum mucosa correlated with an increase in synergistic (with respect to *Salmonella* infection; *Akkermansia*) or opportunistically pathogenic bacteria (*Citrobacter*) and a depletion in anaerobic bacteria (*Clostridium spp.*, *Ruminococcus*, *or Dialliser*). Predictive functional analysis, together with metabolomic analysis revealed changes in glucose and lipid metabolism in infected pigs. The observed changes in commensal healthy microbiota, including the growth of synergistic or potentially pathogenic bacteria and depletion of beneficial or competing bacteria, could contribute to the pathogen's ability to colonize the gut successfully. The findings from this study could be used to form the basis for further research aimed at creating intervention strategies to mitigate the effects of *Salmonella* infection.**

Salmonella is a major foodborne pathogen with a significant number of cases and outbreaks every year[1]. Poultry meat, eggs and pork are the main sources of human salmonellosis. As a consequence of the successful establishment of control programmes in poultry, the relative proportion of human cases attributed to pork consumption has increased[2].

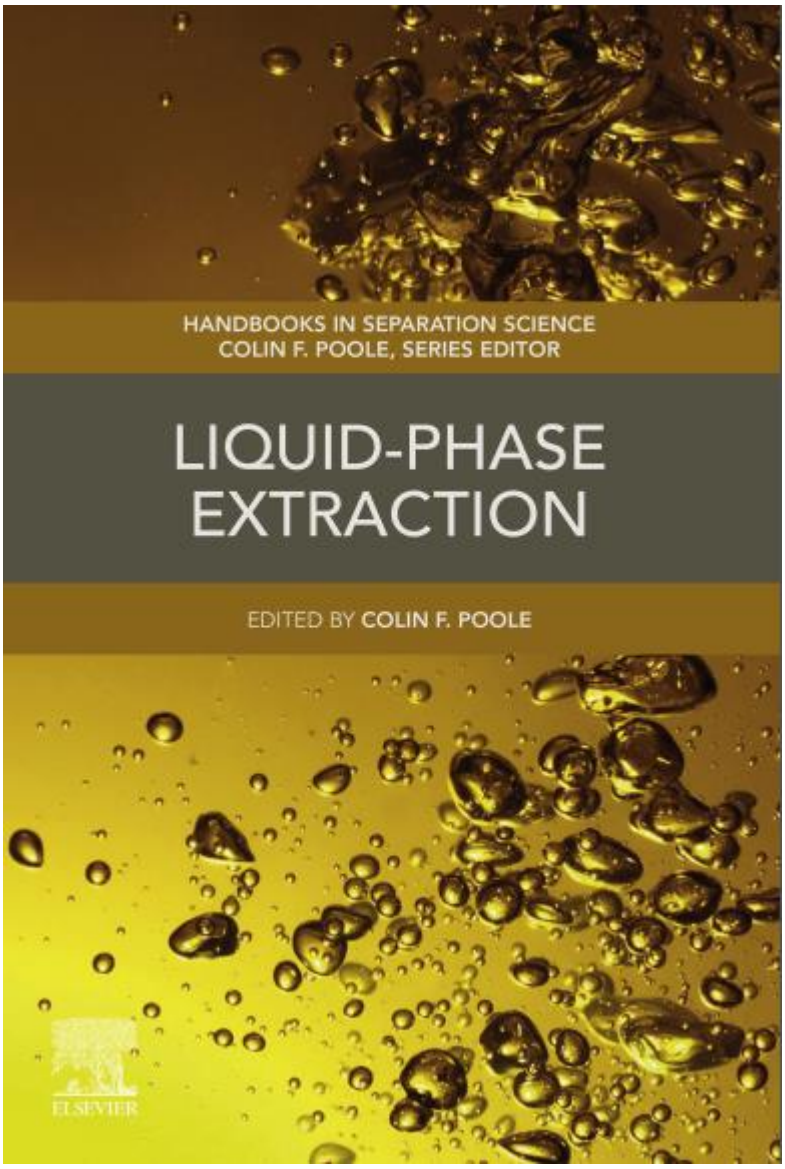
Swine are a natural host for *Salmonella spp*. Apart from infections caused by *S.* Cholerasuis, the host adapted serovar to swine, *Salmonella* infections in pigs usually vary from subclinical to a self-limiting diarrhoea[3]. Among serovars affecting swine, *Salmonella* Typhimurium stands out due to its zoonotic importance in human salmonellosis[1] and its widespread distribution in pork production[4,5]. The relevance of *S.* Typhimurium infections in swine is reflected by the vast literature published with respect to the pathogenesis[6], the host immune response[7,8],

[1]Grupo de Genómica y Mejora Animal, Departamento de Genética, Facultad de Veterinaria, Universidad de Córdoba, 14047, Córdoba, Spain. [2]GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350, Jouy-en-Josas, France. [3]Departamento de Sanidad Animal, Facultad de Veterinaria, Universidad de León, 24071, León, Spain. [4]Departamento de Química Analítica Universidad de Córdoba, Córdoba, CeiA3 Campus de Excelencia Agroalimentaria, Universidad de Córdoba, 14047, Córdoba, Spain. [5]Teagasc, Food Research Centre, Moorepark, Fermoy, Co., Cork, Ireland. [6]APC Microbiome Institute, Cork, Ireland. Correspondence and requests for materials should be addressed to H.A. (email: ge2arroh@uco.es)

ANNEX II

Book chapter on a subject non-related to the Thesis

Book chapter on a subject non-related to the Thesis



Chapter 11

Soxhlet Extraction

M.A. López-Bascón, M.D. Luque de Castro
Department of Analytical Chemistry, University of Córdoba, Córdoba, Spain

Chapter Outline

Liquid-Phase Extraction. https://doi.org/10.1016/B978-0-12-816911-7.00011-4

ANNEX III

Oral and poster communications in national or international meetings

1. Discrimination of Iberian pigs as a function of the feeding regime by combined analysis of fatty acids and stable carbon isotope ratio in subcutaneous fat

M.A. López-Bascón, F. Priego-Capote, M. Calderón-Santiago, V. Sánchez de Medina, J.M. Moreno-Rojas, J.M. García-Casco, M.D. Luque de Castro

IV CONGRESO CIENTÍFICO DE INVESTIGADORES EN FORMACIÓN DE LA UNIVERSIDAD DE CÓRDOBA

Córdoba, 2014

Tipo de comunicación: Comunicación oral **Ámbito:** Regional

2. Análisis lipidómico del tejido adiposo en pacientes con obesidad y resistencia a insulina

M.M. Malagón, J. Sánchez, A. Fernández, R. Guzmán, M. Calderón, M.A. López-Bascón, J. López, R. Vázquez

XII CONGRESO NACIONAL DE LA SOCIEDAD ESPAÑOLA PARA EL ESTUDIO DE LA OBESIDAD (SEEDO)

Málaga, 2015

Tipo de comunicación: Póster en Congreso **Ámbito:** Nacional

3. Confirmatory and quantitative analysis of fatty acid esters of hydroxy fatty acids in serum by solid phase extraction coupled to liquid chromatography tandem mass spectrometry

M.A. López-Bascón, M. Calderón-Santiago, F. Priego-Capote, M.D. Luque de Castro

VII JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2016

Tipo de comunicación: Póster en Congreso **Ámbito:** Internacional

4. A lipidomic approach to study adipose tissue from obese and insulin resistant patients

A. Fernández-Vega, J. Sánchez-Ceinos, R. Guzmán-Ruiz, S. García Navarro, M. Calderón-Santiago, M.A. López-Bascón, F. Priego Capote, J. López-Miranda, R. Vázquez-Martínez, M.M. Malagón

VIII JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2017

Tipo de comunicación: Póster en Congreso **Ámbito:** Internacional

5. Influencia de la preparación de muestra en el análisis lipidómico de tejido adipose

M.A. López-Bascón, M. Calderón-Santiago, A. Fernández-Vega, J. Sánchez-Ceinos, M.M. Malagon, J. López-Miranda, F. Priego-Capote

IV CONGRESO CIENTÍFICO DE INVESTIGADORES EN FORMACIÓN DE LA UNIVERSIDAD DE CÓRDOBA

Córdoba, 2016

Tipo de comunicación: Comunicación oral **Ámbito:** Regional

6. Influence of sample preparation in lipidomic analysis of adipose tissue

M.A. López-Bascón, M. Calderón-Santiago, A. Fernández-Vega, J. Sánchez-Ceinos, M.M. Malagon, J. López-Miranda, F. Priego-Capote

XV REUNIÓN DEL GRUPO REGIONAL ANDALUZ DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA (GRASEQA)

Almeria, 2016

Tipo de comunicación: Póster en Congreso **Ámbito:** Regional

7. Analytical platforms for qualitative/quantitative analysis of microbiota related metabolites in nutritional studies

M.A. López-Bascón, M. Calderón-Santiago, F. Priego-Capote

VIII JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2017

Tipo de comunicación: Póster en Congreso **Ámbito:** Internacional

8. Postprandial response to the oral-fat tolerance test (OFTT) on plasma metabolomics profile

M.A. López-Bascón, M. Calderón-Santiago, A. Camargo, J. López-Miranda, F. Priego-Capote

IX JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2018

Tipo de comunicación: Comunicación oral **Ámbito:** Internacional

9. Determination of primary fatty acid amides in different biofluids by LC–MS/MS: use of synthesized deuterated standards and study of sample preparation

L.S. Castillo-Peinado, M.A. López-Bascón, A. Mena-Bravo, M.D. Luque de Castro, F. Priego-Capote

IX JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2018

Tipo de comunicación: Comunicación oral **Ámbito:** Internacional

10. Contribuciones del grupo FQM227 en metabolómica clínica y agroalimentaria

M.A. López-Bascón, M. Calderón-Santiago, M.M. Delgado Povedano, A. Mena Bravo, C.A. Ledesma Escobar, L.S. Castillo Peinado, A. Díaz Lozano, D. Luque Córdoba, I. Criado Navarro, M.D. Luque de Castro, F. Priego-Capote

VII JORNADAS DE DIVULGACIÓN DE LA INVESTIGACIÓN EN BIOLOGÍA MOLECULAR, GENÉTICA Y BIOTECNOLOGÍA

Córdoba, 2018

Tipo de comunicación: Comunicación oral **Ámbito:** Nacional

11. Lipid fingerprint of the adipose tissue in obesity and insulin resistance

A. Fernández-Vega, E. Chicano-Gálvez, B. Prentice, F. Priego-Capote, M.A. López-Bascón, J. Teso-Rodríguez, L. Molero-Murillo, M.S. Avendaño, O.A. Rangel-Zuñiga, M. Tena-Sempere, J. López-Miranda, R. Guzmán-Ruiz, R. Caprioli, M.M. Malagón

52ND ANNUAL SCIENTIFIC MEETING OF THE EUROPEAN SOCIETY FOR CLINICAL INVESTIGATION

Barcelona, 2018

Tipo de comunicación: Póster en congreso **Ámbito:** Internacional

12. Potenciación de la autonomía del estudiante en el laboratorio analítico: experiencia en la asignatura "Química Analítica Aplicada"

M.L. Soriano, J. Ríos-Gómez, M.A. López-Bascón, L. Arce Jiménez

IV JORNADA SOBRE ESTRATEGIAS PARA LA INNOVACIÓN DE LA ACTIVIDAD DOCENTE EN QUÍMICA ANALÍTICA

Barcelona, 2018

Tipo de comunicación: Póster en congreso **Ámbito:** Nacional

13. Postprandial response to the oral-fat tolerance test (OFTT) on plasma metabolomics profile

M.A. López-Bascón, M. Calderón-Santiago, A. Camargo, J. López-Miranda, F. Priego-Capote

1ST NORDIC METABOLOMICS CONFERENCE

Örebro, Suecia, 2018

Tipo de comunicación: Póster en congreso **Ámbito:** Internacional

14. Postprandial response to the oral-fat tolerance test (OFTT) on plasma metabolomic profile to predict diabetes

M.A. López-Bascón, M. Calderón-Santiago, A. Camargo, J. López-Miranda, F. Priego-Capote

XVI REUNIÓN DEL GRUPO REGIONAL ANDALUZ DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA (GRASEQA)

Granada, 2018

Tipo de comunicación: Comunicación oral **Ámbito:** Regional

15. Postprandial response to the oral-fat tolerance test (OFTT) on plasma metabolomic profile to predict diabetes

M.A. López-Bascón, M. Calderón-Santiago, A. Camargo, J. López-Miranda, F. Priego-Capote

CLINICAL METABOLOMICS COPENHAGEN CONFERENCE

Copenhagen, Dinamarca, 2018

Tipo de comunicación: Póster en congreso **Ámbito:** Internacional

16. Sphingolipids and phospholipids in insulin resistance and obesity. New insights from MALDI Imaging analysis

A. Fernández-Vega, E. Chicano-Gálvez, B. PrenticeB, F. Priego-Capote, M.A. López-Bascón, J. Teso-Rodríguez, L. Molero-Murillo, M.S. Avendaño, O.A.

Rangel-Zuñiga, M. Tena-Sempere, J. López-Miranda, R. Guzmán-Ruiz, R. Caprioli, M.M. Malagón

VI JORNADAS DE JÓVENES INVESTIGADORES EN PROTEÓMICA

Madrid, 2019

Tipo de comunicación: Póster en congreso **Ámbito:** Nacional

17. Massive determination of polar lipids in biological samples by LC–MS/MS

M.A. López-Bascón, A. Díaz-Lozano, M. Calderón-Santiago, F. Priego-Capote

IX JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2019

Tipo de comunicación: Póster en congreso **Ámbito:** Internacional

18. Lipid biomarkers of the adipose tissue in obesity-induced insulin resistance

A. Fernández-Vega, E. Chicano-Gálvez, B. Prentice, F. Priego-Capote, M.A. López-Bascón, J. Teso-Rodríguez, L. Molero-Murillo, M.S. Avendaño, O.A. Rangel-Zuñiga, M. Tena-Sempere, J. López-Miranda, R. Guzmán-Ruiz, R. Caprioli, M.M. Malagón

IX JORNADA DE JÓVENES INVESTIGADORES DEL IMIBIC

Córdoba, 2019

Tipo de comunicación: Comunicación oral **Ámbito:** Internacional

19. Improving the identification coverage in metabolomics analysis of pig fecal samples by chromatographic techniques coupled to mass spectrometry in high resolution mode: influence of sample preparation

M.A. López-Bascón, M. Calderón-Santiago, H. Argüello, L. Morera, J.J. Garrido, F. Priego-Capote

XXII REUNIÓN DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA

Valladolid, 2019

Tipo de comunicación: Póster en congreso **Ámbito:** Nacional

20. Massive determination of polar lipids in plasma by LC–MS/MS

M.A. López-Bascón, A. Díaz-Lozano, M. Calderón-Santiago, F. Priego-Capote

XXII REUNIÓN DE LA SOCIEDAD ESPAÑOLA DE QUÍMICA ANALÍTICA

Valladolid, 2019

Tipo de comunicación: Póster en congreso **Ámbito:** Nacional

ANNEX IV

Co-direction of two Final Degree Projects (TFGs) of Degree in Chemistry (University of Córdoba, Spain)

1. Co-direction of Final Degree Project (TFG) of Chemistry Degree (2017/2018)

Student: Sara María Gil Navarro

Title: Estudio del efecto matriz para la determinación de compuestos fenólicos en aceite de oliva virgen mediante cromatografía de líquidos acoplada a espectrometría de masas en tándem.

Mark: 8.5/10.

2. Co-direction of Final Degree Project (TFG) of Chemistry Degree (2018/2019)

Student: María de la Sierra Serrano Muñoz

Title: Análisis del valor saludable del aceite de oliva virgen extra en base a la Directiva Europea 432/2012.

Mark: 9.8/10.

UNIVERSIDAD D CORDOBA

Facultad de Ciencias

MARTA ROSEL PÉREZ MORALES, SECRETARIA DE LA FACULTAD DE CIENCIAS DE LA UNIVERSIDAD DE CÓRDOBA,

CERTIFICA:

Que **Dª. Mª ASUNCIÓN LÓPEZ BASCÓN,** con D.N.I. 15451742C, ha sido tutora del siguiente Trabajo de Fin de Grado:

- QM17-17-QAN "ESTUDIO DEL EFECTO MATRIZ PARA LA DETERMINACIÓN DE COMPUESTOS FENÓLICOS EN ACEITE DE OLIVA VIRGEN MEDIANTE CROMATOGRAFÍA DE LÍQUIDOS ACOPLADA A ESPECTROMETRÍA DE MASAS EN TÁNDEM". Dicho trabajo fue defendido en la convocatoria de junio del curso 2017/2018 y obtuvo la calificación de 8,5 Notable.

Y para que conste y surta los efectos oportunos, firmo el presente en Córdoba, a 7 de mayo de 2019.

Código Seguro de Verificación	V72EZ3RAX4XN2UMDPSSNA46JWU	Fecha y Hora	08/05/2019 08:15:22
Normativa	Este documento incorpora firma electrónica reconocida de acuerdo a la ley 59/2003, 19 de diciembre, de firma electrónica		
Firmado por	MARTA ROSEL PEREZ MORALES		
Url de verificación	https://sede.uco.es/verifirma/	Página	1/1

UNIVERSIDAD DE CÓRDOBA

Facultad de Ciencias

MARTA ROSEL PÉREZ MORALES, SECRETARIA DE LA FACULTAD DE CIENCIAS DE LA UNIVERSIDAD DE CÓRDOBA,

CERTIFICA:

Que **Dª. Mª ASUNCIÓN LÓPEZ BASCÓN** con D.N.I.: 15451742C, ha sido tutora del siguiente Trabajo de Fin de Grado:

- QM18-24-QAN ANÁLISIS DEL VALOR SALUDABLE DE ACEITE DE OLIVA VIRGEN EXTRA EN BASE A LA DIRECTIVA EUROPEA 432/2012. Dicho trabajo fue defendido en la convocatoria de junio del curso 2018/2019 y obtuvo la calificación de 9,8 Matrícula de honor.

Y para que conste y surta los efectos oportunos, firmo el presente en Córdoba, a 22 de julio de 2019.

Código Seguro de Verificación	VV5WQ7F324RKYXPPSSIBZMNC6E	**Fecha y Hora**	22/07/2019 14:06:54
Normativa	Este documento incorpora firma electrónica reconocida de acuerdo a la ley 59/2003, 19 de diciembre, de firma electrónica		
Firmado por	MARTA ROSEL PEREZ MORALES		
Url de verificación	https://sede.uco.es/verifirma/	**Página**	1/1

ANNEX V

Co-direction of two Final Master Projects (TFMs) of Master in Chemistry (University of Córdoba, Spain)

1. **Co-direction of Final Master Project (TFM) of Master's Degree in Chemistry (2017/2018)**

Student: Laura de los Santos Castillo Peinado

Title: Determinación de amidas primarias de ácidos grasos en diferentes biofluidos mediante LC-MS/MS: síntesis de estándares deuterados y estudio de la preparación de la muestra.

Mark: 9.6/10.

2. **Co-direction of Final Master Project (TFM) of Master's Degree in Chemistry (2017/2018)**

Student: Laura Aguilar García

Title: Desarrollo de un método de análisis lipidómico cualitativo/cuantitativo mediante LC-MS/MS con analizador de triple cuadruplo para su aplicación a muestras biológicas.

Mark: 9.6/10.

Julieta Mérida García, Vicerrectora de Posgrado e Innovación Docente de la Universidad de Córdoba

Certifica:

Que, López Bascón, María Asunción, con DNI 15451742-C

· Ha sido Director/a de la asignatura Trabajo Fin de Máster, incluida en el plan de estudios de Máster Universitario en Química, en el curso 2017/2018, con el título: Determinación de amidas primarias de ácidos grasos en diferentes biofluidos mediante LC-MS/MS: síntesis de estándares deuterados y estudio de la preparación de la muestra, elaborado por Laura de los Santos Castillo Peinado con la calificación Sobresaliente.

· Ha sido Director/a de la asignatura Trabajo Fin de Máster, incluida en el plan de estudios de Máster Universitario en Química, en el curso 2017/2018, con el título: Desarrollo de un método de análisis lipidómico cualitativo/cuantitativo mediante LC-MS/MS con analizador de triple cuádruplo para su aplicación a muestras biológicas, elaborado por Laura Aguilar García con la calificación Sobresaliente

Para que conste, a los efectos oportunos, a petición del interesado/a, firmo la presente en CÓRDOBA a 26 de Febrero del 2019

Código Seguro de Verificación	UEDRBSEPHMR24TUWECSCI5B7PI	**Fecha y Hora**	27/02/2019 13:58:56
Normativa	Este documento incorpora firma electrónica reconocida de acuerdo a la ley 59/2003, 19 de diciembre, de firma electrónica		
Firmado por	JULIETA MERIDA GARCIA		
Url de verificación	https://sede.uco.es/verifirma/	**Página**	1/1

ANNEX VI

Simultaneous research in education

Simultaneous research in education, which has provided 1 article published and participation in a teaching innovation project

1. Promotion of self-employed work in the laboratory practices of students of the Degree in Chemistry: case study.
   ***M. A. López-Bascón**, N. Jurado-Campos, E. Romera-García, F. Priego-Capote, L. Arce*
   Accepted in Revista de Innovación y Buenas Prácticas Docentes 193 (2019)

2. Participation in a teaching project of University of Córdoba
   Title: Fomento del trabajo autónomo en las prácticas de laboratorio de alumnos del grado de Química
   Code: 2018-2-2002

REVISTA DE INNOVACIÓN Y BUENAS PRÁCTICAS DOCENTES

PROMOTION OF SELF-EMPLOYED WORK IN THE LABORATORY PRACTICES OF STUDENTS OF THE DEGREE IN CHEMISTRY: CASE STUDY

FOMENTO DEL TRABAJO AUTÓNOMO EN LAS PRÁCTICAS DE LABORATORIO DE ALUMNOS DEL GRADO DE QUÍMICA: CASO DE ESTUDIO

María Asunción López-Bascón, Natividad Jurado-Campos, Encarnación Romera-García, Feliciano Priego-Capote, Lourdes Arce

Abstract:

In this study we have sought to increase the participation of students in laboratory practices. The main reason has been to prevent the student from memorizing the theoretical concepts without being able to apply them to the resolution of real problems, and thus acquire useful competences for when entering into the working world. The new methodology was developed in the optional subject of the Degree in Chemistry, Applied Analytical Chemistry. In laboratory sessions, the common is to carry out a practical exercise following a script established by the teaching staff. Therefore, it was proposed that they themselves elaborate the protocols of three practices, on three subjects proposed by the teaching staff, and that they carried them out independently in the laboratory.

The evaluation of the students' work was carried out through three activities: preparation of the practical protocols, carrying out of the laboratory practices (using a rubric) and interactive questionnaire on theoretical concepts (using the Kahoot tool). Finally, the degree of acceptance of the new teaching methodology by the students was evaluated. This allowed to detect the aspects to be improved for the implementation of this methodology in the next academic year.

Key words: laboratory work, competences, chemistry, criterion-referenced evaluation, self-employed work

ANEXO V. MEMORIA FINAL DE PROYECTOS. MODALIDADES 1, 2, 3 Y 4

CURSO ACADÉMICO 2018/2019

DATOS IDENTIFICATIVOS:

1. Título del Proyecto

Fomento del trabajo autónomo en las prácticas de laboratorio de alumnos del grado de Química

2. Código del Proyecto

2018-2-2002

3. Resumen del Proyecto

En este proyecto se ha perseguido incrementar la participación de los alumnos en las prácticas de laboratorio. El motivo principal para plantear este proyecto ha sido evitar que el alumno simplemente memorice los conceptos teóricos, y que por tanto los comprenda y sea capaz de aplicarlos a la resolución de problemas reales, adquiriendo compentencias útiles para cuando se introduzca en el mundo laboral.

El proyecto se desarrolló en las clases prácticas de aula y laboratorio de una asignatura optativa del Grado de Química, Química Analítica Aplicada. En este tipo de clases, lo más común es realizar un ejercicio práctico siguiendo un guión. Sin embargo, el alumno no es consciente del trabajo que supone elaborar dicho guión para que la práctica se desarrolle con éxito, seleccionando los procedimientos más correctos desde un punto de vista analítico, medioambiental y económico. Por consiguiente, se les propuso que ellos mismos elaboraran los protocolos de tres prácticas, sobre tres temas concretos propuestos por el profesorado, y que las llevaran a cabo de forma autónoma en el laboratorio.

La evaluación del trabajo de los alumnos se realizó mediante tres actividades: elaboración de los protocolos de prácticas, realización de las prácticas de laboratorio (evaluada mediante rúbrica) y cuestionario interactivo sobre conceptos teóricos (mediante el uso de la herramienta Kahoot!). Como resultado, el 93.75% del alumnado obtuvo una calificación superior a 8.5 (sobre 10). La actividad en la que más discreparon sus resultados fue en la actividad programada con Kahoot!. Por otra parte, se evaluó el grado de aceptación de la metodología docente propuesta en este proyecto por parte del alumnado y el resultado general fue positivo. Se pudieron detectar los aspectos a mejorar para la implementación de esta metodología en el año académico próximo, siendo estos: el horario, el tiempo para desarrollar la actividad y la dificultad para encontrar información.

LIST OF ABBREVIATIONS

AAG, alanine, aspartate and glutamate metabolism

ACE, automatic cartridge exchanger

ACN, acetonitrile

ADA, American Diabetes Association

ANOVA, analysis of variance

APCI, atmospheric pressure chemical ionization

APPI, athmospheric pressure photo ionization

AUC, area under the curve

BCAA, branched chain amino acid

BMI, body mass index

BPC, base peak chromatogram

BSTFA, bis-(trimethylsilyl)-fluoroacetamide

CA, clustering analysis

CCM, central carbon metabolism

CE, capillary electrophoresis

CI, chemical ionization

Ci, confidence interval

CID, collision-induced dissociation

CHO, carbohydrate

CMPF, 3-carboxy-4-methyl-5-propyl-2- furanpropionic acid

CORDIOPREV, Coronary Diet Intervention with Olive Oil and Cardiovascular Prevention study

CTRL, control

DDA, data-dependent acquisition

DDM, data-dependent methods

DEG, differential expression of genes

DG, diglyceride

D-GAP, Diabetes-Genes, Autoimmunity and Prevention

DI, direct infusion

DIPP, Type 1 Diabetes Prediction and Prevention study

D-PBS, Dubelcco's phosphate buffered saline

EA, elemental analyzer

EFA, esterified fatty acid

EI, electron impact ionization

EIC, extracted ion chromatogram

EPA, eicosapentaenoic acid

ES, external standard

ESI, electrospray ionization

FA, fatty acid

FAHFA, fatty acid ester of hydroxy fatty acid

FAME, fatty acid methyl ester

FC, fold change

FDR, false discovery rate

FID, flame ionization detector

FT-IR, Fourier transformation infrared

FWHM, full width at half maximum

GADA, glutamic acid decarboxylase

GC, gas chromatography

GEMs, genome-scale metabolic models

GHT, global hypergeometric test

GPF, gas phase fractionation

GSMM, genome-scale metabolic modeling

GST, glycine, serine and threonine metabolism

HbA1c, glycated hemoglobin A1c

HDL, high density lipoprotein

HESI, heated electrospray ionization

HETE, hydroxyeicosatetraenoic acid

HILIC, hydrophilic interaction liquid chromatography

HLA, human leucocyte antigen

HMDB, Human Metabolome Database

HODE, hydroxyoctadecadienoic acid

HP-921, hexakis (1H, 1H, 3H-tetrafluoropropoxy) phosphazine

HPD, high pressure dispenser

HpODE, hydroperoxyoctadecadienoic acid

HR, hazard ratio

IAA, autoantibody against insulin

IA-2A, insulinoma-associated antigen

IIS, International Isotope Standard

INIA, Institute of Agriculture and Food Research and Technology

IRMS, isotope ratio mass spectrometry

IRP2, iron regulatory protein 2

IS, internal standard

IPA, 2-isopropanol

JDFU, Juvenile Diabetes Foundation Unit

KEGG, Kyoto Encyclopedia of Genes and Genomes

KNN, *k*-nearest neighbor

KS, Kolmogorov-Smirnov

LC, liquid chromatography

LDA, linear discriminant analysis

LF, Lilliefors

LLE, liquid–liquid extraction

LOD, limit of detection

LOESS, local polynomial regression

LOQ, limit of quantitation

LPC, lysophosphatidylcholine

LPE, lysophosphatidylethanolamine

MDS, multidimensional scaling

MeOH, methanol

METLIN, Metabolites and Tandem MS Database

MetS, metabolic syndrome

MF, molecular feature

MG, monoacylglycerol

MIDAS, autosampler

Mo, month

MPP, mass profiler professional

MRM, multiple reaction monitoring

MS, mass spectrometry

MSTUS, mass spectrometry total useful signal

MTBE, methyl *tert*-butyl ether

MUFA, monounsaturated fatty acid

N, observations

NIRS, near-infrared spectroscopy

NIST, National Institute of Standards and Technology

NLS, neutral loss scan mode

NMR, nuclear magnetic resonance

OA, oleate

OAHOA, oleic acid-hydroxy-oleic acid

OAHSA, oleic acid-hydroxy-stearic acid

OFTT, oral fat tolerance test

OGTT, oral glucose tolerance test

OLTT, oral lipid tolerance test

PA, glycerophosphatidic acid

PAHOA, palmitic acid-hydroxy-oleic acid

PAHPA, palmitic acid-hydroxy-palmitic acid

PAHSA, palmitic acid hydroxystearic acid

PBMC, peripheral blood mononuclear cell

PO, palmitoleate

POHPA, palmitoleic acid-hydroxy-palmitic acid

POHPO, palmitoleic acid-hydroxy-palmitoleic acid

pAUC, partial area under the curve

PC, glycerophosphatidylcholine

PCA, principal component analysis

PE, glycerophosphatidylethanolamine

PFTBA, perfluorotri-*n*-butylamine

PG, glycerophosphatidylglyceride

PIS, product ion scanning

PISc, pathway impact score

PLS, partial least squares

PLS-DA, partial least squares discriminant analysis

PNN, probabilistic neural networks

POA, pathway overrepresentation analysis

PrIS, precursor ion scanning

PS, glycerophosphatidylserine

PT1D, children that developed type 1 diabetes mellitus

PUFA, polyunsaturated fatty acid

P1Ab, children not diagnosed with type 1 diabetes mellitus

QC, quality control

QqQ, triple quadrupole

Q1, first quadrupole

q2, collision cell (second quadrupole)

Q3, third quadrupole

QTOF, quadrupole–time of flight

RI, retention index

RM, reporter metabolite

ROC, receiver-operating characteristic

ROS, reactive oxygen specie

RSD, relative standard deviation

RT, retention time

RU, relative unit

SA, stearate

SAHOA, stearic acid-hydroxy-oleic acid

SCFA, short chain fatty acid

SD, standard deviation

SFA, saturated fatty acid

SIM, selected ion monitoring

SLE, solid–liquid extraction

SM, sphingomyelin

SMM, sphingolipid metabolism

SOM, self-organizing map

SPE, solid-phase extraction

SPME, solid-phase microextraction

SRM, selected reaction monitoring

SVM, support vector machine

SW, Shapiro-Wilk

T1DM, type 1 diabetes mellitus

T2DM, type 2 diabetes mellitus

TG, triglyceride

TIC, total ion current

TMCS, trimethylchlorosilane

TMS, trimethylsilyl

TPP, triphenyl phosphate

TUDCA, tauroursodeoxycholic acid

UPLC, ultra-performance liquid chromatography

V-PDB, Vienna Pee Dee Belemnite

VIP, Variable Importance in Projection

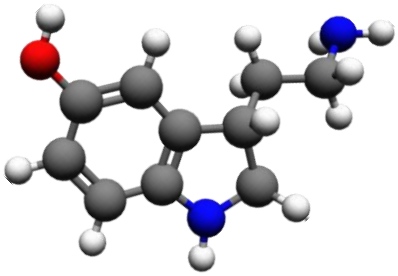
1H NMR, proton nuclear magnetic resonance

12-PAHSA, palmitic acid-12-hydroxy-stearic acid

12-POHSA, palmitoleic acid-12-hydroxy-stearic acid

12-SAHSA, stearic acid-12-hydroxy-stearic acid

12-OAHSA, oleic acid-12-hydroxy-stearic acid



IMIBIC
INSTITUTO MAIMÓNIDES DE INVESTIGACIÓN BIOMÉDICA DE CÓRDOBA

UNIVERSIDAD DE CÓRDOBA

ceiA3
campus de excelencia internacional agroalimentario